THE ASTRONOMER'S GUIDE TO THE UNIVERSE



Alessandro Baldi
Butte College

LibreTexts™

The Astronomer's Guide to the Universe

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

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

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



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

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

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

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

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

This text was compiled on 08/23/2026

TABLE OF CONTENTS

Licensing

1: The Modern Cosmos

1.1: A Tour of the Universe

1.2: Consequences of Light Travel Time

1.3: The Universe on the Large Scale

1.4: Numbers in Astronomy

2: The Universe through Your Eyes

2.1: Thinking Ahead

2.2: The Sky Above

2.3: Earth and Sky

2.4: The Seasons

2.5: Phases and Motions of the Moon

2.6: Ocean Tides and the Moon

2.7: Eclipses of the Sun and Moon

3: The Science of Astronomy

3.1: Ancient Astronomy

3.2: Astrology and Astronomy

3.3: The Birth of Modern Astronomy

3.4: The Laws of Planetary Motion

3.5: The Nature of Science

4: Making Sense of the Universe

4.1: Newton's Great Synthesis

4.2: Newton's Universal Law of Gravitation

5: Light- Messenger of the Cosmos

5.1: The Behavior of Light

5.2: The Electromagnetic Spectrum

5.3: Spectroscopy in Astronomy

5.4: The Structure of the Atom

5.5: Formation of Spectral Lines

5.6: The Doppler Effect

5.7: Telescopes

5.8: Telescopes Today

5.9: Visible-Light Detectors and Instruments

5.10: Radio Telescopes

5.11: Observations outside Earth's Atmosphere

5.12: The Future of Large Telescopes

6: Introduction to the Solar System

6.1: Overview of Our Planetary System

6.2: Composition and Structure of Planets

6.3: Dating Planetary Surfaces

6.4: Origin of the Solar System

6.5: Formation of the Solar System

7: Worlds Around Other Stars

7.1: Evidence That Planets Form around Other Stars

7.2: Planets beyond the Solar System- Search and Discovery

7.3: Exoplanets Everywhere - What we are Learning

7.4: New Perspectives on Planet Formation

8: Terrestrial Planets

8.1: Life, Chemical Evolution, and Climate Change

8.2: The Massive Atmosphere of Venus

8.3: Water and Life on Mars

8.4: Divergent Planetary Evolution

9: Jovian Planets

9.1: Exploring the Outer Planets

9.2: The Giant Planets

9.3: The Gailean Moons of Jupiter

9.4: Titan and Triton

10: Asteroids, Comets, and Dwarf Planets

10.1: Asteroids

10.2: Meteors

10.3: Meteorites - Stones from Heaven

10.4: The "Long-Haired" Comets

10.5: The Origin and Fate of Comets and Related Objects

10.6: Pluto and Charon

10.7: Cosmic Influences on the Evolution of Earth

11: The Sun, Our Star

11.1: Sources of Sunshine- Thermal and Gravitational Energy

11.2: Mass, Energy, and the Theory of Relativity

11.3: The Structure and Composition of the Sun

11.4: The Solar Interior - Theory

11.5: The Solar Interior - Observations

11.6: The Solar Cycle

11.7: Solar Activity above the Photosphere

11.8: Space Weather

12: The Stellar Census

12.1: The Brightness of Stars

12.2: Surveying the Stars

12.3: Colors of Stars

12.4: The Spectra of Stars (and Brown Dwarfs)

12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion

12.6: A Stellar Census

12.7: Measuring Stellar Masses

12.8: Diameters of Stars

12.9: The H-R Diagram

12.10: Star Clusters

12.11: Checking Out the Theory

13: Stellar Evolution

13.1: Star Formation

13.2: The H-R and the Study of Stellar Evolution

13.3: Evolution from the Main Sequence to Red Giants

13.4: Further Evolution of Stars

13.5: The Evolution of More Massive Stars

14: White Dwarfs, Neutron Stars, and Black Holes

14.1: The Death of Low-Mass Stars

14.2: Evolution of Massive Stars- An Explosive Finish

14.3: Supernova Observations

14.4: Pulsars and the Discovery of Neutron Stars

14.5: The Evolution of Binary Star Systems

14.6: Black Holes

14.7: Evidence for Black Holes

15: The Milky Way, Our Island Universe

15.1: The Architecture of the Galaxy

15.2: The Interstellar Medium

15.3: Interstellar Gas

15.4: Cosmic Dust

15.5: The Life Cycle of Cosmic Material

15.6: Spiral Structure

15.7: The Mass of the Galaxy

15.8: Stellar Populations in the Galaxy

15.9: The Formation of the Galaxy

15.10: The Center of the Galaxy

16: The Expanding Universe of Galaxies

16.1: The Discovery of Galaxies

16.2: The Extragalactic Distance Scale
16.3: The Expanding Universe
16.4: Types of Galaxies
16.5: Properties of Galaxies
16.6: Quasars
16.7: Supermassive Black Holes- What Quasars really are
16.8: Quasars as Probes of Evolution in the Universe

17: The Big Bang and the Origin of the Cosmos

17.1: The Beginning of the Universe
17.2: The Cosmic Microwave Background
17.3: The Inflationary Universe

18: Dark Matter and Dark Energy

18.1: The Challenge of Dark Matter
18.2: The Age of the Universe
18.3: A Model of the Universe
18.4: What Is the Universe Really Made Of?

19: Life in the Universe

19.1: The Cosmic Context for Life
19.2: Astrobiology
19.3: Searching for Life beyond Earth
19.4: The Search for Extraterrestrial Intelligence

Index

Glossary

Detailed Licensing

Detailed Licensing

This page titled The Astronomer's Guide to the Universe is shared under a license and was authored, remixed, and/or curated by Alessandro Baldi.

Licensing

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

This page titled Licensing is shared under a license and was authored, remixed, and/or curated by Alessandro Baldi.

1: THE MODERN COSMOS

1.1: A Tour of the Universe
1.2: Consequences of Light Travel Time
1.3: The Universe on the Large Scale
1.4: Numbers in Astronomy

This page titled 1: The Modern Cosmos is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

1.1: A Tour of the Universe

We can now take a brief introductory tour of the universe as astronomers understand it today to get acquainted with the types of objects and distances you will encounter throughout the text. We begin at home with Earth, a nearly spherical planet about 13,000 kilometers in diameter (Figure 1.6). A space traveler entering our planetary system would easily distinguish Earth from the other planets in our solar system by the large amount of liquid water that covers some two thirds of its crust. If the traveler had equipment to receive radio or television signals, or came close enough to see the lights of our cities at night, she would soon find signs that this watery planet has sentient life.

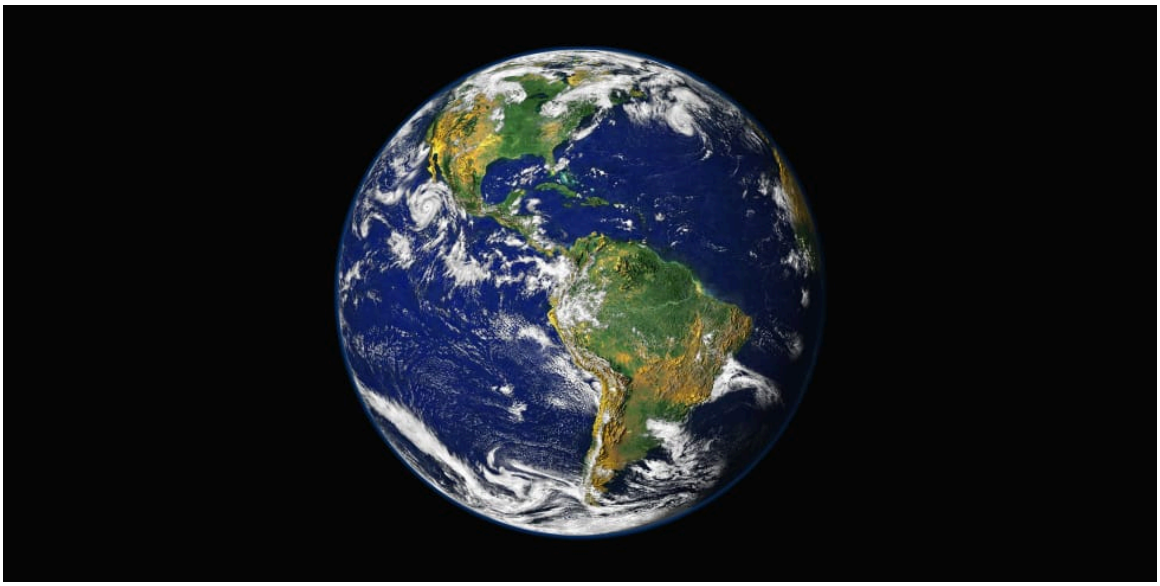


Figure 1.6 : Humanity's Home Base. This image shows the Western hemisphere as viewed from space 35,400 kilometers (about 22,000 miles) above Earth. Data about the land surface from one satellite was combined with another satellite's data about the clouds to create the image. (credit: modification of work by R. Stockli, A. Nelson, F. Hasler, NASA/ GSFC/ NOAA/ USGS)

Our nearest astronomical neighbor is Earth's satellite, commonly called the *Moon*. Figure 1.7 shows Earth and the Moon drawn to scale on the same diagram. Notice how small we have to make these bodies to fit them on the page with the right scale. The Moon's distance from Earth is about 30 times Earth's diameter, or approximately 384,000 kilometers, and it takes about a month for the Moon to revolve around Earth. The Moon's diameter is 3476 kilometers, about one fourth the size of Earth.



Figure 1.7 : Earth and Moon, Drawn to Scale. This image shows Earth and the Moon shown to scale for both size and distance. (credit: modification of work by NASA)

Light (or radio waves) takes 1.3 seconds to travel between Earth and the Moon. If you've seen videos of the Apollo flights to the Moon, you may recall that there was a delay of about 3 seconds between the time Mission Control asked a question and the time the astronauts responded. This was not because the astronauts were thinking slowly, but rather because it took the radio waves almost 3 seconds to make the round trip.

Earth revolves around our star, the Sun, which is about 150 million kilometers away—approximately 400 times as far away from us as the Moon. We call the average Earth–Sun distance an *astronomical unit* (AU) because, in the early days of astronomy, it was the most important measuring standard. Light takes slightly more than 8 minutes to travel 1 astronomical unit, which means the latest news we receive from the Sun is always 8 minutes old. The diameter of the Sun is about 1.5 million kilometers; Earth could fit comfortably inside one of the minor eruptions that occurs on the surface of our star. If the Sun were reduced to the size of a

basketball, Earth would be a small apple seed about 30 meters from the ball.

It takes Earth 1 year (3×10^7 seconds) to go around the Sun at our distance; to make it around, we must travel at approximately 110,000 kilometers per hour. (If you, like many students, still prefer miles to kilometers, you might find the following trick helpful. To convert kilometers to miles, just multiply kilometers by 0.6. Thus, 110,000 kilometers per hour becomes 66,000 miles per hour.) Because gravity holds us firmly to Earth and there is no resistance to Earth's motion in the vacuum of space, we participate in this extremely fast-moving trip without being aware of it day to day.

Earth is only one of eight planets that revolve around the Sun. These planets, along with their moons and swarms of smaller bodies such as dwarf planets, make up the solar system (Figure 1.8). A planet is defined as a body of significant size that orbits a star and does not produce its own light. (If a large body consistently produces its own light, it is then called a *star*.) Later in the book this definition will be modified a bit, but it is perfectly fine for now as you begin your voyage.

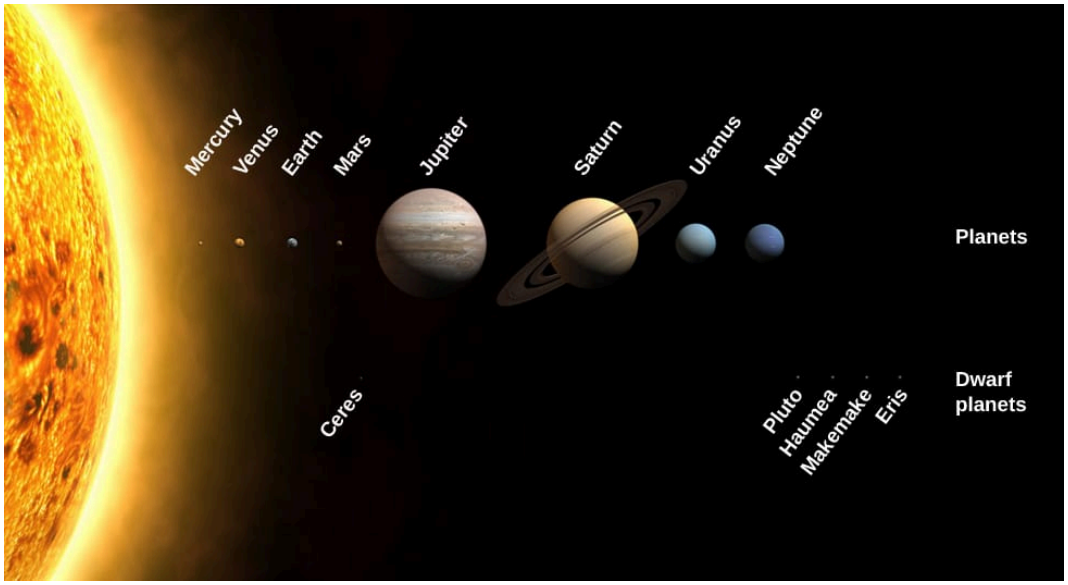


Figure 1.8 : Our Solar Family. The Sun, the planets, and some dwarf planets are shown with their sizes drawn to scale. The orbits of the planets are much more widely separated than shown in this drawing. Notice the size of Earth compared to the giant planets. (credit: modification of work by NASA)

We are able to see the nearby planets in our skies only because they reflect the light of our local star, the Sun. If the planets were much farther away, the tiny amount of light they reflect would usually not be visible to us. The planets we have so far discovered orbiting other stars were found from the pull their gravity exerts on their parent stars, or from the light they block from their stars when they pass in front of them. We can't see most of these planets directly, although a few are now being imaged directly.

The Sun is our local star, and all the other stars are also enormous balls of glowing gas that generate vast amounts of energy by nuclear reactions deep within. We will discuss the processes that cause stars to shine in more detail later in the book. The other stars look faint only because they are so very far away. If we continue our basketball analogy, Proxima Centauri, the nearest star beyond the Sun, which is 4.3 light-years away, would be almost 7000 kilometers from the basketball.

When you look up at a star-filled sky on a clear night, all the stars visible to the unaided eye are part of a single collection of stars we call the *Milky Way Galaxy*, or simply the *Galaxy*. (When referring to the Milky Way, we capitalize *Galaxy*; when talking about other galaxies of stars, we use lowercase *galaxy*.) The Sun is one of hundreds of billions of stars that make up the Galaxy; its extent, as we will see, staggers the human imagination. Within a sphere 10 light-years in radius centered on the Sun, we find roughly ten stars. Within a sphere 100 light-years in radius, there are roughly 10,000 (10^4) stars—far too many to count or name—but we have still traversed only a tiny part of the Milky Way Galaxy. Within a 1000-light-year sphere, we find some ten million (10^7) stars; within a sphere of 100,000 light-years, we finally encompass the entire Milky Way Galaxy.

Our Galaxy looks like a giant disk with a small ball in the middle. If we could move outside our Galaxy and look down on the disk of the Milky Way from above, it would probably resemble the galaxy in Figure 1.9, with its spiral structure outlined by the blue light of hot adolescent stars.



Figure 1.9 : Spiral Galaxy. This galaxy of billions of stars, called by its catalog number NGC 1073, is thought to be similar to our own Milky Way Galaxy. Here we see the giant wheel-shaped system with a bar of stars across its middle. (credit: NASA, ESA)

The Sun is somewhat less than 30,000 light-years from the center of the Galaxy, in a location with nothing much to distinguish it. From our position inside the Milky Way Galaxy, we cannot see through to its far rim (at least not with ordinary light) because the space between the stars is not completely empty. It contains a sparse distribution of gas (mostly the simplest element, hydrogen) intermixed with tiny solid particles that we call *interstellar dust*. This gas and dust collect into enormous clouds in many places in the Galaxy, becoming the raw material for future generations of stars. Figure 1.10 shows an image of the disk of the Galaxy as seen from our vantage point.



Figure 1.10 : Milky Way Galaxy. Because we are inside the Milky Way Galaxy, we see its disk in cross-section flung across the sky like a great milky white avenue of stars with dark “rifts” of dust. In this dramatic image, part of it is seen above Trona Pinnacles in the California desert. (credit: Ian Norman)

Typically, the interstellar material is so extremely sparse that the space between stars is a much better vacuum than anything we can produce in terrestrial laboratories. Yet, the dust in space, building up over thousands of light-years, can block the light of more distant stars. Like the distant buildings that disappear from our view on a smoggy day in Los Angeles, the more distant regions of the Milky Way cannot be seen behind the layers of interstellar smog. Luckily, astronomers have found that stars and raw material shine with various forms of light, some of which do penetrate the smog, and so we have been able to develop a pretty good map of the Galaxy.

Recent observations, however, have also revealed a rather surprising and disturbing fact. There appears to be more—much more—to the Galaxy than meets the eye (or the telescope). From various investigations, we have evidence that much of our Galaxy is made of material we cannot currently observe directly with our instruments. We therefore call this component of the Galaxy *dark matter*. We know the dark matter is there by the pull its gravity exerts on the stars and raw material we can observe, but what this dark matter is made of and how much of it exists remain a mystery. Furthermore, this dark matter is not confined to our Galaxy; it appears to be an important part of other star groupings as well.

By the way, not all stars live by themselves, as the Sun does. Many are born in double or triple systems with two, three, or more stars revolving about each other. Because the stars influence each other in such close systems, multiple stars allow us to measure characteristics that we cannot discern from observing single stars. In a number of places, enough stars have formed together that we recognized them as star clusters (Figure 1.11). Some of the largest of the star clusters that astronomers have cataloged contain hundreds of thousands of stars and take up volumes of space hundreds of light-years across.



Figure 1.11 : Star Cluster. This large star cluster is known by its catalog number, M9. It contains some 250,000 stars and is seen more clearly from space using the Hubble Space Telescope. It is located roughly 25,000 light-years away. (credit: NASA, ESA)

You may hear stars referred to as “eternal,” but in fact no star can last forever. Since the “business” of stars is making energy, and energy production requires some sort of fuel to be used up, eventually all stars run out of fuel. This news should not cause you to panic, though, because our Sun still has at least 5 or 6 billion years to go. Ultimately, the Sun and all stars will die, and it is in their death throes that some of the most intriguing and important processes of the universe are revealed. For example, we now know that many of the atoms in our bodies were once inside stars. These stars exploded at the ends of their lives, recycling their material back into the reservoir of the Galaxy. In this sense, all of us are literally made of recycled “star dust.”

This page titled 1.1: A Tour of the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 1.7: A Tour of the Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

1.2: Consequences of Light Travel Time

There is another reason the speed of light is such a natural unit of distance for astronomers. Information about the universe comes to us almost exclusively through various forms of light, and all such light travels at the speed of light—that is, 1 light-year every year. This sets a limit on how quickly we can learn about events in the universe. If a star is 100 light-years away, the light we see from it tonight left that star 100 years ago and is just now arriving in our neighborhood. The soonest we can learn about any changes in that star is 100 years after the fact. For a star 500 light-years away, the light we detect tonight left 500 years ago and is carrying 500-year-old news.

Because many of us are accustomed to instant news from the Internet, some might find this frustrating.

"You mean, when I see that star up there," you ask, "I won't know what's actually happening there for another 500 years?"

But this isn't the most helpful way to think about the situation. For astronomers, *now* is when the light reaches us here on Earth. There is no way for us to know anything about that star (or other object) until its light reaches us.

But what at first may seem a great frustration is actually a tremendous benefit in disguise. If astronomers really want to piece together what has happened in the universe since its beginning, they must find evidence about each epoch (or period of time) of the past. Where can we find evidence today about cosmic events that occurred billions of years ago?

The delay in the arrival of light provides an answer to this question. The farther out in space we look, the longer the light has taken to get here, and the longer ago it left its place of origin. By looking billions of light-years out into space, astronomers are actually seeing billions of years into the past. In this way, we can reconstruct the history of the cosmos and get a sense of how it has evolved over time.

This is one reason why astronomers strive to build telescopes that can collect more and more of the faint light in the universe. The more light we collect, the fainter the objects we can observe. On average, fainter objects are farther away and can, therefore, tell us about periods of time even deeper in the past. Instruments such as the Hubble Space Telescope (Figure 1.5) and the Very Large Telescope in Chile (which you will learn about in the chapter on Astronomical Instruments), are giving astronomers views of deep space and deep time better than any we have had before.



Figure 1.5 : Telescope in Orbit. The Hubble Space Telescope, shown here in orbit around Earth, is one of many astronomical instruments in space. (credit: modification of work by European Space Agency)

This page titled 1.2: Consequences of Light Travel Time is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 1.6: Consequences of Light Travel Time by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

1.3: The Universe on the Large Scale

In a very rough sense, you could think of the solar system as your house or apartment and the Galaxy as your town, made up of many houses and buildings. In the twentieth century, astronomers were able to show that, just as our world is made up of many, many towns, so the universe is made up of enormous numbers of galaxies. (We define the universe to be everything that exists that is accessible to our observations.) Galaxies stretch as far into space as our telescopes can see, many billions of them within the reach of modern instruments. When they were first discovered, some astronomers called galaxies *island universes*, and the term is aptly descriptive; galaxies do look like islands of stars in the vast, dark seas of intergalactic space.

The nearest galaxy, discovered in 1993, is a small one that lies 70,000 light-years from the Sun in the direction of the constellation Sagittarius, where the smog in our own Galaxy makes it especially difficult to discern. (A constellation, we should note, is one of the 88 sections into which astronomers divide the sky, each named after a prominent star pattern within it.) Beyond this Sagittarius dwarf galaxy lie two other small galaxies, about 160,000 light-years away. First recorded by Magellan's crew as he sailed around the world, these are called the *Magellanic Clouds* (Figure 1.12). All three of these small galaxies are satellites of the Milky Way Galaxy, interacting with it through the force of gravity. Ultimately, all three may even be swallowed by our much larger Galaxy, as other small galaxies have been over the course of cosmic time.



Figure 1.12 : Neighbor Galaxies. This image shows both the Large Magellanic Cloud and the Small Magellanic Cloud above the telescopes of the Atacama Large Millimeter/Submillimeter Array (ALMA) in the Atacama Desert of northern Chile. (credit: ESO, C. Malin)

The nearest large galaxy is a spiral quite similar to our own, located in the constellation of Andromeda, and is thus called the Andromeda galaxy; it is also known by one of its catalog numbers, M31 (Figure 1.13). M31 is a little more than 2 million light-years away and, along with the Milky Way, is part of a small cluster of more than 50 galaxies referred to as the *Local Group*.

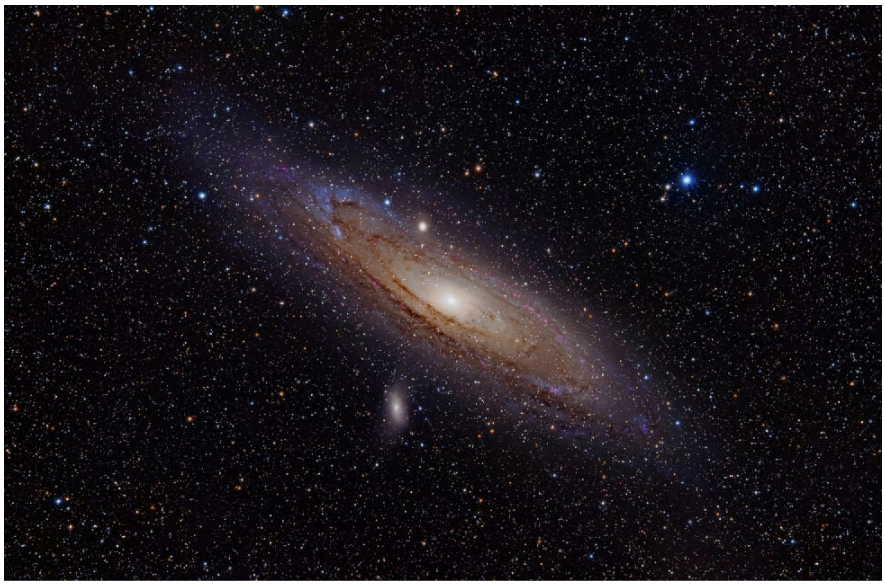


Figure 1.13 : Closest Spiral Galaxy. The Andromeda galaxy (M31) is a spiral-shaped collection of stars similar to our own Milky Way. (credit: Adam Evans)

At distances of 10 to 15 million light-years, we find other small galaxy groups, and then at about 50 million light-years there are more impressive systems with thousands of member galaxies. We have discovered that galaxies occur mostly in clusters, both large and small (Figure 1.14).



Figure 1.14 : Fornax Cluster of Galaxies. In this image, you can see part of a cluster of galaxies located about 60 million light-years away in the constellation of Fornax. All the objects that are not pinpoints of light in the picture are galaxies of billions of stars. (credit: ESO, J. Emerson, VISTA. Acknowledgment: Cambridge Astronomical Survey Unit)

Some of the clusters themselves form into larger groups called *superclusters*. The Local Group is part of a supercluster of galaxies, called the Virgo Supercluster, which stretches over a diameter of 110 million light-years. We are just beginning to explore the structure of the universe at these enormous scales and are already encountering some unexpected findings.

At even greater distances, where many ordinary galaxies are too dim to see, we find *quasars*. These are brilliant centers of galaxies, glowing with the light of an extraordinarily energetic process. The enormous energy of the quasars is produced by gas that is heated to a temperature of millions of degrees as it falls toward a massive black hole and swirls around it. The brilliance of quasars makes them the most distant beacons we can see in the dark oceans of space. They allow us to probe the universe 10 billion light-years away or more, and thus 10 billion years or more in the past.

With quasars we can see way back close to the Big Bang explosion that marks the beginning of time. Beyond the quasars and the most distant visible galaxies, we have detected the feeble glow of the explosion itself, filling the universe and thus coming to us from all directions in space. The discovery of this “afterglow of creation” is considered to be one of the most significant events in twentieth-century science, and we are still exploring the many things it has to tell us about the earliest times of the universe.

Measurements of the properties of galaxies and quasars in remote locations require large telescopes, sophisticated light-amplifying devices, and painstaking labor. Every clear night, at observatories around the world, astronomers and students are at work on such mysteries as the birth of new stars and the large-scale structure of the universe, fitting their results into the tapestry of our understanding.

This page titled 1.3: The Universe on the Large Scale is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 1.8: The Universe on the Large Scale by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

1.4: Numbers in Astronomy

In astronomy we deal with distances on a scale you may never have thought about before, with numbers larger than any you may have encountered. We adopt two approaches that make dealing with astronomical numbers a little bit easier. First, we use a system for writing large and small numbers called *scientific notation* (or sometimes *powers-of-ten notation*). This system is very appealing because it eliminates the many zeros that can seem overwhelming to the reader. In scientific notation, if you want to write a number such as 500,000,000, you express it as Math input error *Système International d'Unités*). The metric system is summarized in Appendix D (see Example 1.2).

Link to Learning

Watch this brief PBS animation that explains how scientific notation works and why it's useful.

A common unit astronomers use to describe distances in the universe is a light-year, which is the distance light travels during one year. Because light always travels at the same speed, and because its speed turns out to be the fastest possible speed in the universe, it makes a good standard for keeping track of distances. You might be confused because a "light-year" seems to imply that we are measuring time, but this mix-up of time and distance is common in everyday life as well. For example, when your friend asks where the movie theater is located, you might say "about 20 minutes from downtown."

So, how many kilometers are there in a light-year? Light travels at the amazing pace of Math input error

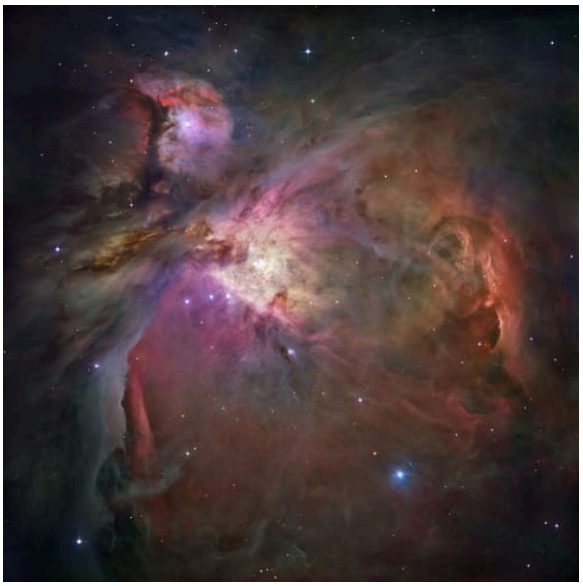


Figure 1.4 : Orion Nebula. This beautiful cloud of cosmic raw material (gas and dust from which new stars and planets are being made) called the Orion Nebula is about 1400 light-years away. That's a distance of roughly 1.34×10^{16} kilometers—a pretty big number. The gas and dust in this region are illuminated by the intense light from a few extremely energetic adolescent stars. (credit: NASA, ESA, M. Robberto (Space Telescope Science Institute/ESA) and the Hubble Space Telescope Orion Treasury Project Team)

Example 1.1

Scientific Notation

In 2015, the richest human being on our planet had a net worth of $79.2 billion. Some might say this is an astronomical sum of money. Express this amount in scientific notation.

Solution

$79.2 billion can be written $79,200,000,000. Expressed in scientific notation it becomes $\$7.92 \times 10^{10}$

Example 1.2

Getting Familiar with a Light-Year

How many kilometers are there in a light-year?

Solution

Light travels 3×10^5 km in 1 s. So, let's calculate how far it goes in a year:

- There are 60 $\left(6 \times 10^1\right)$ s in 1 min, and 6×10^1 min in 1 h.
- Multiply these together and you find that there are 3.6×10^3 s/h.
- Thus, light covers $3 \times 10^5 \ \mathrm{km/s} \times 3.6 \times 10^3 \ \mathrm{s/h} = 1.08 \times 10^9 \ \mathrm{km/h}$.
- There are 24 or 2.4×10^1 h in a day, and 365.25 $\left(3.65 \times 10^2\right)$ days in 1 y.
- The product of these two numbers is 8.77×10^3 h/y.
- Multiplying this by 1.08×10^9 km/h gives 9.46×10^{12} km/light-year.

That's almost 10,000,000,000,000 km that light covers in a year. To help you imagine how long this distance is, we'll mention that a string 1 light-year long could fit around the circumference of Earth 236 million times.

This page titled 1.4: Numbers in Astronomy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 1.5: Numbers in Astronomy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2: THE UNIVERSE THROUGH YOUR EYES

2.1: Thinking Ahead
2.2: The Sky Above
2.3: Earth and Sky
2.4: The Seasons
2.5: Phases and Motions of the Moon
2.6: Ocean Tides and the Moon
2.7: Eclipses of the Sun and Moon

This page titled 2: The Universe through Your Eyes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.1: Thinking Ahead



Figure 2.1 : Night Sky. In this panoramic photograph of the night sky from the Atacama Desert in Chile, we can see the central portion of the Milky Way Galaxy arcing upward in the center of the frame. On the left, the Large Magellanic Cloud and the Small Magellanic Cloud (smaller galaxies that orbit the Milky Way Galaxy) are easily visible from the Southern Hemisphere. (credit: modification of work by ESO/Y. Beletsky)

Much to your surprise, a member of the Flat Earth Society moves in next door. He believes that Earth is flat and all the NASA images of a spherical Earth are either faked or simply show the round (but flat) disk of Earth from above. How could you prove to your new neighbor that Earth really is a sphere? (When you've thought about this on your own, you can check later in the chapter for some suggested answers.)

Today, few people really spend much time looking at the night sky. In ancient days, before electric lights robbed so many people of the beauty of the sky, the stars and planets were an important aspect of everyone's daily life. All the records that we have—on paper and in stone—show that ancient civilizations around the world noticed, worshipped, and tried to understand the lights in the sky and fit them into their own view of the world. These ancient observers found both majestic regularity and never-ending surprise in the motions of the heavens. Through their careful study of the planets, the Greeks and later the Romans laid the foundation of the science of astronomy.

This page titled 2.1: Thinking Ahead is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax.

- 2.1: Thinking Ahead by OpenStax is licensed CC BY 4.0.

2.2: The Sky Above

Learning Objectives

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

- Define the main features of the celestial sphere
- Explain the system astronomers use to describe the sky
- Describe how motions of the stars appear to us on Earth
- Describe how motions of the Sun, Moon, and planets appear to us on Earth
- Understand the modern meaning of the term *constellation*

Our senses suggest to us that Earth is the center of the universe—the hub around which the heavens turn. This geocentric (Earth-centered) view was what almost everyone believed until the European Renaissance. After all, it is simple, logical, and seemingly self-evident. Furthermore, the geocentric perspective reinforced those philosophical and religious systems that taught the unique role of human beings as the central focus of the cosmos. However, the geocentric view happens to be wrong. One of the great themes of our intellectual history is the overthrow of the geocentric perspective. Let us, therefore, take a look at the steps by which we reevaluated the place of our world in the cosmic order.

The Celestial Sphere

If you go on a camping trip or live far from city lights, your view of the sky on a clear night is pretty much identical to that seen by people all over the world before the invention of the telescope. Gazing up, you get the impression that the sky is a great hollow dome with you at the center (Figure 2.2), and all the stars are an equal distance from you on the surface of the dome. The top of that dome, the point directly above your head, is called the zenith, and where the dome meets Earth is called the horizon. From the sea or a flat prairie, it is easy to see the horizon as a circle around you, but from most places where people live today, the horizon is at least partially hidden by mountains, trees, buildings, or smog.

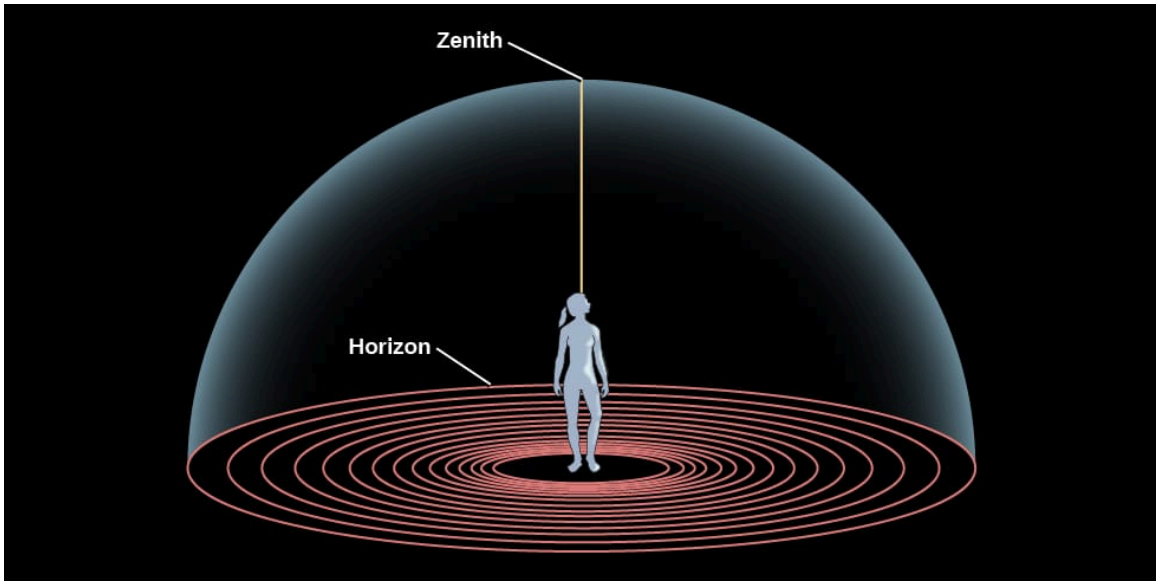


Figure 2.2 : The Sky around Us. The horizon is where the sky meets the ground; an observer's zenith is the point directly overhead.

If you lie back in an open field and observe the night sky for hours, as ancient shepherds and travelers regularly did, you will see stars rising on the eastern horizon (just as the Sun and Moon do), moving across the dome of the sky in the course of the night, and setting on the western horizon. Watching the sky turn like this night after night, you might eventually get the idea that the dome of the sky is really part of a great sphere that is turning around you, bringing different stars into view as it turns. The early Greeks regarded the sky as just such a celestial sphere (Figure 2.3). Some thought of it as an actual sphere of transparent

crystalline material, with the stars embedded in it like tiny jewels.

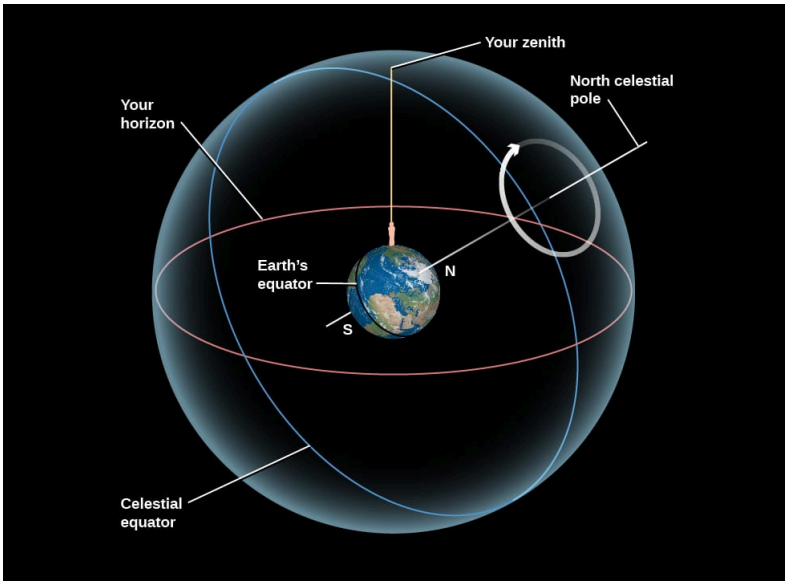


Figure 2.3 : Circles on the Celestial Sphere. Here we show the (imaginary) celestial sphere around Earth, on which objects are fixed, and which rotates around Earth on an axis. In reality, it is Earth that turns around this axis, creating the illusion that the sky revolves around us. Note that Earth in this picture has been tilted so that your location is at the top and the North Pole is where the N is. The apparent motion of celestial objects in the sky around the pole is shown by the circular arrow.

Today, we know that it is not the celestial sphere that turns as night and day proceed, but rather the planet on which we live. We can put an imaginary stick through Earth's North and South Poles, representing our planet's axis. It is because Earth turns on this axis every 24 hours that we see the Sun, Moon, and stars rise and set with clockwork regularity. Today, we know that these celestial objects are not really on a dome, but at greatly varying distances from us in space. Nevertheless, it is sometimes still convenient to talk about the celestial dome or sphere to help us keep track of objects in the sky. There is even a special theater, called a *planetarium*, in which we project a simulation of the stars and planets onto a white dome.

As the celestial sphere rotates, the objects on it maintain their positions with respect to one another. A grouping of stars such as the Big Dipper has the same shape during the course of the night, although it turns with the sky. During a single night, even objects we know to have significant motions of their own, such as the nearby planets, seem fixed relative to the stars. Only meteors—brief "shooting stars" that flash into view for just a few seconds—move appreciably with respect to other objects on the celestial sphere. (This is because they are not stars at all. Rather, they are small pieces of cosmic dust, burning up as they hit Earth's atmosphere.) We can use the fact that the entire celestial sphere seems to turn together to help us set up systems for keeping track of what things are visible in the sky and where they happen to be at a given time.

Celestial Poles and Celestial Equator

To help orient us in the turning sky, astronomers use a system that extends Earth's axis points into the sky. Imagine a line going through Earth, connecting the North and South Poles. This is Earth's axis, and Earth rotates about this line. If we extend this imaginary line outward from Earth, the points where this line intersects the celestial sphere are called the *north celestial pole* and the *south celestial pole*. As Earth rotates about its axis, the sky appears to turn in the opposite direction around those celestial poles (Figure 2.4). We also (in our imagination) throw Earth's equator onto the sky and call this the celestial equator. It lies halfway between the celestial poles, just as Earth's equator lies halfway between our planet's poles.



Figure 2.4 : Circling the South Celestial Pole. This long-exposure photo shows trails left by stars as a result of the apparent rotation of the celestial sphere around the south celestial pole. (In reality, it is Earth that rotates.) (Credit: ESO/Iztok Bončina)

Now let's imagine how riding on different parts of our spinning Earth affects our view of the sky. The apparent motion of the celestial sphere depends on your latitude (position north or south of the equator). First of all, notice that Earth's axis is pointing at the celestial poles, so these two points in the sky do not appear to turn.

If you stood at the North Pole of Earth, for example, you would see the north celestial pole overhead, at your zenith. The celestial equator, 90° from the celestial poles, would lie along your horizon. As you watched the stars during the course of the night, they would all circle around the celestial pole, with none rising or setting. Only that half of the sky north of the celestial equator is ever visible to an observer at the North Pole. Similarly, an observer at the South Pole would see only the southern half of the sky.

If you were at Earth's equator, on the other hand, you see the celestial equator (which, after all, is just an "extension" of Earth's equator) pass overhead through your zenith. The celestial poles, being 90° from the celestial equator, must then be at the north and south points on your horizon. As the sky turns, all stars rise and set; they move straight up from the east side of the horizon and set straight down on the west side. During a 24-hour period, all stars are above the horizon exactly half the time. (Of course, during some of those hours, the Sun is too bright for us to see them.)

What would an observer in the latitudes of the United States or Europe see? Remember, we are neither at Earth's pole nor at the equator, but in between them. For those in the continental United States and Europe, the north celestial pole is neither overhead nor on the horizon, but in between. It appears above the northern horizon at an angular height, or altitude, equal to the observer's latitude. In San Francisco, for example, where the latitude is 38° N, the north celestial pole is 38° above the northern horizon.

For an observer at 38° N latitude, the south celestial pole is 38° below the southern horizon and, thus, never visible. As Earth turns, the whole sky seems to pivot about the north celestial pole. For this observer, stars within 38° of the North Pole can never set. They are always above the horizon, day and night. This part of the sky is called the north circumpolar zone. For observers in the continental United States, the Big Dipper, Little Dipper, and Cassiopeia are examples of star groups in the north circumpolar zone. On the other hand, stars within 38° of the south celestial pole never rise. That part of the sky is the south circumpolar zone. To most U.S. observers, the Southern Cross is in that zone. (Don't worry if you are not familiar with the star groups just mentioned; we will introduce them more formally later on.)

At this particular time in Earth's history, there happens to be a star very close to the north celestial pole. It is called Polaris, the pole star, and has the distinction of being the star that moves the least amount as the northern sky turns each day. Because it moved so little while the other stars moved much more, it played a special role in the mythology of several Native American tribes, for example (some called it the "fastener of the sky").

Astronomy Basics: What's Your Angle?

Astronomers measure how far apart objects appear in the sky by using angles. By definition, there are 360° in a circle, so a circle stretching completely around the celestial sphere contains 360°. The half-sphere or dome of the sky then contains 180° from horizon to opposite horizon. Thus, if two stars are 18° apart, their separation spans about 1/10 of the dome of the sky. To give you a sense of how big a degree is, the full Moon is about half a degree across. This is about the width of your smallest finger (pinkie) seen at arm's length.

Rising and Setting of the Sun

We described the movement of stars in the night sky, but what about during the daytime? The stars continue to circle during the day, but the brilliance of the Sun makes them difficult to see. (The Moon can often be seen in the daylight, however.) On any given day, we can think of the Sun as being located at some position on the hypothetical celestial sphere. When the Sun rises—that is, when the rotation of Earth carries the Sun above the horizon—sunlight is scattered by the molecules of our atmosphere, filling our sky with light and hiding the stars above the horizon.

For thousands of years, astronomers have been aware that the Sun does more than just rise and set. It changes position gradually on the celestial sphere, moving each day about 1° to the east relative to the stars. Very reasonably, the ancients thought this meant the Sun was slowly moving around Earth, taking a period of time we call 1 year to make a full circle. Today, of course, we know it is Earth that is going around the Sun, but the effect is the same: the Sun's position in our sky changes day to day. We have a similar experience when we walk around a campfire at night; we see the flames appear in front of each person seated about the fire in turn.

The path the Sun appears to take around the celestial sphere each year is called the ecliptic (Figure 2.6). Because of its motion on the ecliptic, the Sun rises about 4 minutes later each day with respect to the stars. Earth must make just a bit more than one complete rotation (with respect to the stars) to bring the Sun up again.

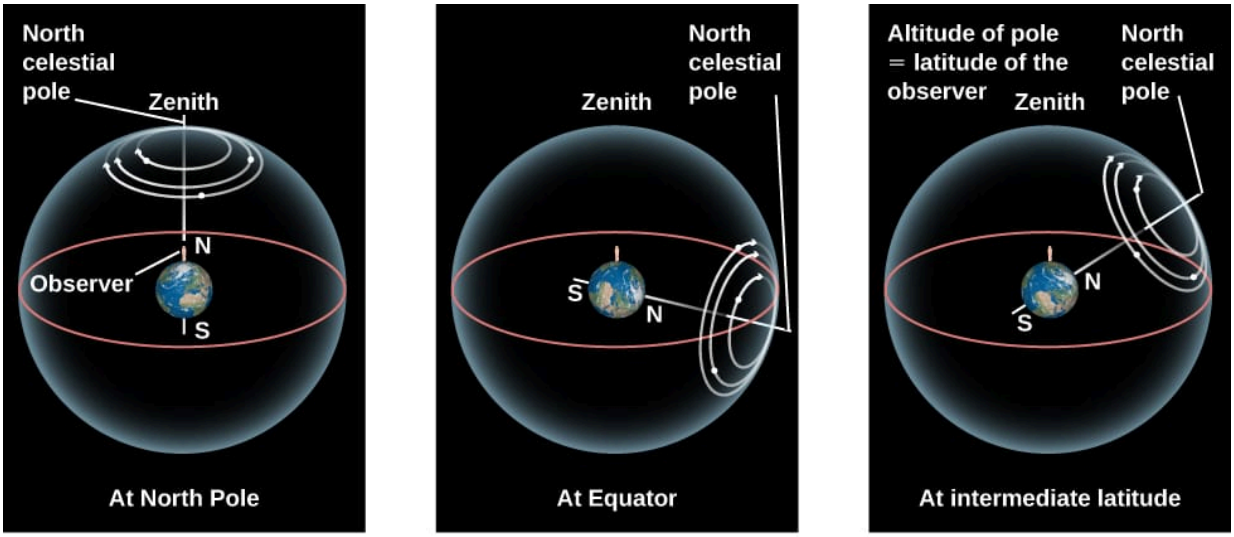


Figure 2.5 : Star Circles at Different Latitudes. The turning of the sky looks different depending on your latitude on Earth. The red circle in each case is your horizon. Your zenith is the point above your head. (a) At the North Pole, the stars circle the zenith and do not rise and set. (b) At the equator, the celestial poles are on the horizon, and the stars rise straight up and set straight down. (c) At intermediate latitudes, the north celestial pole is at some position between overhead and the horizon. Its angle above the horizon turns out to be equal to the observer's latitude. Stars rise and set at an angle to the horizon.

As the months go by and we look at the Sun from different places in our orbit, we see it projected against different places in our orbit, and thus against different stars in the background (Figure 2.6 and Table 2.1)—or we would, at least, if we could see the stars in the daytime. In practice, we must deduce which stars lie behind and beyond the Sun by observing the stars visible in the opposite direction at night. After a year, when Earth has completed one trip around the Sun, the Sun will appear to have completed one circuit of the sky along the ecliptic.

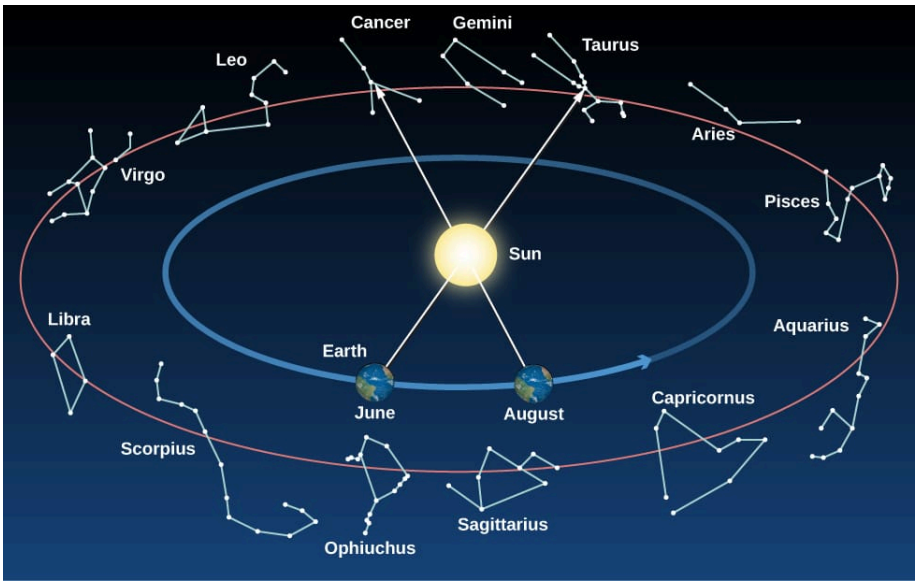


Figure 2.6 : Constellations on the Ecliptic. As Earth revolves around the Sun, we sit on "platform Earth" and see the Sun moving around the sky. The circle in the sky that the Sun appears to make around us in the course of a year is called the *ecliptic*. This circle (like all circles in the sky) goes through a set of constellations. The ancients thought these constellations, which the Sun (and the Moon and planets) visited, must be special and incorporated them into their system of astrology. Note that at any given time of the year, some of the constellations crossed by the ecliptic are visible in the night sky; others are in the day sky and are thus hidden by the brilliance of the Sun. (The term constellation will be defined more precisely in the next subsection.)

Table 2.1: Constellations on the Ecliptic

Constellation on the Ecliptic	Dates When the Sun Crosses It
Capricornus	January 21–February 16
Aquarius	February 16–March 11
Pisces	March 11–April 18
Aries	April 18–May 13
Taurus	May 13–June 22
Gemini	June 22–July 21
Cancer	July 21–August 10
Leo	August 10–September 16
Virgo	September 16–October 31
Libra	October 31–November 23
Scorpius	November 23–November 29
Ophiuchus	November 29–December 18
Sagittarius	December 18–January 21

The ecliptic does not lie along the celestial equator but is inclined to it at an angle of about 23.5°. In other words, the Sun's annual path in the sky is not linked with Earth's equator. This is because our planet's axis of rotation is tilted by about 23.5° from a vertical line sticking out of the plane of the ecliptic (Figure 2.7). Being tilted from "straight up" is not at all unusual among celestial bodies; Uranus and Pluto are actually tilted so much that they orbit the Sun "on their side."

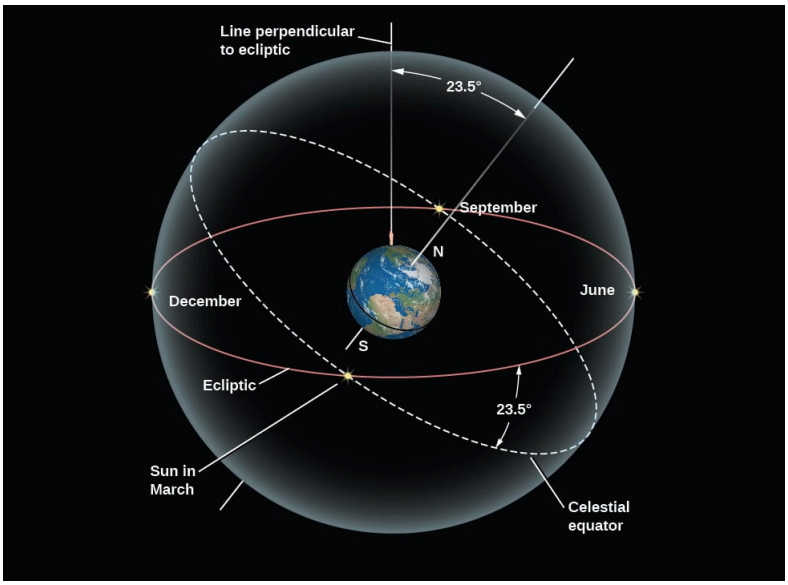


Figure 2.7 : The Celestial Tilt. The celestial equator is tilted by 23.5° to the ecliptic. As a result, North Americans and Europeans see the Sun north of the celestial equator and high in our sky in June, and south of the celestial equator and low in the sky in December.

The inclination of the ecliptic is the reason the Sun moves north and south in the sky as the seasons change. In Earth, Moon, and Sky, we discuss the progression of the seasons in more detail.

Fixed and Wandering Stars

The Sun is not the only object that moves among the fixed stars. The Moon and each of the planets that are visible to the unaided eye—Mercury, Venus, Mars, Jupiter, Saturn, and Uranus (although just barely)—also change their positions slowly from day to day. During a single day, the Moon and planets all rise and set as Earth turns, just as the Sun and stars do. But like the Sun, they have independent motions among the stars, superimposed on the daily rotation of the celestial sphere. Noticing these motions, the Greeks of 2000 years ago distinguished between what they called the *fixed stars*—those that maintain fixed patterns among themselves through many generations—and the *wandering stars*, or planets. The word “planet,” in fact, means “wanderer” in ancient Greek.

Today, we do not regard the Sun and Moon as planets, but the ancients applied the term to all seven of the moving objects in the sky. Much of ancient astronomy was devoted to observing and predicting the motions of these celestial wanderers. They even dedicated a unit of time, the week, to the seven objects that move on their own; that’s why there are 7 days in a week. The Moon, being Earth’s nearest celestial neighbor, has the fastest apparent motion; it completes a trip around the sky in about 1 month (or *moonth*). To do this, the Moon moves about 12°, or 24 times its own apparent width on the sky, each day.

Example 2.1: Angles in the Sky

A circle consists of 360 degrees (°). When we measure the angle in the sky that something moves, we can use this formula:

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

This is true whether the motion is measured in kilometers per hour or degrees per hour; we just need to use consistent units.

As an example, let’s say you notice the bright star Sirius due south from your observing location in the Northern Hemisphere. You note the time, and then later, you note the time that Sirius sets below the horizon. You find that Sirius has traveled an angular distance of about 75° in 5 h. About how many hours will it take for Sirius to return to its original location?

Solution

The speed of Sirius is $\frac{75^\circ}{5\,\mathrm{h}} = \frac{15^\circ}{1\,\mathrm{h}}$. If we want to know the time required for Sirius to return to its original location, we need to wait until it goes around a full circle, or 360°. Rearranging the formula for speed we were originally given, we find:

$$\text{time} = \frac{\text{distance}}{\text{speed}} = \frac{360^\circ}{15^\circ/\mathrm{h}} = 24\,\mathrm{h}$$

The actual time is a few minutes shorter than this, and we will explore why in a later chapter.

Exercise 2.2.1

The Moon moves in the sky relative to the background stars (in addition to moving with the stars as a result of Earth's rotation.) Go outside at night and note the position of the Moon relative to nearby stars. Repeat the observation a few hours later. How far has the Moon moved? (For reference, the diameter of the Moon is about 0.5°.) Based on your estimate of its motion, how long will it take for the Moon to return to the position relative to the stars in which you first observed it?

Answer available at the end of this section.

The individual paths of the Moon and planets in the sky all lie close to the ecliptic, although not exactly on it. This is because the paths of the planets about the Sun, and of the Moon about Earth, are all in nearly the same plane, as if they were circles on a huge sheet of paper. The planets, the Sun, and the Moon are thus always found in the sky within a narrow 18-degree-wide belt, centered on the ecliptic, called the zodiac (Figure 2.6). (The root of the term "zodiac" is the same as that of the word "zoo" and means a collection of animals; many of the patterns of stars within the zodiac belt reminded the ancients of animals, such as a fish or a goat.)

How the planets appear to move in the sky as the months pass is a combination of their actual motions plus the motion of Earth about the Sun; consequently, their paths are somewhat complex. As we will see, this complexity has fascinated and challenged astronomers for centuries.

Constellations

The backdrop for the motions of the "wanderers" in the sky is the canopy of stars. If there were no clouds in the sky and we were on a flat plain with nothing to obstruct our view, we could see about 3000 stars with the unaided eye. To find their way around such a multitude, the ancients found groupings of stars that made some familiar geometric pattern or (more rarely) resembled something they knew. Each civilization found its own patterns in the stars, much like a modern Rorschach test in which you are asked to discern patterns or pictures in a set of inkblots. The ancient Chinese, Egyptians, and Greeks, among others, found their own groupings—or constellations—of stars. These were helpful in navigating among the stars and in passing their star lore on to their children.

You may be familiar with some of the old star patterns we still use today, such as the Big Dipper, Little Dipper, and Orion the hunter, with his distinctive belt of three stars (Figure 2.8). However, many of the stars we see are not part of a distinctive star pattern at all, and a telescope reveals millions of stars too faint for the eye to see. Therefore, during the early decades of the 20th century, astronomers from many countries decided to establish a more formal system for organizing the sky.

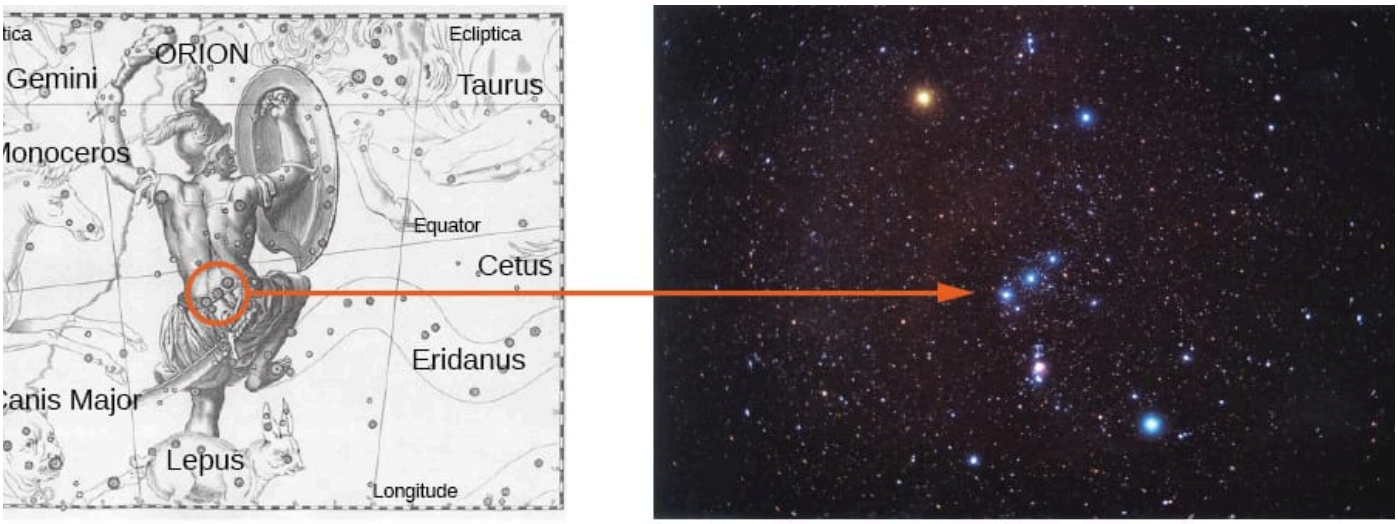


Figure 2.8 : Orion. (a) The winter constellation of Orion, the hunter, is surrounded by neighboring constellations, as illustrated in the seventeenth-century atlas by Hevelius. (b) A photograph shows the Orion region in the sky. Note the three blue stars that make up the belt of the hunter. The bright red star above the belt denotes his armpit and is called Betelgeuse (pronounced “Beetel-juice”). The bright blue star below the belt is his foot and is called Rigel. (credit a: modification of work by Johannes Hevelius; b: modification of work by Matthew Spinelli)

Today, we use the term *constellation* to mean one of 88 sectors into which we divide the sky, much as the United States is divided into 50 states. The modern boundaries between the constellations are imaginary lines in the sky running north–south and east–west, so that each point in the sky falls in a specific constellation, although, like the states, not all constellations are the same size. All the constellations are listed in Appendix L. Whenever possible, we have named each modern constellation after the Latin translations of one of the ancient Greek star patterns that lies within it. Thus, the modern constellation of Orion is a kind of box on the sky, which includes, among many other objects, the stars that made up the ancient picture of the hunter. Some people use the term *asterism* to denote an especially noticeable star pattern within a constellation (or sometimes spanning parts of several constellations). For example, the Big Dipper is an asterism within the constellation of Ursa Major, the Big Bear.

Students are sometimes puzzled because the constellations seldom resemble the people or animals for which they were named. In all likelihood, the Greeks themselves did not name groupings of stars because they looked like actual people or subjects (any more than the outline of Washington state resembles George Washington). Rather, they named sections of the sky in honor of the characters in their mythology and then fit the star configurations to the animals and people as best they could.

Link to Learning

This website about objects in the sky allows users to construct a detailed sky map showing the location and information about the Sun, Moon, planets, stars, constellations, and even satellites orbiting Earth. Begin by setting your observing location using the option in the menu in the upper right corner of the screen. An excellent website called Figures in the Sky shows the constellation figures imagined by cultures around the world.

Answers

Exercise 2.2.1

The speed of the moon is 0.5°/1 h. To move a full 360°, the moon needs 720 h: $\frac{0.5^\circ}{1\,\mathrm{h}} = \frac{360^\circ}{720\,\mathrm{h}}$. Dividing 720 h by the conversion factor of 24 h/day reveals the lunar cycle is about 30 days.

This page titled 2.2: The Sky Above is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 2.2: The Sky Above by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.3: Earth and Sky

Learning Objectives

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

- Describe how latitude and longitude are used to map Earth
- Explain how right ascension and declination are used to map the sky

In order to create an accurate map, a mapmaker needs a way to uniquely and simply identify the location of all the major features on the map, such as cities or natural landmarks. Similarly, astronomical mapmakers need a way to uniquely and simply identify the location of stars, galaxies, and other celestial objects. On Earth maps, we divide the surface of Earth into a grid, and each location on that grid can easily be found using its *latitude* and *longitude* coordinate. Astronomers have a similar system for objects on the sky. Learning about these can help us understand the apparent motion of objects in the sky from various places on Earth.

Locating Places on Earth

Let's begin by fixing our position on the surface of planet Earth. As we discussed in Observing the Sky: The Birth of Astronomy, Earth's axis of rotation defines the locations of its North and South Poles and of its equator, halfway between. Two other directions are also defined by Earth's motions: east is the direction toward which Earth rotates, and west is its opposite. At almost any point on Earth, the four directions—north, south, east, and west—are well defined, despite the fact that our planet is round rather than flat. The only exceptions are exactly at the North and South Poles, where the directions east and west are ambiguous (because points exactly at the poles do not turn).

We can use these ideas to define a system of coordinates attached to our planet. Such a system, like the layout of streets and avenues in Manhattan or Salt Lake City, helps us find where we are or want to go. Coordinates on a sphere, however, are a little more complicated than those on a flat surface. We must define circles on the sphere that play the same role as the rectangular grid that you see on city maps.

A great circle is any circle on the surface of a sphere whose center is at the center of the sphere. For example, Earth's equator is a great circle on Earth's surface, halfway between the North and South Poles. We can also imagine a series of great circles that pass through both the North and South Poles. Each of the circles is called a meridian; they are each perpendicular to the equator, crossing it at right angles.

Any point on the surface of Earth will have a meridian passing through it (Figure 4.2). The meridian specifies the east-west location, or longitude, of the place. By international agreement (and it took many meetings for the world's countries to agree), longitude is defined as the number of degrees of arc along the equator between your meridian and the one passing through Greenwich, England, which has been designated as the Prime Meridian. The longitude of the Prime Meridian is defined as 0°.

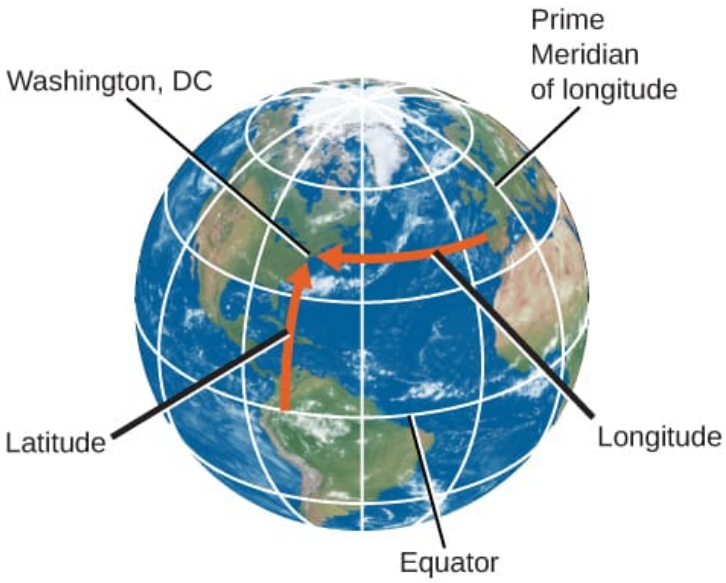


Figure 4.2 : Latitude and Longitude of Washington, DC. We use latitude and longitude to find cities like Washington, DC, on a globe. Latitude is the number of degrees north or south of the equator, and longitude is the number of degrees east or west of the Prime Meridian. Washington, DC's coordinates are 38° N and 77° W.

Why Greenwich, you might ask? Every country wanted 0° longitude to pass through its own capital. Greenwich, the site of the old Royal Observatory (Figure 4.3), was selected because it was between continental Europe and the United States, and because it was the site for much of the development of the method to measure longitude at sea. Longitudes are measured either to the east or to the west of the Greenwich meridian from 0° to 180°. As an example, the longitude of the clock-house benchmark of the U.S. Naval Observatory in Washington, DC, is 77.066° W.



Figure 4.3 : Royal Observatory in Greenwich, England. At the internationally agreed-upon zero point of longitude at the Royal Observatory Greenwich, tourists can stand and straddle the exact line where longitude "begins."(credit left: modification of work by "pdbreen"/Flickr; credit right: modification of work by Ben Sutherland)

Your latitude (or north-south location) is the number of degrees of arc you are away from the equator along your meridian. Latitudes are measured either north or south of the equator from 0° to 90°. (The latitude of the equator is 0°.) As an example, the latitude of the previously mentioned Naval Observatory benchmark is 38.921° N. The latitude of the South Pole is 90° S, and the latitude of the North Pole is 90° N.

Locating Places in the Sky

Positions in the sky are measured in a way that is very similar to the way we measure positions on the surface of Earth. Instead of latitude and longitude, however, astronomers use coordinates called declination and right ascension. To denote positions of objects in the sky, it is often convenient to make use of the fictitious celestial sphere. We saw in Observing the Sky: The Birth of Astronomy that the sky appears to rotate about points above the North and South Poles of Earth—points in the sky called the north celestial pole and the south celestial pole. Halfway between the celestial poles, and thus 90° from each pole, is the *celestial equator,* a great circle on the celestial sphere that is in the same plane as Earth's equator. We can use these markers in the sky to set up a system of celestial coordinates.

Declination on the celestial sphere is measured the same way that latitude is measured on the sphere of Earth: from the celestial equator toward the north (positive) or south (negative). So Polaris, the star near the north celestial pole, has a declination of almost +90°.

Right ascension (RA) is like longitude, except that instead of Greenwich, the arbitrarily chosen point where we start counting is the *vernal equinox*, a point in the sky where the *ecliptic* (the Sun's path) crosses the celestial equator. RA can be expressed either in units of angle (degrees) or in units of time. This is because the celestial sphere appears to turn around Earth once a day as our planet turns on its axis. Thus the 360° of RA that it takes to go once around the celestial sphere can just as well be set equal to 24 hours. Then each 15° of arc is equal to 1 hour of time. For example, the approximate celestial coordinates of the bright star Capella are RA $5h = 75°$ and declination $+50°$.

One way to visualize these circles in the sky is to imagine Earth as a transparent sphere with the terrestrial coordinates (latitude and longitude) painted on it with dark paint. Imagine the celestial sphere around us as a giant ball, painted white on the inside. Then imagine yourself at the center of Earth, with a bright light bulb in the middle, looking out through its transparent surface to the sky. The terrestrial poles, equator, and meridians will be projected as dark shadows on the celestial sphere, giving us the system of coordinates in the sky.

The Turning Earth

Why do many stars rise and set each night? Why, in other words, does the night sky seem to turn? We have seen that the apparent rotation of the celestial sphere could be accounted for either by a daily rotation of the sky around a stationary Earth or by the rotation of Earth itself. Since the seventeenth century, it has been generally accepted that it is Earth that turns, but not until the nineteenth century did the French physicist Jean Foucault provide an unambiguous demonstration of this rotation. In 1851, he suspended a 60-meter pendulum with a mass of about 25 kilograms from the dome of the Pantheon in Paris and started the pendulum swinging evenly. If Earth had not been turning, there would have been no alteration of the pendulum's plane of oscillation, and so it would have continued tracing the same path. Yet after a few minutes Foucault could see that the pendulum's plane of motion was turning. Foucault explained that it was not the pendulum that was shifting, but rather Earth that was turning beneath it (Figure 4.4). You can now find such pendulums in many science centers and planetariums around the world.

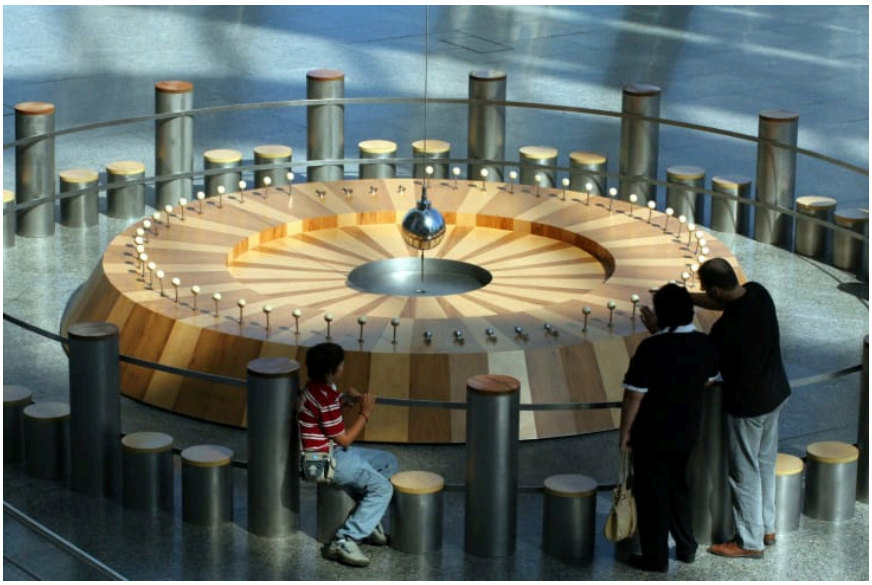


Figure 4.4 : Foucault's Pendulum. As Earth turns, the plane of oscillation of the Foucault pendulum shifts gradually so that over the course of 12 hours, all the targets in the circle at the edge of the wooden platform are knocked over in sequence. (credit: Manuel M. Vicente)

Can you think of other pieces of evidence that indicate that it is Earth and not the sky that is turning? (See Collaborative Group Activity A at the end of this chapter.)

This page titled 2.3: Earth and Sky is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 4.2: Earth and Sky by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.4: The Seasons

Learning Objectives

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

- Describe how the tilt of Earth's axis causes the seasons
- Explain how seasonal differences on Earth vary with latitude

One of the fundamental facts of life at Earth's midlatitudes, where most of this book's readers live, is that there are significant variations in the heat we receive from the Sun during the course of the year. We thus divide the year into *seasons*, each with its different amount of sunlight. The difference between seasons gets more pronounced the farther north or south from the equator we travel, and the seasons in the Southern Hemisphere are the opposite of what we find on the northern half of Earth. With these observed facts in mind, let us ask what causes the seasons.

Many people have believed that the seasons were the result of the changing distance between Earth and the Sun. This sounds reasonable at first: it should be colder when Earth is farther from the Sun. But the facts don't bear out this hypothesis. Although Earth's orbit around the Sun is an ellipse, its distance from the Sun varies by only about 3%. That's not enough to cause significant variations in the Sun's heating. To make matters worse for people in North America who hold this hypothesis, Earth is actually closest to the Sun in January, when the Northern Hemisphere is in the middle of winter. And if distance were the governing factor, why would the two hemispheres have opposite seasons? As we shall show, the seasons are actually caused by the 23.5° tilt of Earth's axis.

The Seasons and Sunshine

Figure 4.5 shows Earth's annual path around the Sun, with Earth's axis tilted by 23.5°. Note that our axis continues to point the same direction in the sky throughout the year. As Earth travels around the Sun, in June the Northern Hemisphere "leans into" the Sun and is more directly illuminated. In December, the situation is reversed: the Southern Hemisphere leans into the Sun, and the Northern Hemisphere leans away. In September and March, Earth leans "sideways"—neither into the Sun nor away from it—so the two hemispheres are equally favored with sunshine.

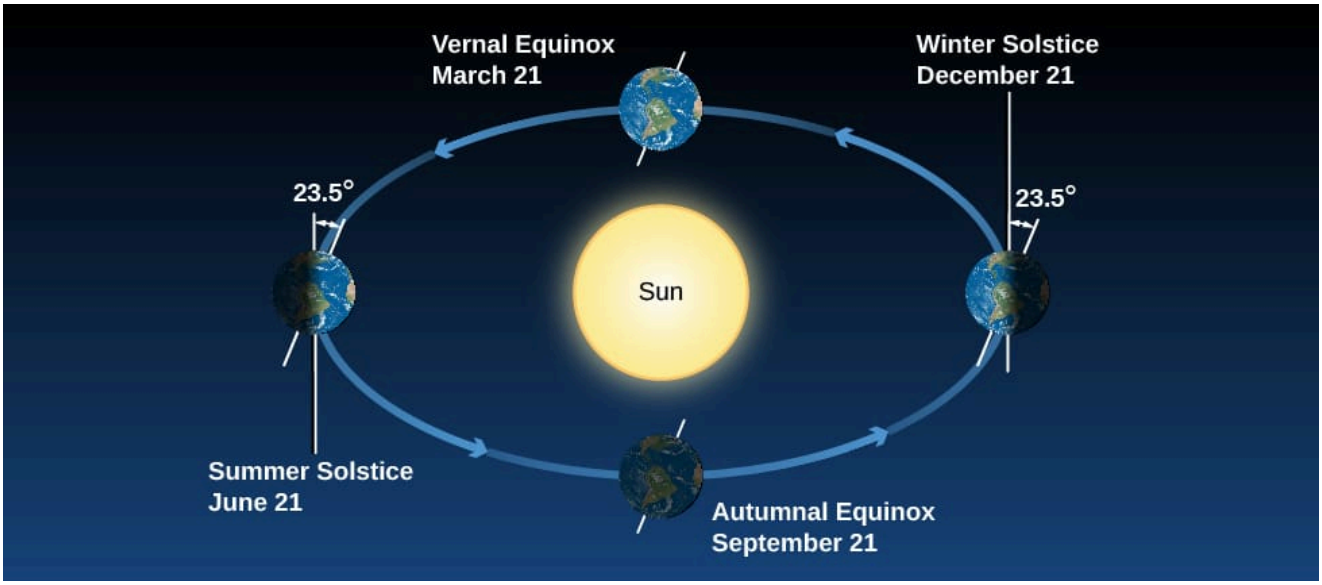


Figure 4.5 : Seasons. We see Earth at different seasons as it circles the Sun. In June, the Northern Hemisphere "leans into" the Sun, and those in the North experience summer and have longer days. In December, during winter in the Northern Hemisphere, the Southern Hemisphere "leans into" the Sun and is illuminated more directly. In spring and autumn, the two hemispheres receive more equal shares of sunlight.[1]

How does the Sun's favoring one hemisphere translate into making it warmer for us down on the surface of Earth? There are two effects we need to consider. When we lean into the Sun, sunlight hits us at a more direct angle and is more effective at heating Earth's surface (Figure 4.6). You can get a similar effect by shining a flashlight onto a wall. If you shine the flashlight straight on, you get an intense spot of light on the wall. But if you hold the flashlight at an angle (if the wall "leans out" of the beam), then the spot of light is more spread out. Like the straight-on light, the sunlight in June is more direct and intense in the Northern Hemisphere, and hence more effective at heating.

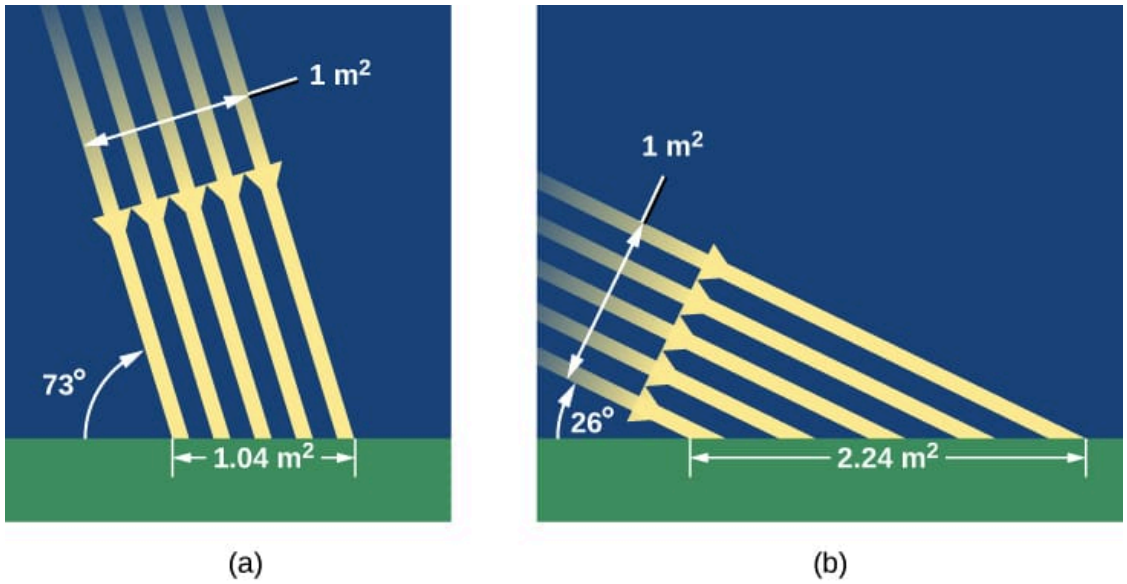


Figure 4.6 : The Sun's Rays in Summer and Winter. (a) In summer, the Sun appears high in the sky and its rays hit Earth more directly, spreading out less. (b) In winter, the Sun is low in the sky and its rays spread out over a much wider area, becoming less effective at heating the ground.

The second effect has to do with the length of time the Sun spends above the horizon (Figure 4.7). Even if you've never thought about astronomy before, we're sure you have observed that the hours of daylight increase in summer and decrease in winter. Let's see why this happens.

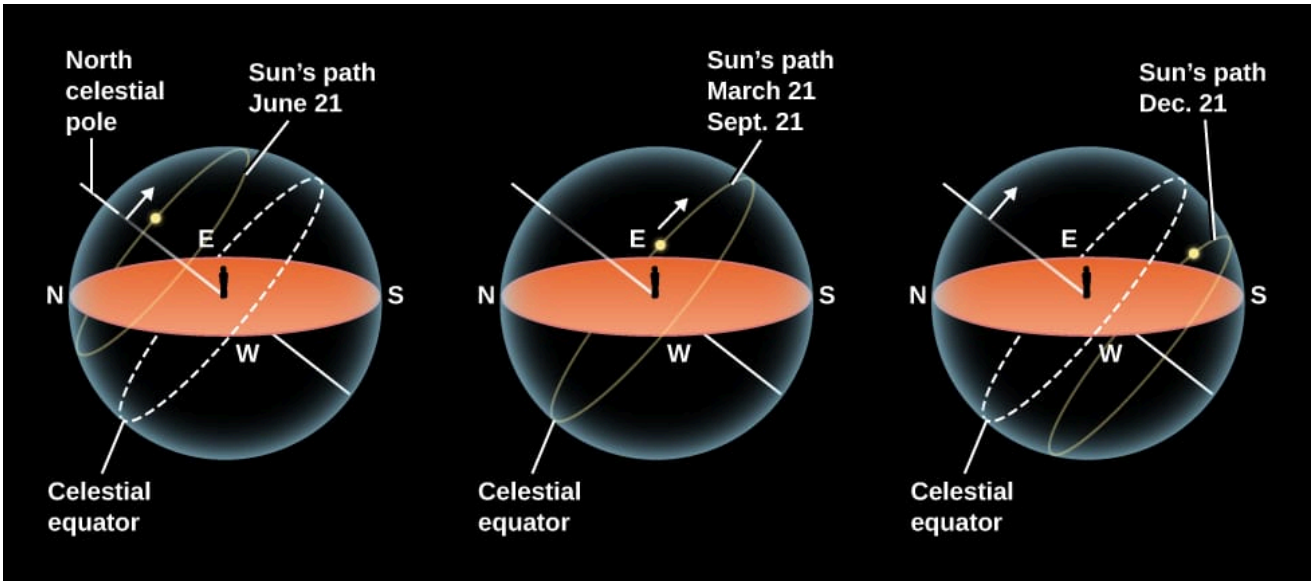


Figure 4.7 : The Sun's Path in the Sky for Different Seasons. On June 21, the Sun rises north of east and sets north of west. For observers in the Northern Hemisphere of Earth, the Sun spends about 15 hours above the horizon in the United States, meaning more hours of daylight. On December 21, the Sun rises south of east and sets south of west. It spends 9 hours above the horizon in the United States, which means fewer hours of daylight and more hours of night in northern lands (and a strong need for people to hold celebrations to cheer themselves up). On March 21 and September 21, the Sun spends equal amounts of time above and below the horizon in both hemispheres.

As we saw in Observing the Sky: The Birth of Astronomy, an equivalent way to look at our path around the Sun each year is to pretend that the Sun moves around Earth (on a circle called the ecliptic). Because Earth's axis is tilted, the ecliptic is tilted by about 23.5° relative to the celestial equator (review Figure 2.7). As a result, where we see the Sun in the sky changes as the year wears on.

In June, the Sun is north of the celestial equator and spends more time with those who live in the Northern Hemisphere. It rises high in the sky and is above the horizon in the United States for as long as 15 hours. Thus, the Sun not only heats us with more direct rays, but it also has more time to do it each day. (Notice in Figure 4.7 that the Northern Hemisphere's gain is the Southern Hemisphere's loss. There the June Sun is low in the sky, meaning fewer daylight hours. In Chile, for example, June is a colder, darker time of year.) In December, when the Sun is south of the celestial equator, the situation is reversed.

Link to Learning

The Motions of the Sun Simulator from Columbia University's Center for Teaching and Learning provides a demonstration of the Sun's apparent motion in the sky as the Earth rotates each day, and its changing altitude with latitude and time of year.

Let's look at what the Sun's illumination on Earth looks like at some specific dates of the year, when these effects are at their maximum. On or about June 21 (the date we who live in the Northern Hemisphere call the *summer solstice* or sometimes the first day of summer), the Sun shines down most directly upon the Northern Hemisphere of Earth. It appears about 23° north of the equator, and thus, on that date, it passes through the zenith of places on Earth that are at 23° N latitude. The situation is shown in detail in Figure 4.8. To a person at 23° N (near Hawaii, for example), the Sun is directly overhead at noon. This latitude, where the Sun can appear at the zenith at noon on the first day of summer, is called the *Tropic of Cancer*.

We also see in Figure 4.8 that the Sun's rays shine down all around the North Pole at the solstice. As Earth turns on its axis, the North Pole is continuously illuminated by the Sun; all places within 23° of the pole have sunshine for 24 hours. The Sun is as far north on this date as it can get; thus, 90° – 23° (or 67° N) is the southernmost latitude where the Sun can be seen for a full 24-hour period (sometimes called the "land of the midnight Sun"). That circle of latitude is called the *Arctic Circle*.

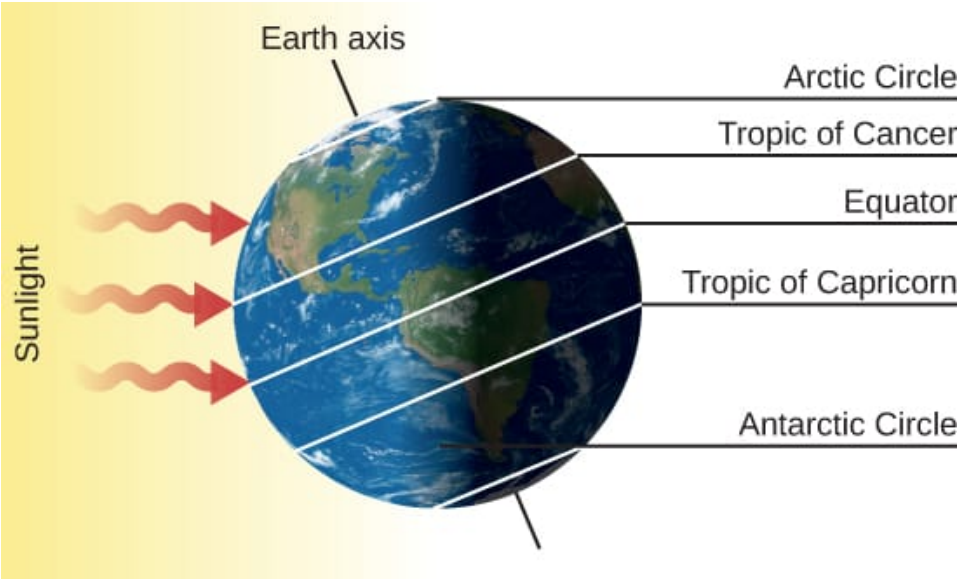


Figure 4.8 : Earth on June 21. This is the date of the summer solstice in the Northern Hemisphere. Note that as Earth turns on its axis (the line connecting the North and South Poles), the North Pole is in constant sunlight while the South Pole is veiled in 24 hours of darkness. The Sun is at the zenith for observers on the Tropic of Cancer.

Many early cultures scheduled special events around the summer solstice to celebrate the longest days and thank their gods for making the weather warm. This required people to keep track of the lengths of the days and the northward trek of the Sun in order to know the right day for the "party." (You can do the same thing by watching for several weeks, from the same observation point, where the Sun rises or sets relative to a fixed landmark. In spring, the Sun will rise farther and farther north of east, and set farther and farther north of west, reaching the maximum around the summer solstice.)

Now look at the South Pole in Figure 4.8. On June 21, all places within 23° of the South Pole—that is, south of what we call the *Antarctic Circle*—do not see the Sun at all for 24 hours.

The situation is reversed 6 months later, about December 21 (the date of the *winter solstice*, or the first day of winter in the Northern Hemisphere), as shown in Figure 4.9. Now it is the Arctic Circle that has the 24-hour night and the Antarctic Circle that has the midnight Sun. At latitude 23° S, called the *Tropic of Capricorn*, the Sun passes through the zenith at noon. Days are longer in the Southern Hemisphere and shorter in the north. In the United States and Southern Europe, there may be only 9 or 10 hours of sunshine during the day. It is winter in the Northern Hemisphere and summer in the Southern Hemisphere.

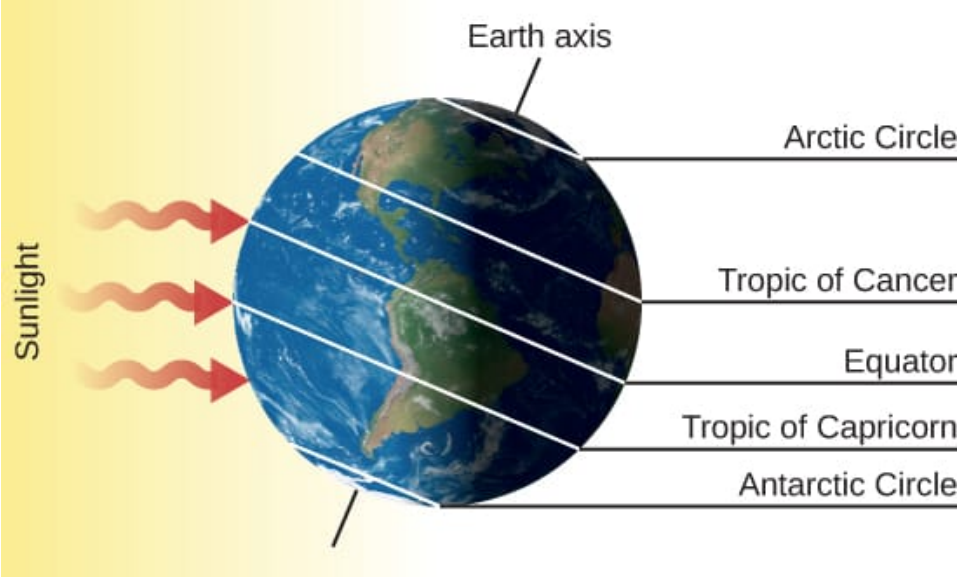


Figure 4.9 : Earth on December 21. This is the date of the winter solstice in the Northern Hemisphere. Now the North Pole is in darkness for 24 hours and the South Pole is illuminated. The Sun is at the zenith for observers on the Tropic of Capricorn and thus is low in the sky for the residents of the Northern Hemisphere.

Example 4.1: Seasonal Variations

As you can see in Figure 4.8, the Tropic of Cancer is the latitude for which the Sun is directly overhead on the summer solstice. At this time, the Sun is at a declination of 23° N of the celestial equator, and the corresponding latitude on Earth is 23° N of the equator. If Earth were tilted a bit less, then the Tropic of Cancer would be at a lower latitude, closer to the equator.

The Arctic Circle marks the southernmost latitude for which the day length is 24 hours on the day of the summer solstice. This is located at 90° – 23° = 67° N of Earth's equator. If Earth were tilted a bit less, then the Arctic Circle would move farther North. In the limit at which Earth is not tilted at all (its axis is perpendicular to the ecliptic), the Tropic of Cancer would be right on Earth's equator, and the Arctic Circle would simply be the North Pole. Suppose the tilt of Earth's axis were tilted only 5°. What would be the effect on the seasons and the locations of the Tropic of Cancer and Arctic Circle?

Solution

If Earth were tilted less, the seasons would be less extreme. The variation in day length and direct sunlight would be very small over the course of a year, and the Sun's daily path in the sky would not vary much. If Earth were tilted by 5°, the Sun's position on the day of the summer solstice would be 5° N of the celestial equator, so the Tropic of Cancer would be at the corresponding latitude on Earth of 5° N of the Equator. The Arctic Circle would be located at 90° – 5° = 85° N of the equator.

Exercise 2.4.1

Suppose the tilt of Earth's axis were 16°. What, then, would be the difference in latitude between the Arctic Circle and the Tropic of Cancer? What would be the effect on the seasons compared with that produced by the actual tilt of 23°?

Answer available at the end of this section.

Link to Learning

The Columbia University's Center for Teaching and Learning's Motions of the Sun Simulator allows you to see how the Sun's changing altitude over the course of the year changes the intensity of lighting experienced on Earth. Select the "Step by day" option before running the animation to illustrate this effect.

Many cultures that developed some distance north of the equator have a celebration around December 21 to help people deal with the depressing lack of sunlight and the often dangerously cold temperatures. Originally, this was often a time for huddling with family and friends, for sharing the reserves of food and drink, and for rituals asking the gods to return the light and heat and turn the cycle of the seasons around. Many cultures constructed elaborate devices for anticipating when the shortest day of the year was coming. Stonehenge in England, built long before the invention of writing, is probably one such device. In our own time, we continue the winter solstice tradition with various holiday celebrations around that December date.

Halfway between the solstices, on about March 21 and September 21, the Sun is on the celestial equator. From Earth, it appears above our planet's equator and favors neither hemisphere. Every place on Earth then receives roughly 12 hours of sunshine and 12 hours of night. The points where the Sun crosses the celestial equator are called the *vernal* (spring) and *autumnal* (fall) *equinoxes*.

The Seasons at Different Latitudes

The seasonal effects are different at different latitudes on Earth. Near the equator, for instance, all seasons are much the same. Every day of the year, the Sun is up half the time, so there are approximately 12 hours of sunshine and 12 hours of night. Local residents define the seasons by the amount of rain (wet season and dry season) rather than by the amount of sunlight. As we travel north or south, the seasons become more pronounced, until we reach extreme cases in the Arctic and Antarctic.

At the North Pole, all celestial objects that are north of the celestial equator are always above the horizon and, as Earth turns, circle around parallel to it. The Sun is north of the celestial equator from about March 21 to September 21, so at the North Pole, the Sun rises when it reaches the vernal equinox and sets when it reaches the autumnal equinox. Each year there are 6 months of sunshine at each pole, followed by 6 months of darkness.

Example 4.2: The Position of the Sun in the Sky

The Sun's coordinates on the celestial sphere range from a declination of 23° N of the celestial equator (or +23°) to a declination 23° S of the celestial equator (or –23°). So, the Sun's altitude at noon, when it crosses the meridian, varies by a total of 46°. What is the altitude of the Sun at noon on March 21, as seen from a place on Earth's equator? What is its altitude on June 21, as seen from a place on Earth's equator?

Solution

On Earth's equator, the celestial equator passes through the zenith. On March 21, the Sun is crossing the celestial equator, so it should be found at the zenith (90°) at noon. On June 21, the Sun is 23° N of the celestial equator, so it will be 23° away from the zenith at noon. The altitude above the horizon will be 23° less than the altitude of the zenith (90°), so it is 90° – 23° = 67° above the horizon.

Exercise 2.4.1

What is the altitude of the Sun at noon on December 21, as seen from a place on the Tropic of Cancer?

Answer available at the end of this section.

Clarifications about the Real World

In our discussions so far, we have been describing the rising and setting of the Sun and stars as they would appear if Earth had little or no atmosphere. In reality, however, the atmosphere has the curious effect of allowing us to see a little way "over the horizon." This effect is a result of *refraction*, the bending of light passing through air or water, something we will discuss in Astronomical Instruments. Because of this atmospheric refraction (and the fact that the Sun is not a point of light but a disk), the Sun appears to rise earlier and to set later than it would if no atmosphere were present.

In addition, the atmosphere scatters light and provides some twilight illumination even when the Sun is below the horizon. Astronomers define morning twilight as beginning when the Sun is 18° below the horizon, and evening twilight extends until the Sun sinks more than 18° below the horizon.

These atmospheric effects require small corrections in many of our statements about the seasons. At the equinoxes, for example, the Sun appears to be above the horizon for a few minutes longer than 12 hours, and below the horizon for fewer than 12 hours. These effects are most dramatic at Earth's poles, where the Sun actually can be seen more than a week before it reaches the celestial equator.

You probably know that the summer solstice (June 21) is not the warmest day of the year, even if it is the longest. The hottest months in the Northern Hemisphere are July and August. This is because our weather involves the air and water covering Earth's surface, and these large reservoirs do not heat up instantaneously. You have probably observed this effect for yourself; for example, a pond does not get warm the moment the Sun rises but is warmest late in the afternoon, after it has had time to absorb the Sun's heat. In the same way, Earth gets warmer after it has had a chance to absorb the extra sunlight that is the Sun's summer gift to us. And the coldest times of winter are a month or more after the winter solstice.

Footnotes

- 1Note that the dates indicated for the solstices and equinoxes are approximate; depending on the year, they may occur a day or two earlier or later.

Answers

Exercise 2.4.1

The Tropic of Cancer is at a latitude equal to Earth's tilt, so in this case, it would be at 16° N latitude. The Arctic Circle is at a latitude equal to 90° minus Earth's tilt, or 90° – 16° = 74°. The difference between these two latitudes is 74° – 16° = 58°. Since the tilt of Earth is less, there would be less variation in the tilt of Earth and less variation in the Sun's paths throughout the year, so there would be milder seasonal changes.

Exercise 2.4.1

On the day of the winter solstice, the Sun is located about 23° S of the celestial equator. From the Tropic of Cancer, a latitude of 23° N, the zenith would be a declination of 23° N. The difference in declination between zenith and the position of the Sun is 46°, so the Sun would be 46° away from the zenith. That means it would be at an altitude of 90° – 46° = 44°.

This page titled 2.4: The Seasons is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 4.3: The Seasons by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.5: Phases and Motions of the Moon

Learning Objectives

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

- Explain the cause of the lunar phases
- Understand how the Moon rotates and revolves around Earth

After the Sun, the Moon is the brightest and most obvious object in the sky. Unlike the Sun, it does not shine under its own power, but merely glows with reflected sunlight. If you were to follow its progress in the sky for a month, you would observe a cycle of phases (different appearances), with the Moon starting dark and getting more and more illuminated by sunlight over the course of about two weeks. After the Moon's disk becomes fully bright, it begins to fade, returning to dark about two weeks later.

These changes fascinated and mystified many early cultures, which came up with marvelous stories and legends to explain the cycle of the Moon. Even in the modern world, many people don't understand what causes the phases, thinking that they are somehow related to the shadow of Earth. Let us see how the phases can be explained by the motion of the Moon relative to the bright light source in the solar system, the Sun.

Lunar Phases

Although we know that the Sun moves 1/12 of its path around the sky each month, for purposes of explaining the phases, we can assume that the Sun's light comes from roughly the same direction during the course of a four-week lunar cycle. The Moon, on the other hand, moves completely around Earth in that time. As we watch the Moon from our vantage point on Earth, how much of its face we see illuminated by sunlight depends on the angle the Sun makes with the Moon.

Here is a simple experiment to show you what we mean: stand about 6 feet in front of a bright electric light in a completely dark room (or outdoors at night) and hold in your hand a small round object such as a tennis ball or an orange. Your head can then represent Earth, the light represents the Sun, and the ball the Moon. Move the ball around your head (making sure you don't cause an eclipse by blocking the light with your head). You will see phases just like those of the Moon on the ball. (Another good way to get acquainted with the phases and motions of the Moon is to follow our satellite in the sky for a month or two, recording its shape, its direction from the Sun, and when it rises and sets.)

Let's examine the Moon's cycle of phases using Figure 4.14, which depicts the Moon's behavior for the entire month. The trick to this figure is that you must imagine yourself standing on Earth, facing the Moon in each of its phases. So, for the position labeled "New," you are on the right side of Earth and it's the middle of the day; for the position "Full," you are on the left side of Earth in the middle of the night. Note that in every position on Figure 4.14, the Moon is half illuminated and half dark (as a ball in sunlight should be). The difference at each position has to do with what part of the Moon faces Earth.

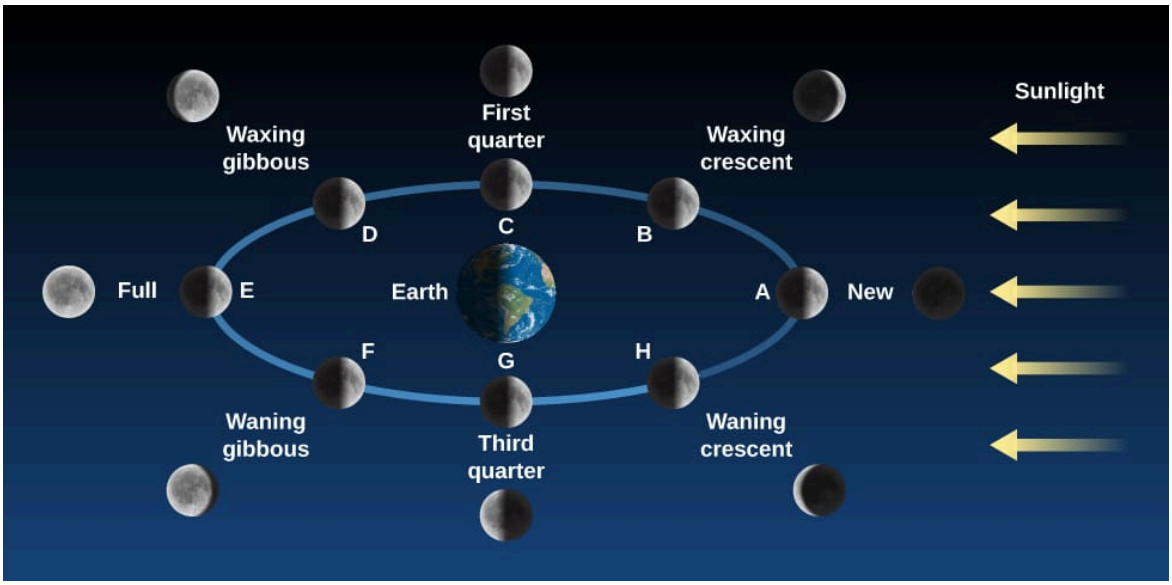


Figure 4.14 : Phases of the Moon. The appearance of the Moon changes over the course of a complete monthly cycle. The pictures of the Moon on the white circle show the perspective from space, with the Sun off to the right in a fixed position. The outer images show how the Moon appears to you in the sky from each point in the orbit. Imagine yourself standing on Earth, facing the Moon at each stage. In the position "New," for example, you are facing the Moon from the right side of Earth in the middle of the day. (Note that the distance of the Moon from Earth is not to scale in this diagram: the Moon is roughly 30 Earth-diameters away from us.) (credit: modification of work by NASA)

The Moon is said to be *new* when it is in the same general direction in the sky as the Sun (position A). Here, its illuminated (bright) side is turned away from us and its dark side is turned toward us. You might say that the Sun is shining on the "wrong " side of the Moon from our perspective. In this phase the Moon is invisible to us; its dark, rocky surface does not give off any light of its own. Because the new moon is in the same part of the sky as the Sun, it rises at sunrise and sets at sunset.

But the Moon does not remain in this phase long because it moves eastward each day in its monthly path around us. Since it takes about 30 days to orbit Earth and there are 360° in a circle, the Moon will move about 12° in the sky each day (or about 24 times its own diameter). A day or two after the new phase, the thin *crescent* first appears, as we begin to see a small part of the Moon's illuminated hemisphere. It has moved into a position where it now reflects a little sunlight toward us along one side. The bright crescent increases in size on successive days as the Moon moves farther and farther around the sky away from the direction of the Sun (position B). Because the Moon is moving eastward away from the Sun, it rises later and later each day (like a student during summer vacation).

After about one week, the Moon is one-quarter of the way around its orbit (position C) and so we say it is at the *first quarter* phase. Half of the Moon's illuminated side is visible to Earth observers. Because of its eastward motion, the Moon now lags about one-quarter of the day behind the Sun, rising around noon and setting around midnight.

During the week after the first quarter phase, we see more and more of the Moon's illuminated hemisphere (position D), a phase that is called *waxing* (or growing) gibbous (from the Latin *gibbus*, meaning hump). Eventually, the Moon arrives at position E in our figure, where it and the Sun are opposite each other in the sky. The side of the Moon turned toward the Sun is also turned toward Earth, and we have the *full* phase.

When the Moon is full, it is opposite the Sun in the sky. The Moon does the opposite of what the Sun does, rising at sunset and setting at sunrise. Note what that means in practice: the completely illuminated (and thus very noticeable) Moon rises just as it gets dark, remains in the sky all night long, and sets as the Sun's first rays are seen at dawn. Its illumination throughout the night helps lovers on a romantic stroll and students finding their way back to their dorms after a long night in the library or an off-campus party.

And when is the full moon highest in the sky and most noticeable? At midnight, a time made famous in generations of horror novels and films. (Note how the behavior of a vampire like Dracula parallels the behavior of the full Moon: Dracula rises at sunset, does his worst mischief at midnight, and must be back down in his coffin by sunrise. The old legends were a way of personifying the behavior of the Moon, which was a much more dramatic part of people's lives in the days before electric lights and television.)

Folklore has it that more crazy behavior is seen during the time of the full moon (the Moon even gives a name to crazy

behavior—"lunacy"). But, in fact, statistical tests of this "hypothesis" involving thousands of records from hospital emergency rooms and police files do not reveal any correlation of human behavior with the phases of the Moon. For example, homicides occur at the same rate during the new moon or the crescent moon as during the full moon. Most investigators believe that the real story is not that more crazy behavior happens on nights with a full moon, but rather that we are more likely to notice or remember such behavior with the aid of a bright celestial light that is up all night long.

During the two weeks following the full moon, the Moon goes through the same phases again in reverse order (points F, G, and H in Figure 4.14), returning to new phase after about 29.5 days. About a week after the full moon, for example, the Moon is at *third quarter*, meaning that it is three-quarters of the way around (not that it is three-quarters illuminated—in fact, half of the visible side of the Moon is again dark). At this phase, the Moon is now rising around midnight and setting around noon.

Note that there is one thing quite misleading about Figure 4.14. If you look at the Moon in position E, although it is full in theory, it appears as if its illumination would in fact be blocked by a big fat Earth, and hence we would not see anything on the Moon except Earth's shadow. In reality, the Moon is nowhere near as close to Earth (nor is its path so identical with the Sun's in the sky) as this diagram (and the diagrams in most textbooks) might lead you to believe.

The Moon is actually 30 *Earth-diameters* away from us; Science and the Universe: A Brief Tour contains a diagram that shows the two objects to scale. And, since the Moon's orbit is tilted relative to the path of the Sun in the sky, Earth's shadow misses the Moon most months. That's why we regularly get treated to a full moon. The times when Earth's shadow does fall on the Moon are called lunar eclipses and are discussed in Eclipses of the Sun and Moon.

Making Connections: Astronomy and the Days of the Week

The week seems independent of celestial motions, although its length may have been based on the time between quarter phases of the Moon. In Western culture, the seven days of the week are named after the seven "wanderers" that the ancients saw in the sky: the Sun, the Moon, and the five planets visible to the unaided eye (Mercury, Venus, Mars, Jupiter, and Saturn).

In English, we can easily recognize the names Sun-day (Sunday), Moon-day (Monday), and Saturn-day (Saturday), but the other days are named after the Norse equivalents of the Roman gods that gave their names to the planets. In languages more directly related to Latin, the correspondences are clearer. Wednesday, Mercury's day, for example, is *mercoledi* in Italian, *mercredi* in French, and *miércoles* in Spanish. Mars gives its name to Tuesday (*martes* in Spanish), Jupiter or Jove to Thursday (*giovedi* in Italian), and Venus to Friday (*vendredi* in French).

There is no reason that the week has to have seven days rather than five or eight. It is interesting to speculate that if we had lived in a planetary system where more planets were visible without a telescope, the Beatles could have been right and we might well have had "Eight Days a Week."

Link to Learning

This simulation of the Moon orbiting the Earth shows both a "top-down" view and the view from Earth's surface.

The Moon's Revolution and Rotation

The Moon's sidereal period—that is, the period of its revolution about Earth measured with respect to the stars—is a little over 27 days: the sidereal month is 27.3217 days to be exact. The time interval in which the phases repeat—say, from full to full—is the solar month, 29.5306 days. The difference results from Earth's motion around the Sun. The Moon must make more than a complete turn around the moving Earth to get back to the same phase with respect to the Sun. As we saw, the Moon changes its position on the celestial sphere rather rapidly: even during a single evening, the Moon creeps visibly eastward among the stars, traveling its own width in a little less than 1 hour. The delay in moonrise from one day to the next caused by this eastward motion averages about 50 minutes.

The Moon *rotates* on its axis in exactly the same time that it takes to *revolve* about Earth. As a consequence, the Moon always keeps the same face turned toward Earth (Figure 4.15). You can simulate this yourself by "orbiting" your roommate or another volunteer. Start by facing your roommate. If you make one rotation (spin) with your shoulders in the exact same time that you revolve around him or her, you will continue to face your roommate during the whole "orbit." As we will see in coming chapters,

our Moon is not the only world that exhibits this behavior, which scientists call synchronous rotation.

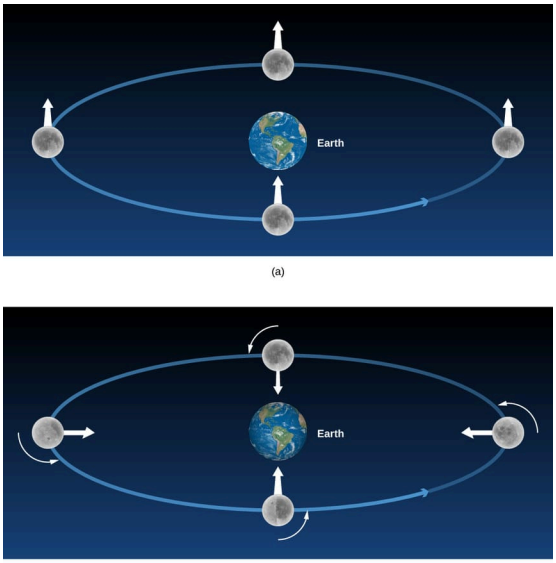


Figure 4.15 : The Moon without and with Rotation. In this figure, we stuck a white arrow into a fixed point on the Moon to keep track of its sides. (a) If the Moon did not rotate as it orbited Earth, it would present all of its sides to our view; hence the white arrow would point directly toward Earth only in the bottom position on the diagram. (b) Actually, the Moon rotates in the same period that it revolves, so we always see the same side (the white arrow keeps pointing to Earth).

The differences in the Moon's appearance from one night to the next are due to changing illumination by the Sun, not to its own rotation. You sometimes hear the back side of the Moon (the side we never see) called the "dark side." This is a misunderstanding of the real situation: which side is light and which is dark changes as the Moon moves around Earth. The back side is dark no more frequently than the front side. Since the Moon rotates, the Sun rises and sets on all sides of the Moon. With apologies to Pink Floyd, there is simply no regular "Dark Side of the Moon."

This page titled 2.5: Phases and Motions of the Moon is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 4.6: Phases and Motions of the Moon by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6: Ocean Tides and the Moon

Learning Objectives

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

- Describe what causes tides on Earth
- Explain why the amplitude of tides changes during the course of a month

Anyone living near the sea is familiar with the twice-daily rising and falling of the tides. Early in history, it was clear that tides must be related to the Moon because the daily delay in high tide is the same as the daily delay in the Moon's rising. A satisfactory explanation of the tides, however, awaited the theory of gravity, supplied by Newton.

The Pull of the Moon on Earth

The gravitational forces exerted by the Moon at several points on Earth are illustrated in Figure 4.16. These forces differ slightly from one another because Earth is not a point, but has a certain size: all parts are not equally distant from the Moon, nor are they all in exactly the same direction from the Moon. Moreover, Earth is not perfectly rigid. As a result, the differences among the forces of the Moon's attraction on different parts of Earth (called *differential forces*) cause Earth to distort slightly. The side of Earth nearest the Moon is attracted toward the Moon more strongly than is the center of Earth, which in turn is attracted more strongly than is the side opposite the Moon. Thus, the differential forces tend to stretch Earth slightly into an *oblate spheroid* (which is something like an American football shape), with its long diameter pointed toward the Moon.

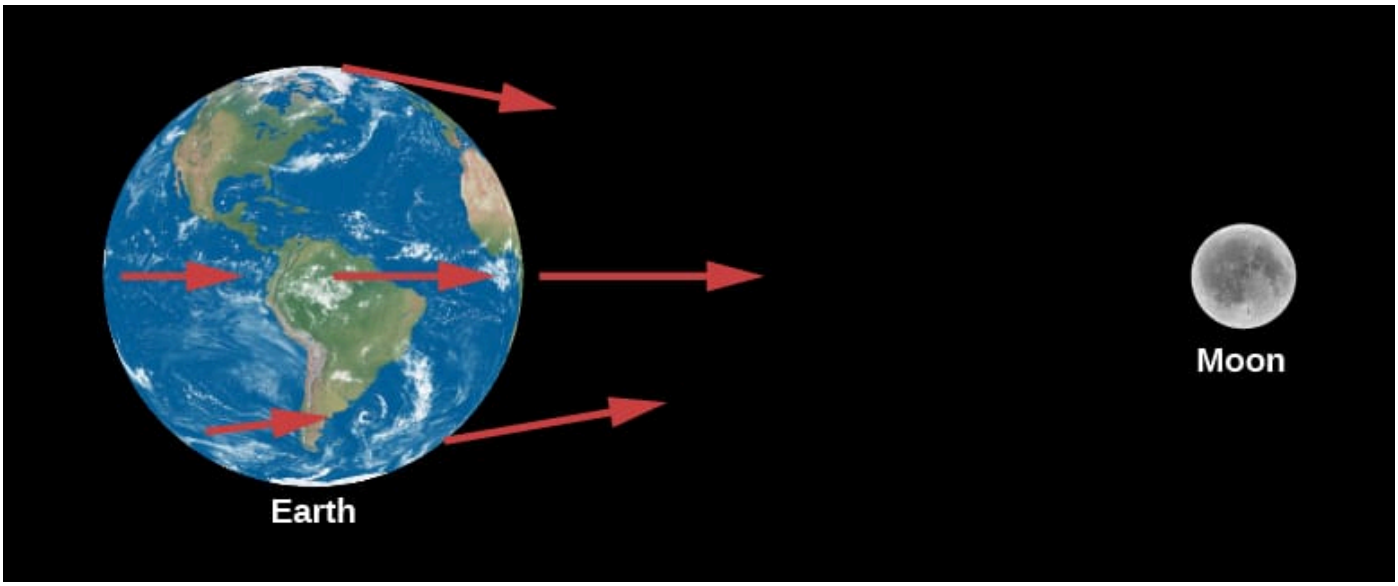


Figure 4.16 : Pull of the Moon. The Moon's differential attraction is shown on different parts of Earth. (Note that the differences have been exaggerated for educational purposes.)

If Earth were made of water, it would distort until the Moon's differential forces over different parts of its surface came into balance with Earth's own gravitational forces pulling it together. Calculations show that in this case, Earth would distort from a sphere by amounts ranging up to nearly 1 meter. Measurements of the actual deformation of Earth show that the solid Earth does distort, but only about one-third as much as water would, because of the greater rigidity of Earth's interior.

Because the tidal distortion of the solid Earth amounts—at its greatest—to only about 20 centimeters, Earth does not distort enough to balance the Moon's differential forces with its own gravity. Hence, objects at Earth's surface experience tiny horizontal tugs, tending to make them slide about. These *tide-raising forces* are too insignificant to affect solid objects like astronomy students or rocks in Earth's crust, but they do affect the waters in the oceans.

The Formation of Tides

The tide-raising forces, acting over a number of hours, produce motions of the water that result in measurable tidal bulges in the oceans. Water on the side of Earth facing the Moon flows toward it, with the greatest depths roughly at the point below the Moon. On the side of Earth opposite the Moon, water also flows to produce a tidal bulge (Figure 4.17).

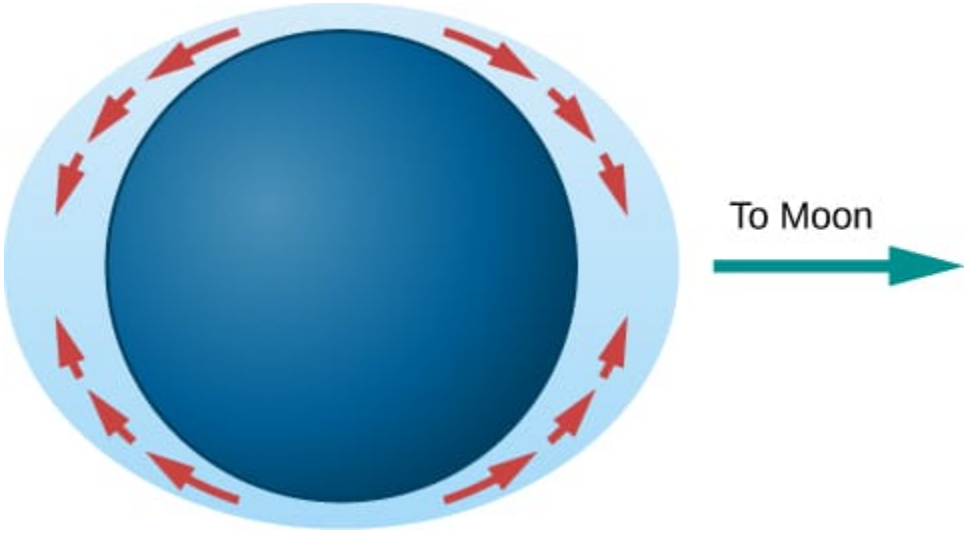


Figure 4.17 : Tidal Bulges in an “Ideal” Ocean. Differences in gravity cause tidal forces that push water in the direction of tidal bulges on Earth.

Link to Learning

You can run this animation for a visual demonstration of the tidal bulge.

Note that the tidal bulges in the oceans do not result from the Moon’s compressing or expanding the water, nor from the Moon’s lifting the water “away from Earth.” Rather, they result from an actual flow of water over Earth’s surface toward the two regions below and opposite the Moon, causing the water to pile up to greater depths at those places (Figure 4.18).



Figure 4.18 : High and Low Tides. This is a side-by-side comparison of the Bay of Fundy in Canada at high and low tides. (credit a, b: modification of work by Dylan Kereluk)

In the idealized (and, as we shall see, oversimplified) model just described, the height of the tides would be only a few feet. The rotation of Earth would carry an observer at any given place alternately into regions of deeper and shallower water. An observer

being carried toward the regions under or opposite the Moon, where the water was deepest, would say, “The tide is coming in”; when carried away from those regions, the observer would say, “The tide is going out.” During a day, the observer would be carried through two tidal bulges (one on each side of Earth) and so would experience two high tides and two low tides.

The Sun also produces tides on Earth, although it is less than half as effective as the Moon at tide raising. The actual tides we experience are a combination of the larger effect of the Moon and the smaller effect of the Sun. When the Sun and Moon are lined up (at new moon or full moon), the tides produced reinforce each other and so are greater than normal (Figure 4.19). These are called spring tides (the name is connected not to the season but to the idea that higher tides “spring up”). Spring tides are approximately the same, whether the Sun and Moon are on the same or opposite sides of Earth, because tidal bulges occur on both sides. When the Moon is at first quarter or last quarter (at right angles to the Sun’s direction), the tides produced by the Sun partially cancel the tides of the Moon, making them lower than usual. These are called neap tides.

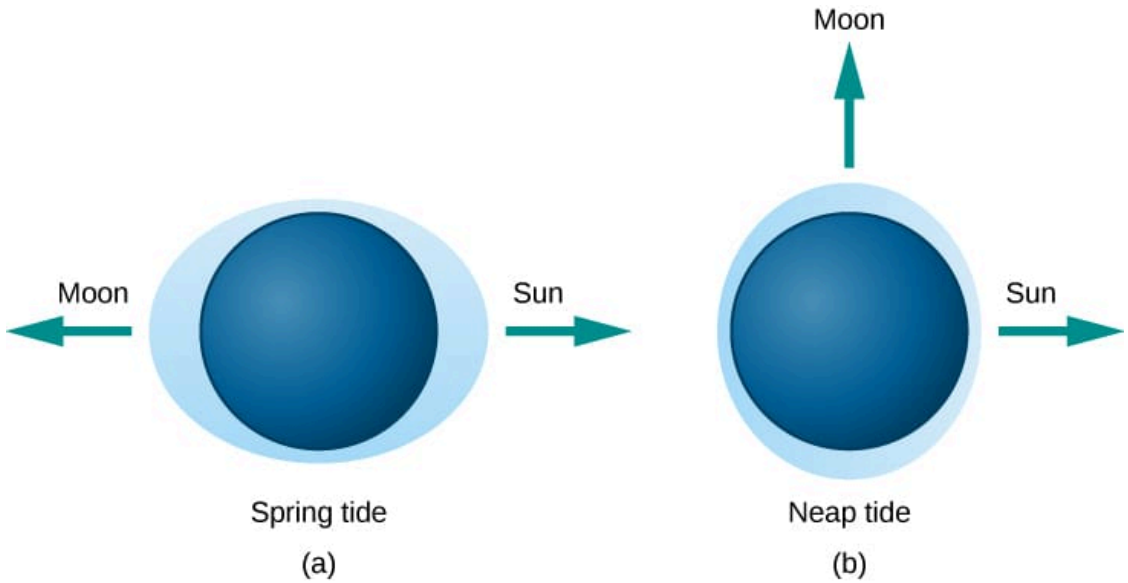


Figure 4.19 : Tides Caused by Different Alignments of the Sun and Moon. (a) In spring tides, the Sun’s and Moon’s pulls reinforce each other. (b) In neap tides, the Sun and the Moon pull at right angles to each other and the resulting tides are lower than usual.

The “simple” theory of tides, described in the preceding paragraphs, would be sufficient if Earth rotated very slowly and were completely surrounded by very deep oceans. However, the presence of land masses stopping the flow of water, the friction in the oceans and between oceans and the ocean floors, the rotation of Earth, the wind, the variable depth of the ocean, and other factors all complicate the picture. This is why, in the real world, some places have very small tides while in other places huge tides become tourist attractions. If you have been in such places, you may know that “tide tables” need to be computed and published for each location; one set of tide predictions doesn’t work for the whole planet. In this introductory chapter, we won’t delve further into these complexities.

Voyagers in Astronomy: George Darwin and the Slowing of Earth

The rubbing of water over the face of Earth involves an enormous amount of energy. Over long periods of time, the friction of the tides is slowing down the rotation of Earth. Our day gets longer by about 0.002 second each century. That seems very small, but such tiny changes can add up over millions and billions of years.

Although Earth’s spin is slowing down, the angular momentum (see Orbits and Gravity) in a system such as the Earth-Moon system cannot change. Thus, some other spin motion must speed up to take the extra angular momentum. The details of what happens were worked out over a century ago by George Darwin, the son of naturalist Charles Darwin. George Darwin (see Figure 4.20) had a strong interest in science but studied law for six years and was admitted to the bar. However, he never practiced law, returning to science instead and eventually becoming a professor at Cambridge University. He was a protégé of Lord Kelvin, one of the great physicists of the nineteenth century, and he became interested in the long-term evolution of the solar system. He specialized in making detailed (and difficult) mathematical calculations of how orbits and motions change over geologic time.

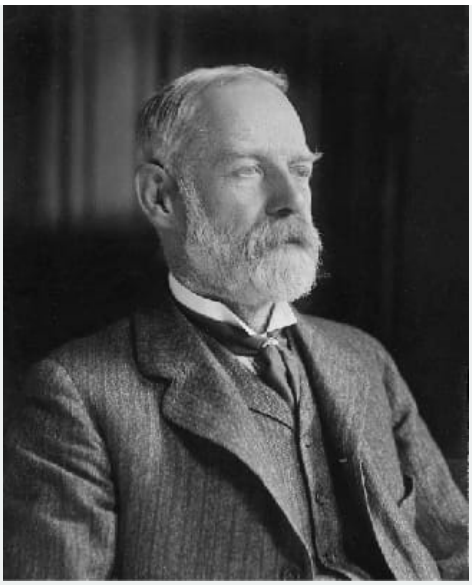


Figure 4.20 : George Darwin (1845–1912). George Darwin is best known for studying Earth's spin in relation to angular momentum.

What Darwin calculated for the Earth-Moon system was that the Moon will slowly spiral outward, away from Earth. As it moves farther away, it will orbit less quickly (just as planets farther from the Sun move more slowly in their orbits). Thus, the month will get longer. Also, because the Moon will be more distant, total eclipses of the Sun will no longer be visible from Earth.

Both the day and the month will continue to get longer, although bear in mind that the effects are very gradual. Darwin's calculations were confirmed by mirrors placed on the Moon by Apollo 11 astronauts. These show that the Moon is moving away by 3.8 centimeters per year, and that ultimately—billions of years in the future—the day and the month will be the same length (about 47 of our present days). At this point the Moon will be stationary in the sky over the same spot on Earth, meaning some parts of Earth will see the Moon and its phases and other parts will never see them. This kind of alignment is already true for Pluto's moon Charon (among others). Its rotation and orbital period are the same length as a day on Pluto.

This page titled 2.6: Ocean Tides and the Moon is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 4.7: Ocean Tides and the Moon by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.7: Eclipses of the Sun and Moon

Learning Objectives

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

- Describe what causes lunar and solar eclipses
- Differentiate between a total and partial solar eclipse
- Explain why lunar eclipses are much more common than solar eclipses

One of the coincidences of living on Earth at the present time is that the two most prominent astronomical objects, the Sun and the Moon, have nearly the same apparent size in the sky. Although the Sun is about 400 times larger in diameter than the Moon, it is also about 400 times farther away, so both the Sun and the Moon have the same angular size—about 1/2°. As a result, the Moon, as seen from Earth, can appear to cover the Sun, producing one of the most impressive events in nature.

Any solid object in the solar system casts a shadow by blocking the light of the Sun from a region behind it. This shadow in space becomes apparent whenever another object moves into it. In general, an *eclipse* occurs whenever any part of either Earth or the Moon enters the shadow of the other. When the Moon's shadow strikes Earth, people within that shadow see the Sun at least partially covered by the Moon; that is, they witness a solar eclipse. When the Moon passes into the shadow of Earth, people on the night side of Earth see the Moon darken in what is called a lunar eclipse. Let's look at how these happen in more detail.

The shadows of Earth and the Moon consist of two parts: a cone where the shadow is darkest, called the *umbra*, and a lighter, more diffuse region of darkness called the *penumbra*. As you can imagine, the most spectacular eclipses occur when an object enters the umbra. Figure 4.21 illustrates the appearance of the Moon's shadow and what the Sun and Moon would look like from different points within the shadow.

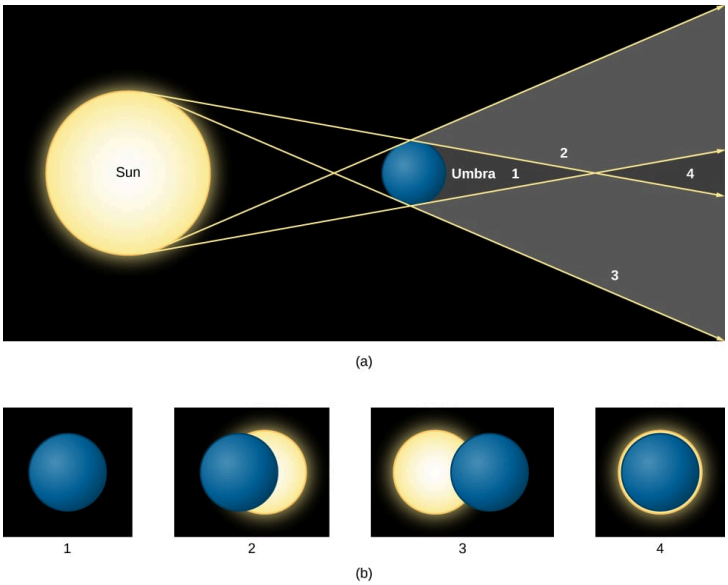


Figure 4.21 : Solar Eclipse. (a) The shadow cast by a spherical body (the Moon, for example) is shown. Notice the dark umbra and the lighter penumbra. Four points in the shadow are labeled with numbers. In (b) you see what the Sun and Moon would look like in the sky at the four labeled points. At position 1, you see a total eclipse. At positions 2 and 3, the eclipse is partial. At position 4, the Moon is farther away and thus cannot cover the Sun completely; a ring of light thus shows around the Sun, creating what is called an "annular" eclipse.

If the path of the Moon in the sky were identical to the path of the Sun (the ecliptic), we might expect to see an eclipse of the Sun and the Moon each month—whenever the Moon got in front of the Sun or into the shadow of Earth. However, as we mentioned, the Moon's orbit is tilted relative to the plane of Earth's orbit about the Sun by about 5° (imagine two hula hoops with a common center, but tilted a bit). As a result, during most months, the Moon is sufficiently above or below the ecliptic plane to avoid an eclipse. But when the two paths cross (twice a year), it is then "eclipse season" and eclipses are possible.

Eclipses of the Sun

The apparent or angular sizes of both the Sun and Moon vary slightly from time to time as their distances from Earth vary. (Figure 4.21 shows the distance of the observer varying at points 1–4, but the idea is the same.) Much of the time, the Moon looks slightly smaller than the Sun and cannot cover it completely, even if the two are perfectly aligned. In this type of “annular eclipse,” there is a ring of light around the dark sphere of the Moon.

However, if an eclipse of the Sun occurs when the Moon is somewhat nearer than its average distance, the Moon can completely hide the Sun, producing a *total* solar eclipse. Another way to say it is that a total eclipse of the Sun occurs at those times when the umbra of the Moon’s shadow reaches the surface of Earth.

The geometry of a total solar eclipse is illustrated in Figure 4.22. If the Sun and Moon are properly aligned, then the Moon’s darkest shadow intersects the ground at a small point on Earth’s surface. Anyone on Earth within the small area covered by the tip of the Moon’s shadow will, for a few minutes, be unable to see the Sun and will witness a total eclipse. At the same time, observers on a larger area of Earth’s surface who are in the penumbra will see only a part of the Sun eclipsed by the Moon: we call this a *partial* solar eclipse.

Between Earth’s rotation and the motion of the Moon in its orbit, the tip of the Moon’s shadow sweeps eastward at about 1500 kilometers per hour along a thin band across the surface of Earth. The thin zone across Earth within which a total solar eclipse is visible (weather permitting) is called the eclipse path. Within a region about 3000 kilometers on either side of the eclipse path, a partial solar eclipse is visible. It does not take long for the Moon’s shadow to sweep past a given point on Earth. The duration of totality may be only a brief instant; it can never exceed about 7 minutes.

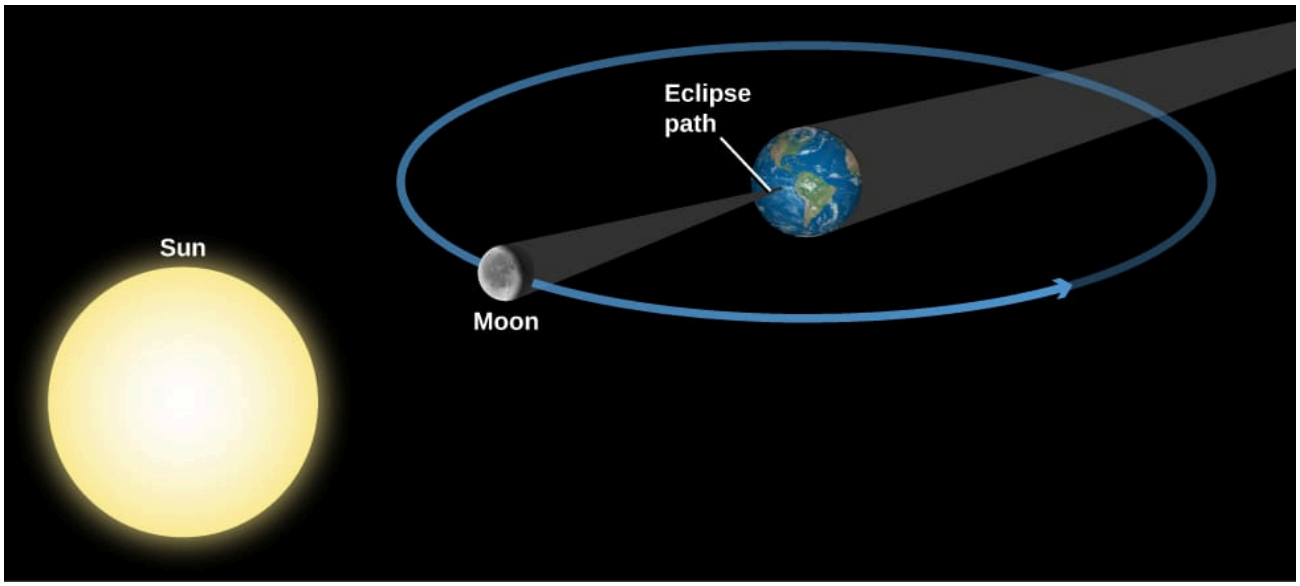


Figure 4.22 : Geometry of a Total Solar Eclipse. Note that our diagram is not to scale. The Moon blocks the Sun during new moon phase as seen from some parts of Earth and casts a shadow on our planet.

Because a total eclipse of the Sun is so spectacular, it is well worth trying to see one if you can. There are some people whose hobby is “eclipse chasing” and who brag about how many they have seen in their lifetimes. Because much of Earth’s surface is water, eclipse chasing can involve lengthy boat trips (and often requires air travel as well). As a result, eclipse chasing is rarely within the budget of a typical college student. Nevertheless, a list of future eclipses is given for your reference in Appendix H, just in case you strike it rich early. (And, as you can see in the Appendix, there was a total eclipse visible in the United States in 2017, and there will be another in 2024, to which college students may be able to afford travel.)

Appearance of a Total Eclipse

What can you see if you are lucky enough to catch a total eclipse? A solar eclipse starts when the Moon just begins to silhouette itself against the edge of the Sun’s disk. A partial phase follows, during which more and more of the Sun is covered by the Moon. About an hour after the eclipse begins, the Sun becomes completely hidden behind the Moon. In the few minutes immediately

before this period of totality begins, the sky noticeably darkens, some flowers close up, and chickens may go to roost. As an eerie twilight suddenly descends during the day, other animals (and people) may get disoriented. During totality, the sky is dark enough that planets become visible in the sky, and usually the brighter stars do as well.

As the bright disk of the Sun becomes entirely hidden behind the Moon, the Sun's remarkable corona flashes into view (Figure 4.23). The *corona* is the Sun's outer atmosphere, consisting of sparse gases that extend for millions of miles in all directions from the apparent surface of the Sun. It is ordinarily not visible because the light of the corona is feeble compared with the light from the underlying layers of the Sun. Only when the brilliant glare from the Sun's visible disk is blotted out by the Moon during a total eclipse is the pearly white corona visible. (We'll talk more about the corona in the chapter on The Sun: A Garden-Variety Star.)

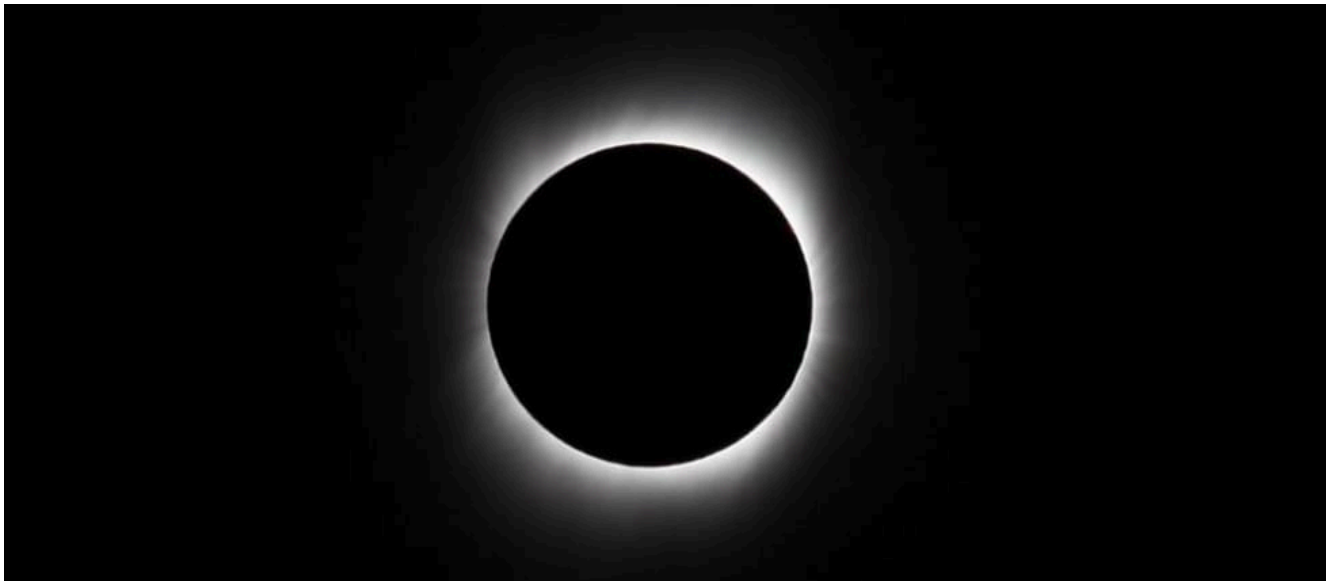


Figure 4.23 : The Sun's Corona. The corona (thin outer atmosphere) of the Sun is visible during a total solar eclipse. (It looks more extensive in photographs than it would to the unaided eye.) (credit: modification of work by Lutfar Rahman Nirjhar)

The total phase of the eclipse ends, as abruptly as it began, when the Moon begins to uncover the Sun. Gradually, the partial phases of the eclipse repeat themselves, in reverse order, until the Moon has completely uncovered the Sun. We should make one important safety point here: while the few minutes of the *total* eclipse are safe to look at, if any part of the Sun is uncovered, you must protect your eyes with safe eclipse glasses[2] or by projecting an image of the Sun (instead of looking at it directly). For more, read the How to Observe Solar Eclipses box in this chapter.

Eclipses of the Moon

A lunar eclipse occurs when the Moon enters the shadow of Earth. The geometry of a lunar eclipse is shown in Figure 4.24. Earth's dark shadow is about 1.4 million kilometers long, so at the Moon's distance (an average of 384,000 kilometers), it could cover about four full moons. Unlike a solar eclipse, which is visible only in certain local areas on Earth, a lunar eclipse is visible to everyone who can see the Moon. Because a lunar eclipse can be seen (weather permitting) from the entire night side of Earth, lunar eclipses are observed far more frequently from a given place on Earth than are solar eclipses.

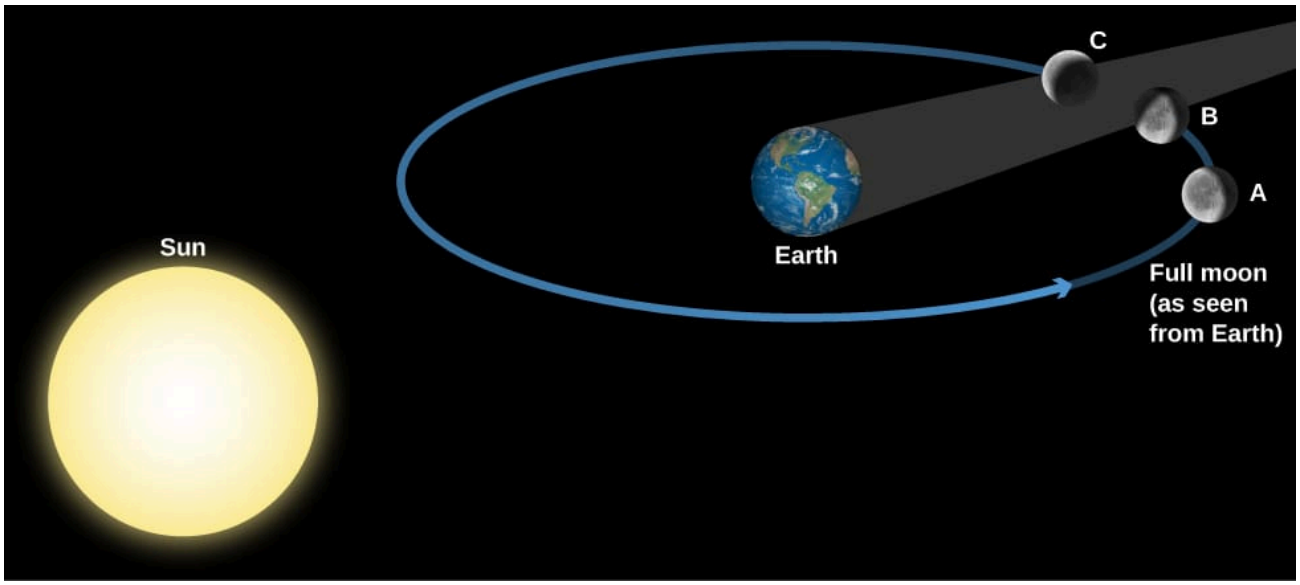


Figure 4.24 : Geometry of a Lunar Eclipse. The Moon is shown moving through the different parts of Earth's shadow during a total lunar eclipse. Note that the distance the Moon moves in its orbit during the eclipse has been exaggerated here for clarity.

An eclipse of the Moon is total only if the Moon's path carries it though Earth's umbra. If the Moon does not enter the umbra completely, we have a partial eclipse of the Moon. But because Earth is larger than the Moon, its umbra is larger, so that lunar eclipses last longer than solar eclipses, as we will discuss below.

A lunar eclipse can take place only when the Sun, Earth, and Moon are in a line. The Moon is opposite the Sun, which means the Moon will be in full phase before the eclipse, making the darkening even more dramatic. About 20 minutes before the Moon reaches the dark shadow, it dims somewhat as Earth partly blocks the sunlight. As the Moon begins to dip into the shadow, the curved shape of Earth's shadow upon it soon becomes apparent.

Even when totally eclipsed, the Moon is still faintly visible, usually appearing a dull coppery red. The illumination on the eclipsed Moon is sunlight that has been bent into Earth's shadow by passing through Earth's atmosphere.

After totality, the Moon moves out of the shadow and the sequence of events is reversed. The total duration of the eclipse depends on how closely the Moon's path approaches the axis of the shadow. For an eclipse where the Moon goes through the center of Earth's shadow, each partial phase consumes at least 1 hour, and totality can last as long as 1 hour and 40 minutes. Eclipses of the Moon are much more "democratic" than solar eclipses. Since the full moon is visible on the entire night side of Earth, the lunar eclipse is visible for all those who live in that hemisphere. (Recall that a total eclipse of the Sun is visible only in a narrow path where the shadow of the umbra falls.) Total eclipses of the Moon occur, on average, about once every two or three years. A list of future total eclipses of the Moon is in Appendix H. In addition, since the lunar eclipse happens to a full moon, and a full moon is not dangerous to look at, everyone can look at the Moon during all the parts of the eclipse without worrying about safety.

Thanks to our understanding of gravity and motion (see Orbits and Gravity), eclipses can now be predicted centuries in advance. We've come a long way since humanity stood frightened by the darkening of the Sun or the Moon, fearing the displeasure of the gods. Today, we enjoy the sky show with a healthy appreciation of the majestic forces that keep our solar system running.

Use the Solar and Lunar Eclipse simulation to explore the conditions of solar and lunar eclipses yourself.

Seeing for Yourself: How to Observe Solar Eclipses

A total eclipse of the Sun is a spectacular sight and should not be missed. However, it is extremely dangerous to look directly at the Sun: even a brief exposure can damage your eyes. Normally, few rational people are tempted to do this because it is painful (and something your mother told you never to do!). But during the partial phases of a solar eclipse, the temptation to

take a look is strong. Think before you give in. The fact that the Moon is covering part of the Sun doesn't make the uncovered part any less dangerous to look at. Still, there are perfectly safe ways to follow the course of a solar eclipse, if you are lucky enough to be in the path of the shadow.

The easiest technique is to make a pinhole projector. Take a piece of cardboard with a small (1 millimeter) hole punched in it, and hold it several feet above a light surface, such as a concrete sidewalk or a white sheet of paper, so that the hole is "aimed" at the Sun. The hole produces a fuzzy but adequate image of the eclipsed Sun. Alternatively, if it's the right time of year, you can let the tiny spaces between a tree's leaves form multiple pinhole images against a wall or sidewalk. Watching hundreds of little crescent Suns dancing in the breeze can be captivating. A kitchen colander also makes an excellent pinhole projector.

Although there are safe filters for looking at the Sun directly, people have suffered eye damage by looking through improper filters, or no filter at all. For example, neutral density photographic filters are not safe because they transmit infrared radiation that can cause severe damage to the retina. Also unsafe are smoked glass, completely exposed color film, sunglasses, and many other homemade filters. Safe filters include welders' goggles and specially designed eclipse glasses.

You should certainly look at the Sun directly when it is totally eclipsed, even through binoculars or telescopes. Unfortunately, the total phase, as we discussed, is all too brief. But if you know when it is coming and going, be sure you look, for it's an unforgettably beautiful sight. And, despite the ancient folklore that presents eclipses as dangerous times to be outdoors, the partial phases of eclipses—as long as you are not looking directly at the Sun—are not any more dangerous than being out in sunlight.

During past eclipses, unnecessary panic has been created by uninformed public officials acting with the best intentions. There were two marvelous total eclipses in Australia in the twentieth century during which townspeople held newspapers over their heads for protection and schoolchildren cowered indoors with their heads under their desks. What a pity that all those people missed what would have been one of the most memorable experiences of their lifetimes.

On August 21, 2017, a total eclipse of the Sun was visible across a large swath of the continental United States, and was seen by millions of people from around the country and the world (a).



Figure 4.25 : 2017 Total Solar and Lunar Eclipse (a) The eclipsed Sun August 21, 2017, showing remarkable detail in the Sun's outer atmosphere. This is a composite of short, medium, and long exposures, as no single exposure can capture the huge range of brightness the Sun exhibits. (b) A total eclipse of the Moon seen over California on January 31, 2018. The Moon moves slowly into the Earth's shadow, looks red when the eclipse is total and red sunlight refracts through the Earth's atmosphere, and then slowly moves out of the shadow. (credit a: modification of work by Rick Fienberg, American Astronomical Socity/TravelQuest International; credit b: modification of work by Brian Day.)

Link to Learning

For the August 2017 eclipse, the United States postal service issued a special commemorative stamp — the first ever with “thermochromic ink” which changes when you touch it.

Footnotes

- 2Eclipse glasses are available in many planetarium and observatory gift stores, and also from the two main U.S. manufacturers: American Paper Optics and Rainbow Symphony.

This page titled 2.7: Eclipses of the Sun and Moon is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 4.8: Eclipses of the Sun and Moon by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

3: THE SCIENCE OF ASTRONOMY

3.1: Ancient Astronomy
3.2: Astrology and Astronomy
3.3: The Birth of Modern Astronomy
3.4: The Laws of Planetary Motion
3.5: The Nature of Science

This page titled 3: The Science of Astronomy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.1: Ancient Astronomy

Learning Objectives

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

- Describe early examples of astronomy around the world
- Explain how Greek astronomers were able to deduce that Earth is spherical
- Explain how Greek astronomers were able to calculate Earth's size
- Describe the motion of Earth called precession
- Describe Ptolemy's geocentric system of planetary motion

Let us now look briefly back into history. Much of modern Western civilization is derived in one way or another from the ideas of the ancient Greeks and Romans, and this is true in astronomy as well. However, many other ancient cultures also developed sophisticated systems for observing and interpreting the sky.

Astronomy around the World

Ancient Babylonian, Assyrian, and Egyptian astronomers knew the approximate length of the year. The Egyptians of 3000 years ago, for example, adopted a calendar based on a 365-day year. They kept careful track of the rising time of the bright star Sirius in the predawn sky, which has a yearly cycle that corresponded with the flooding of the Nile River. The Chinese also had a working calendar; they determined the length of the year at about the same time as the Egyptians. The Chinese also recorded comets, bright meteors, and dark spots on the Sun. (Many types of astronomical objects were introduced in Science and the Universe: A Brief Tour. If you are not familiar with terms like *comets* and *meteors*, you may want to review that chapter.) Later, Chinese astronomers kept careful records of "guest stars"—those that are normally too faint to see but suddenly flare up to become visible to the unaided eye for a few weeks or months. We still use some of these records in studying stars that exploded a long time ago.

The Mayan culture in Mexico and Central America developed a sophisticated calendar based on the planet Venus, and they made astronomical observations from sites dedicated to this purpose a thousand years ago. The Polynesians learned to navigate by the stars over hundreds of kilometers of open ocean—a skill that enabled them to colonize new islands far away from where they began.

In Britain, before the widespread use of writing, ancient people used stones to keep track of the motions of the Sun and Moon. We still find some of the great stone circles they built for this purpose, dating from as far back as 2800 BCE. The best known of these is Stonehenge, which is discussed in Earth, Moon, and Sky.[1]

Early Greek and Roman Cosmology

Our concept of the cosmos—its basic structure and origin—is called cosmology, a word with Greek roots. Before the invention of telescopes, humans had to depend on the simple evidence of their senses for a picture of the universe. The ancients developed cosmologies that combined their direct view of the heavens with a rich variety of philosophical and religious symbolism.

At least 2000 years before Columbus, educated people in the eastern Mediterranean region knew Earth was round. Belief in a spherical Earth may have stemmed from the time of Pythagoras, a philosopher and mathematician who lived 2500 years ago. He believed circles and spheres to be "perfect forms" and suggested that Earth should therefore be a sphere. As evidence that the gods liked spheres, the Greeks cited the fact that the Moon is a sphere, using evidence we describe later.

The writings of Aristotle (384–322 BCE), the tutor of Alexander the Great, summarize many of the ideas of his day. They describe how the progression of the Moon's phases—its apparent changing shape—results from our seeing different portions of the Moon's sunlit hemisphere as the month goes by (see Earth, Moon, and Sky). Aristotle also knew that the Sun has to be farther away from Earth than is the Moon because occasionally the Moon passed exactly between Earth and the Sun and hid the Sun temporarily from view. We call this a *solar eclipse*.

Aristotle cited convincing arguments that Earth must be round. First is the fact that as the Moon enters or emerges from Earth's shadow during an eclipse of the Moon, the shape of the shadow seen on the Moon is always round (Figure 2.9). Only a spherical object always produces a round shadow. If Earth were a disk, for example, there would be some occasions when the sunlight would strike it edge-on and its shadow on the Moon would be a line.



Figure 2.9 : Earth's Round Shadow. A lunar eclipse occurs when the Moon moves into and out of Earth's shadow. Note the curved shape of the shadow—evidence for a spherical Earth that has been recognized since antiquity. (credit: modification of work by Brian Paczkowski)

As a second argument, Aristotle explained that travelers who go south a significant distance are able to observe stars that are not visible farther north. And the height of the North Star—the star nearest the north celestial pole—decreases as a traveler moves south. On a flat Earth, everyone would see the same stars overhead. The only possible explanation is that the traveler must have moved over a curved surface on Earth, showing stars from a different angle. (See the How Do We Know Earth Is Round? feature for more ideas on proving Earth is round.)

One Greek thinker, Aristarchus of Samos (310–230 BCE), even suggested that Earth was moving around the Sun, but Aristotle and most of the ancient Greek scholars rejected this idea. One of the reasons for their conclusion was the thought that if Earth moved about the Sun, they would be observing the stars from different places along Earth's orbit. As Earth moved along, nearby stars should shift their positions in the sky relative to more distant stars. In a similar way, we see foreground objects appear to move against a more distant background whenever we are in motion. When we ride on a train, the trees in the foreground appear to shift their position relative to distant hills as the train rolls by. Unconsciously, we use this phenomenon all of the time to estimate distances around us.

The apparent shift in the direction of an object as a result of the motion of the observer is called parallax. We call the shift in the apparent direction of a star due to Earth's orbital motion *stellar parallax*. The Greeks made dedicated efforts to observe stellar parallax, even enlisting the aid of Greek soldiers with the clearest vision, but to no avail. The brighter (and presumably nearer) stars just did not seem to shift as the Greeks observed them in the spring and then again in the fall (when Earth is on the opposite side of the Sun).

This meant either that Earth was not moving or that the stars had to be so tremendously far away that the parallax shift was immeasurably small. A cosmos of such enormous extent required a leap of imagination that most ancient philosophers were not prepared to make, so they retreated to the safety of the Earth-centered view, which would dominate Western thinking for nearly two millennia.

Astronomy Basics: How Do We Know Earth Is Round?

In addition to the two ways (from Aristotle's writings) discussed in this chapter, you might also reason as follows:

1. Let's watch a ship leave its port and sail into the distance on a clear day. On a flat Earth, we would just see the ship get smaller and smaller as it sails away. But this isn't what we actually observe. Instead, ships sink below the horizon, with the hull disappearing first and the mast remaining visible for a while longer. Eventually, only the top of the mast can be seen as the ship sails around the curvature of Earth. Finally, the ship disappears under the horizon.
2. The International Space Station circles Earth once every 90 minutes or so. Photographs taken from the shuttle and other satellites show that Earth is round from every perspective.
3. Suppose you made a friend in each time zone of Earth. You call all of them at the same hour and ask, "Where is the Sun?" On a flat Earth, each caller would give you roughly the same answer. But on a round Earth you would find that, for some friends, the Sun would be high in the sky whereas for others it would be rising, setting, or completely out of sight (and this last group of friends would be upset with you for waking them up).

Measurement of Earth by Eratosthenes

The Greeks not only knew Earth was round, but also they were able to measure its size. The first fairly accurate determination of Earth's diameter was made in about 200 BCE by Eratosthenes (276–194 BCE), a Greek living in Alexandria, Egypt. His method was a geometric one, based on observations of the Sun.

The Sun is so distant from us that all the light rays that strike our planet approach us along essentially parallel lines. To see why, look at Figure 2.10. Take a source of light near Earth—say, at position A. Its rays strike different parts of Earth along diverging paths. From a light source at B, or at C (which is still farther away), the angle between rays that strike opposite parts of Earth is smaller. The more distant the source, the smaller the angle between the rays. For a source infinitely distant, the rays travel along parallel lines.

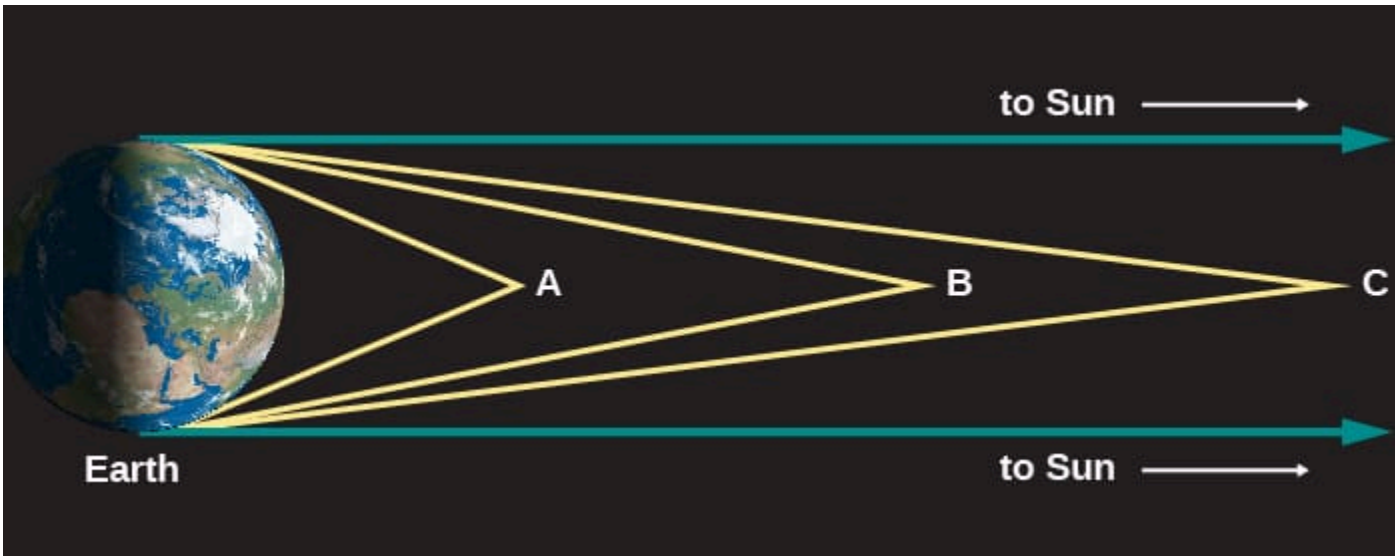


Figure 2.10 : Light Rays from Space. The more distant an object, the more nearly parallel the rays of light coming from it.

Of course, the Sun is not infinitely far away, but given its distance of 150 million kilometers, light rays striking Earth from a point on the Sun diverge from one another by an angle far too small to be observed with the unaided eye. As a consequence, if people all over Earth who could see the Sun were to point at it, their fingers would, essentially, all be parallel to one another. (The same is also true for the planets and stars—an idea we will use in our discussion of how telescopes work.)

Eratosthenes was told that on the first day of summer at Syene, Egypt (near modern Aswan), sunlight struck the bottom of a vertical well at noon. This indicated that the Sun was directly over the well—meaning that Syene was on a direct line from the center of Earth to the Sun. At the corresponding time and date in Alexandria, Eratosthenes observed the shadow a column made and saw that the Sun was not directly overhead, but was slightly south of the zenith, so that its rays made an angle with the vertical equal to about 1/50 of a circle (7°). Because the Sun's rays striking the two cities are parallel to one another, why would the two rays not make the same angle with Earth's surface? Eratosthenes reasoned that the curvature of the round Earth meant that "straight up" was not the same in the two cities. And the measurement of the angle in Alexandria, he realized, allowed him to figure out the size of Earth. Alexandria, he saw, must be 1/50 of Earth's circumference north of Syene (Figure 2.11). Alexandria had been measured to be 5000 stadia north of Syene. (The *stadium* was a Greek unit of length, derived from the length of the racetrack in a stadium.) Eratosthenes thus found that Earth's circumference must be 50×5000, or 250,000 stadia.

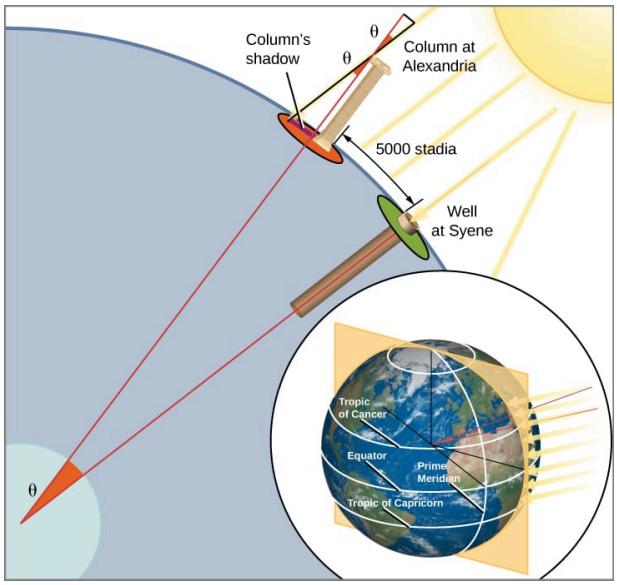


Figure 2.11 : How Eratosthenes Measured the Size of Earth. Eratosthenes measured the size of Earth by observing the angle at which the Sun's rays hit our planet's surface. The Sun's rays come in parallel, but because Earth's surface curves, a ray at Syene comes straight down whereas a ray at Alexandria makes an angle of 7° with the vertical. That means, in effect, that at Alexandria, Earth's surface has curved away from Syene by 7° of 360°, or 1/50 of a full circle. Thus, the distance between the two cities must be 1/50 the circumference of Earth. (credit: modification of work by NOAA Ocean Service Education)

It is not possible to evaluate precisely the accuracy of Eratosthenes solution because there is doubt about which of the various kinds of Greek stadia he used as his unit of distance. If it was the common Olympic stadium, his result is about 20% too large. According to another interpretation, he used a stadium equal to about 1/6 kilometer, in which case his figure was within 1% of the correct value of 40,000 kilometers. Even if his measurement was not exact, his success at measuring the size of our planet by using only shadows, sunlight, and the power of human thought was one of the greatest intellectual achievements in history.

Hipparchus and Precession

Perhaps the greatest astronomer of antiquity was Hipparchus, born in Nicaea in what is present-day Turkey. He erected an observatory on the island of Rhodes around 150 BCE, when the Roman Republic was expanding its influence throughout the Mediterranean region. There he measured, as accurately as possible, the positions of objects in the sky, compiling a pioneering star catalog with about 850 entries. He designated celestial coordinates for each star, specifying its position in the sky, just as we specify the position of a point on Earth by giving its latitude and longitude.

He also divided the stars into apparent magnitudes according to their apparent brightness. He called the brightest ones "stars of the first magnitude"; the next brightest group, "stars of the second magnitude"; and so forth. This rather arbitrary system, in modified form, still remains in use today (although it is less and less useful for professional astronomers).

By observing the stars and comparing his data with older observations, Hipparchus made one of his most remarkable discoveries: the position in the sky of the north celestial pole had altered over the previous century and a half. Hipparchus deduced correctly that this had happened not only during the period covered by his observations, but was in fact happening all the time: the direction around which the sky appears to rotate changes slowly but continuously. Recall from the section on celestial poles and the celestial equator that the north celestial pole is just the projection of Earth's North Pole into the sky. If the north celestial pole is wobbling around, then Earth itself must be doing the wobbling. Today, we understand that the direction in which Earth's axis points does indeed change slowly but regularly—a motion we call precession. If you have ever watched a spinning top wobble, you observed a similar kind of motion. The top's axis describes a path in the shape of a cone, as Earth's gravity tries to topple it (Figure 2.12).

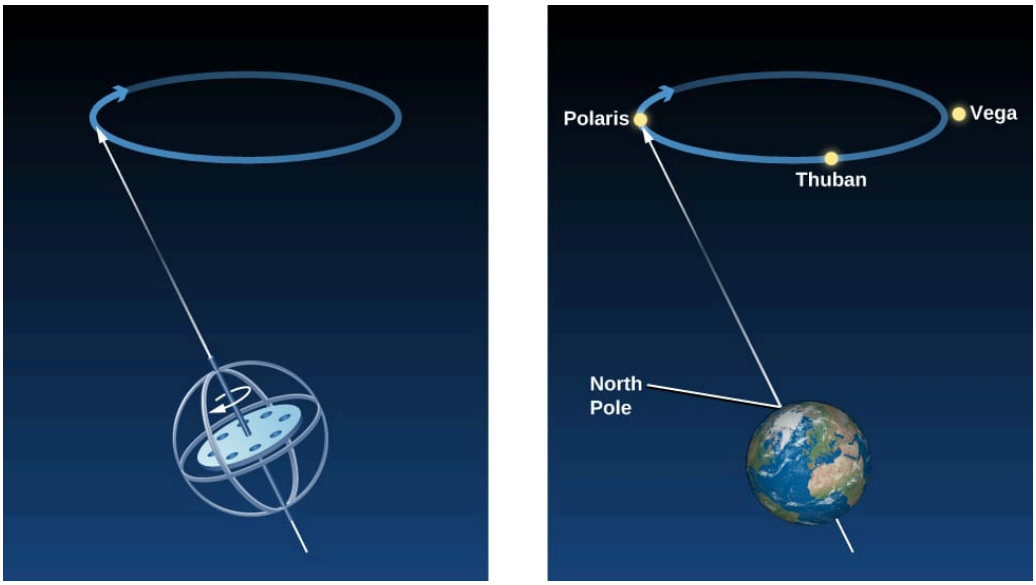


Figure 2.12 : Precession. Just as the axis of a rapidly spinning top wobbles slowly in a circle, so the axis of Earth wobbles in a 26,000-year cycle. Today the north celestial pole is near the star Polaris, but about 5000 years ago it was close to a star called Thuban, and in 14,000 years it will be closest to the star Vega.

Because our planet is not an exact sphere, but bulges a bit at the equator, the pulls of the Sun and Moon cause it to wobble like a top. It takes about 26,000 years for Earth's axis to complete one circle of precession. As a result of this motion, the point where our axis points in the sky changes as time goes on. While Polaris is the star closest to the north celestial pole today (it will reach its closest point around the year 2100), the star Vega in the constellation of Lyra will be the North Star in 14,000 years.

Ptolemy's Model of the Solar System

The last great astronomer of the Roman era was Claudius Ptolemy (or Ptolemaeus), who flourished in Alexandria in about the year 140. He wrote a mammoth compilation of astronomical knowledge, which today is called by its Arabic name, *Almagest* (meaning "The Greatest"). *Almagest* does not deal exclusively with Ptolemy's own work; it includes a discussion of the astronomical achievements of the past, principally those of Hipparchus. Today, it is our main source of information about the work of Hipparchus and other Greek astronomers.

Ptolemy's most important contribution was a geometric representation of the solar system that predicted the positions of the planets for any desired date and time. Hipparchus, not having enough data on hand to solve the problem himself, had instead amassed observational material for posterity to use. Ptolemy supplemented this material with new observations of his own and produced a cosmological model that endured more than a thousand years, until the time of Copernicus.

The complicating factor in explaining the motions of the planets is that their apparent wandering in the sky results from the combination of their own motions with Earth's orbital revolution. As we watch the planets from our vantage point on the moving Earth, it is a little like watching a car race while you are competing in it. Sometimes opponents' cars pass you, but at other times you pass them, making them appear to move backward for a while with respect to you.

Figure 2.13 shows the motion of Earth and a planet farther from the Sun—in this case, Mars. Earth travels around the Sun in the same direction as the other planet and in nearly the same plane, but its orbital speed is faster. As a result, it overtakes the planet periodically, like a faster race car on the inside track. The figure shows where we see the planet in the sky at different times. The path of the planet among the stars is illustrated in the star field on the right side of the figure.

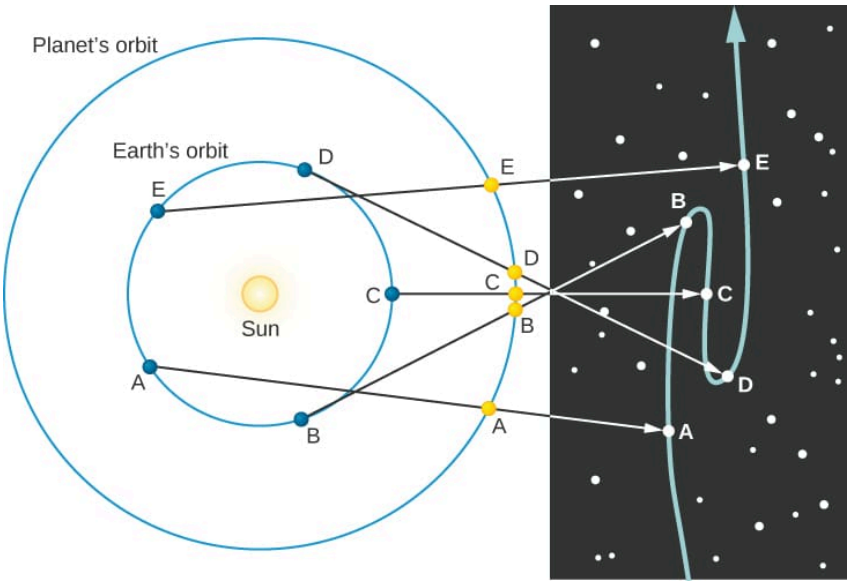


Figure 2.13 : Retrograde Motion of a Planet beyond Earth's Orbit. The letters on the diagram show where Earth and Mars are at different times. By following the lines from each Earth position through each corresponding Mars position, you can see how the retrograde path of Mars looks against the background stars.

Link to Learning

The planetary configurations simulator from Foothill AstroSims allows you to see the usual prograde and occasional retrograde motion of other planets. You can switch back and forth between viewing motion from Earth and Mars (as well as other planets).

Normally, planets move eastward in the sky over the weeks and months as they orbit the Sun, but from positions B to D in Figure 2.13, as Earth passes the planets in our example, it appears to drift backward, moving west in the sky. Even though it is actually moving to the east, the faster-moving Earth has overtaken it and seems, from our perspective, to be leaving it behind. As Earth rounds its orbit toward position E, the planet again takes up its apparent eastward motion in the sky. The temporary apparent westward motion of a planet as Earth swings between it and the Sun is called retrograde motion. Such backward motion is much easier for us to understand today, now that we know Earth is one of the moving planets and not the unmoving center of all creation. But Ptolemy was faced with the far more complex problem of explaining such motion while assuming a stationary Earth.

Furthermore, because the Greeks believed that celestial motions had to be circles, Ptolemy had to construct his model using circles alone. To do it, he needed dozens of circles, some moving around other circles, in a complex structure that makes a modern viewer dizzy. But we must not let our modern judgment cloud our admiration for Ptolemy's achievement. In his day, a complex universe centered on Earth was perfectly reasonable and, in its own way, quite beautiful. However, as Alfonso X, the King of Castile, was reported to have said after having the Ptolemaic system of planet motions explained to him, "If the Lord Almighty had consulted me before embarking upon Creation, I should have recommended something simpler."

Ptolemy solved the problem of explaining the observed motions of planets by having each planet revolve in a small orbit called an epicycle. The center of the epicycle then revolved about Earth on a circle called a *deferent* (Figure 2.14). When the planet is at position *x* in Figure 2.14 on the epicycle orbit, it is moving in the same direction as the center of the epicycle; from Earth, the planet appears to be moving eastward. When the planet is at *y*, however, its motion is in the direction opposite to the motion of the epicycle's center around Earth. By choosing the right combination of speeds and distances, Ptolemy succeeded in having the planet moving westward at the correct speed and for the correct interval of time, thus replicating retrograde motion with his model.

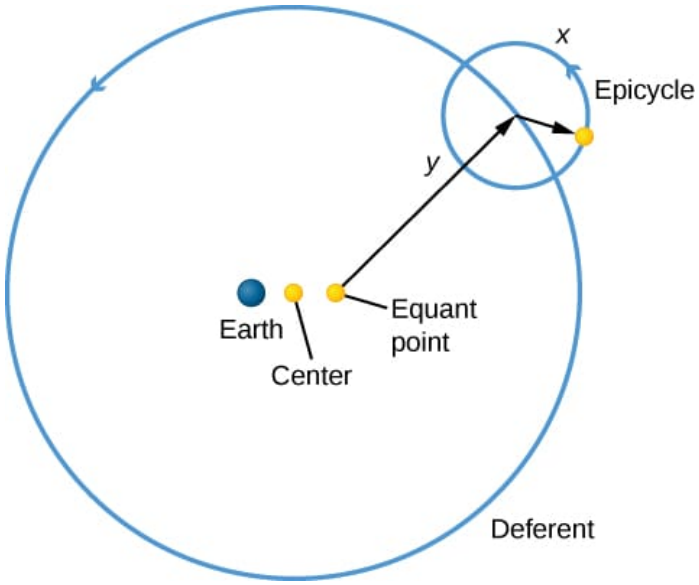


Figure 2.14 : Ptolemy's Complicated Cosmological System. Each planet orbits around a small circle called an *epicycle*. Each epicycle orbits on a larger circle called the *deferent*. This system is not centered exactly on Earth but on an offset point called the *equant*. The Greeks needed all this complexity to explain the actual motions in the sky because they believed that Earth was stationary and that all sky motions had to be circular.

Link to Learning

Use the Ptolemaic System simulator from Foothill AstroSims to explore how Ptolemy's system of deferents and epicycles explained the apparent motion of the planets.

However, we shall see in Orbits and Gravity that the planets, like Earth, travel about the Sun in orbits that are ellipses, not circles. Their actual behavior cannot be represented accurately by a scheme of uniform circular motions. In order to match the observed motions of the planets, Ptolemy had to center the deferent circles, not on Earth, but at points some distance from Earth. In addition, he introduced uniform circular motion around yet another axis, called the *equant point*. All of these considerably complicated his scheme.

It is a tribute to the genius of Ptolemy as a mathematician that he was able to develop such a complex system to account successfully for the observations of planets. It may be that Ptolemy did not intend for his cosmological model to describe reality, but merely to serve as a mathematical representation that allowed him to predict the positions of the planets at any time. Whatever his thinking, his model, with some modifications, was eventually accepted as authoritative in the Muslim world and (later) in Christian Europe.

Footnotes

- 1For resources providing more information about the astronomy of diverse cultures around the world, please see http://bit.ly/astrocultures.

This page titled 3.1: Ancient Astronomy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 2.3: Ancient Astronomy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

3.2: Astrology and Astronomy

Learning Objectives

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

- Explain the origins of astrology
- Explain what a horoscope is
- Summarize the arguments that invalidate astrology as a scientific practice

Many ancient cultures regarded the planets and stars as representatives or symbols of the gods or other supernatural forces that controlled their lives. For them, the study of the heavens was not an abstract subject; it was connected directly to the life-and-death necessity of understanding the actions of the gods and currying favor with them. Before the time of our scientific perspectives, everything that happened in nature—from the weather, to diseases and accidents, to celestial surprises such as eclipses or new comets—was thought to be an expression of the whims or displeasure of the gods. Any signs that helped people understand what these gods had in mind were considered extremely important.

The movements of the seven objects that had the power to "wander" through the realm of the sky—the Sun, the Moon, and five planets visible to the unaided eye—clearly must have special significance in such a system of thinking.

Most ancient cultures associated these seven objects with various supernatural rulers in their pantheon and kept track of them for religious reasons. Even in the comparatively sophisticated Greece of antiquity, the planets had the names of gods and were credited with having the same powers and influences as the gods whose names they bore. From such ideas was born the ancient system called astrology, still practiced by some people today, in which the positions of these bodies among the stars of the zodiac are thought to hold the key to understanding what we can expect from life.

The Beginnings of Astrology

Astrology began in Babylonia about two and half millennia ago. The Babylonians, believing the planets and their motions influenced the fortunes of kings and nations, used their knowledge of astronomy to guide their rulers. When the Babylonian culture was absorbed by the Greeks, astrology gradually came to influence the entire Western world and eventually spread to Asia as well.

By the 2nd century BCE the Greeks democratized astrology by developing the idea that the planets influence every individual. In particular, they believed that the configuration of the Sun, Moon, and planets at the moment of birth affected a person's personality and fortune—a doctrine called *natal astrology*. Natal astrology reached its peak with Ptolemy 400 years later. As famous for his astrology as for his astronomy, Ptolemy compiled the *Tetrabiblos*, a treatise on astrology that remains the "bible" of the subject. It is essentially this ancient religion, older than Christianity or Islam, that is still practiced by today's astrologers.

The Horoscope

The key to natal astrology is the horoscope, a chart showing the positions of the planets in the sky at the moment of an individual's birth. The word "horoscope" comes from the Greek words *hora* (meaning "time") and *skopos* (meaning a "watcher" or "marker"), so "horoscope" can literally be translated as "marker of the hour." When a horoscope is charted, the planets (including the Sun and Moon, classed as *wanderers* by the ancients) must first be located in the zodiac. At the time astrology was set up, the zodiac was divided into 12 sectors called *signs* (Figure 2.15), each 30° long. Each sign was named after a constellation in the sky through which the Sun, Moon, and planets were seen to pass—the sign of Virgo after the constellation of Virgo, for example.



Figure 2.15 : Zodiac Signs. The signs of the zodiac are shown in a medieval woodcut.

When someone today casually asks you your "sign," they are asking for your "sun sign"—which zodiac sign the Sun was in at the moment you were born. However, more than 2000 years have passed since the signs received their names from the constellations. Because of precession, the constellations of the zodiac slide westward along the ecliptic, going once around the sky in about 26,000 years. Thus, today the real stars have slipped around by about 1/12 of the zodiac—about the width of one sign.

In most forms of astrology, however, the signs have remained assigned to the dates of the year they had when astrology was first set up. This means that the astrological signs and the real constellations are out of step; the sign of Aries, for example, now occupies the constellation of Pisces. When you look up your sun sign in a newspaper astrology column, the name of the sign associated with your birthday is no longer the name of the constellation in which the Sun was actually located when you were born. To know that constellation, you must look for the sign before the one that includes your birthday.

A complete horoscope shows the location of not only the Sun, but also the Moon and each planet in the sky by indicating its position in the appropriate sign of the zodiac. However, as the celestial sphere turns (owing to the rotation of Earth), the entire zodiac moves across the sky to the west, completing a circuit of the heavens each day. Thus, the position in the sky (or "house" in astrology) must also be calculated. There are more or less standardized rules for the interpretation of the horoscope, most of which (at least in Western schools of astrology) are derived from the *Tetrabiblos* of Ptolemy. Each sign, each house, and each planet—the last acting as a center of force—is supposed to be associated with particular matters in a person's life.

The detailed interpretation of a horoscope is a very complicated business, and there are many schools of astrological thought on how it should be done. Although some of the rules may be standardized, how each rule is to be weighed and applied is a matter of judgment—and "art." It also means that it is very difficult to tie down astrology to specific predictions or to get the same predictions from different astrologers.

Astrology Today

Astrologers today use the same basic principles laid down by Ptolemy nearly 2000 years ago. They cast horoscopes (a process much simplified by the development of appropriate computer programs) and suggest interpretations. Sun sign astrology (which you read in the newspapers and many magazines) is a recent, simplified variant of natal astrology. Although even professional astrologers do not place much trust in such a limited scheme, which tries to fit everyone into just 12 groups, sun sign astrology is taken seriously by many people (perhaps because it is discussed so commonly in the media).

Today, we know much more about the nature of the planets as physical bodies, as well as about human genetics, than the ancients could. It is hard to imagine how the positions of the Sun, Moon, or planets in the sky at the moment of our birth could have anything to do with our personality or future. There are no known forces, not gravity or anything else, that could cause such effects. (For example, a straightforward calculation shows that the gravitational pull of the obstetrician delivering a newborn baby is greater than that of Mars.) Astrologers thus have to argue there must be unknown forces exerted by the planets that depend

on their configurations with respect to one another and that do not vary according to the distance of the planet—forces for which there is no shred of evidence.

Another curious aspect of astrology is its emphasis on planet configurations at birth. What about the forces that might influence us at conception? Isn't our genetic makeup more important for determining our personality than the circumstances of our birth? Would we really be a different person if we had been born a few hours earlier or later, as astrology claims? (Back when astrology was first conceived, birth was thought of as a moment of magic significance, but today we understand a lot more about the long process that precedes it.)

Actually, very few well-educated people today buy the claim that our entire lives are predetermined by astrological influences at birth, but many people apparently believe that astrology has validity as an indicator of affinities and personality. A surprising number of Americans make judgments about people—whom they will hire, associate with, and even marry—on the basis of astrological information. To be sure, these are difficult decisions, and you might argue that we should use any relevant information that might help us to make the right choices. But does astrology actually provide any useful information on human personality? This is the kind of question that can be tested using the scientific method (see Testing Astrology).

The results of hundreds of tests are all the same: there is no evidence that natal astrology has any predictive power, even in a statistical sense. Why, then, do people often seem to have anecdotes about how well their own astrologer advised them? Effective astrologers today use the language of the zodiac and the horoscope only as the outward trappings of their craft. Mostly they work as amateur therapists, offering simple truths that clients like or need to hear. (Recent studies have shown that just about any sort of short-term therapy makes people feel a little better because the very act of talking about our problems with someone who listens attentively is, in itself, beneficial.)

The scheme of astrology has no basis in scientific fact, however; at best, it can be described as a pseudoscience. It is an interesting historical system, left over from prescientific days and best remembered for the impetus it gave people to learn the cycles and patterns of the sky. From it grew the science of astronomy, which is our main subject for discussion.

Making Connections: Testing Astrology

In response to modern public interest in astrology, scientists have carried out a wide range of statistical tests to assess its predictive power. The simplest of these examine sun sign astrology to determine whether—as astrologers assert—some signs are more likely than others to be associated with some objective measure of success, such as winning Olympic medals, earning high corporate salaries, or achieving elective office or high military rank. (You can devise such a test yourself by looking up the birth dates of all members of Congress, for example, or all members of the U.S. Olympic team.) Are our political leaders somehow selected at birth by their horoscopes and thus more likely to be Leos, say, than Scorpios?

You do not even need to be specific about your prediction in such tests. After all, many practitioners of astrology disagree about which signs go with which personality characteristics. To demonstrate the validity of the astrological hypothesis, it would be sufficient if the birthdays of all our leaders clustered in any one or two signs in some statistically significant way. Dozens of such tests have been performed, and all have come up completely negative: the birth dates of leaders in all fields tested have been found to be distributed randomly among *all* the signs. Sun sign astrology does not predict anything about a person's future occupation or strong personality traits.

In a fine example of such a test, two statisticians examined the reenlistment records of the United States Marine Corps. We suspect you will agree that it takes a certain kind of personality not only to enlist, but also to reenlist in the Marines. If sun signs can predict strong personality traits—as astrologers claim—then those who reenlisted (with similar personalities) should have been distributed preferentially in those one or few signs that matched the personality of someone who loves being a Marine. However, the reenlisted were distributed randomly among all the signs.

More sophisticated studies have also been done, involving full horoscopes calculated for thousands of individuals. The results of all these studies are also negative: none of the systems of astrology has been shown to be at all effective in connecting astrological aspects to personality, success, or finding the right person to love.

Other tests show that it hardly seems to matter what a horoscope interpretation says, as long as it is vague enough, and as long as each subject feels it was prepared personally just for him or her. The French statistician Michel Gauquelin, for example, sent the horoscope interpretation for one of the worst mass murderers in history to 150 people, but told each recipient that it was a “reading” prepared exclusively for him or her. Ninety-four percent of the readers said they recognized themselves in the interpretation of the mass murderer’s horoscope.

Geoffrey Dean, an Australian researcher, reversed the astrological readings of 22 subjects, substituting phrases that were the opposite of what the horoscope actually said. Yet, his subjects said that the resulting readings applied to them just as often (95%) as the people to whom the original phrases were given.

Link to Learning

For more on astrology and science from an astronomer’s point of view, read this article that shines light on the topic through an accessible Q&A.

This page titled 3.2: Astrology and Astronomy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 2.4: Astrology and Astronomy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

3.3: The Birth of Modern Astronomy

Learning Objectives

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

- Explain how Copernicus developed the heliocentric model of the solar system
- Explain the Copernican model of planetary motion and describe evidence or arguments in favor of it
- Describe Galileo's discoveries concerning the study of motion and forces
- Explain how Galileo's discoveries tilted the balance of evidence in favor of the Copernican model

Astronomy made no major advances in strife-torn medieval Europe. The birth and expansion of Islam after the seventh century led to a flowering of Arabic and Jewish cultures that preserved, translated, and added to many of the astronomical ideas of the Greeks. Many of the names of the brightest stars, for example, are today taken from the Arabic, as are such astronomical terms as "zenith."

As European culture began to emerge from its long, dark age, trading with Arab countries led to a rediscovery of ancient texts such as *Almagest* and to a reawakening of interest in astronomical questions. This time of rebirth (in French, "*renaissance*") in astronomy was embodied in the work of Copernicus (Figure 2.16).

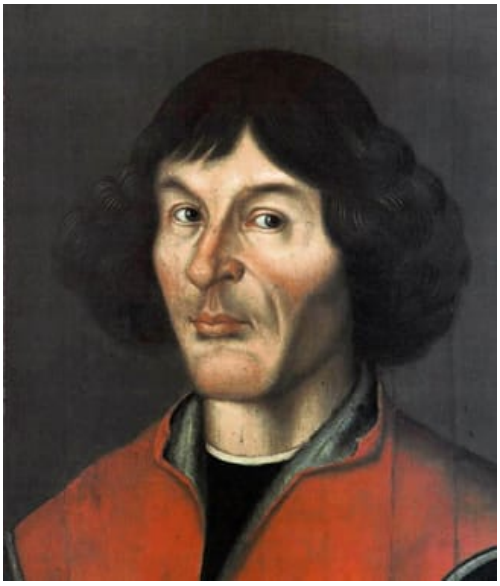


Figure 2.16 : Nicolaus Copernicus (1473–1543). Copernicus was a cleric and scientist who played a leading role in the emergence of modern science. Although he could not prove that Earth revolves about the Sun, he presented such compelling arguments for this idea that he turned the tide of cosmological thought and laid the foundations upon which Galileo and Kepler so effectively built in the following century.

Copernicus

One of the most important events of the Renaissance was the displacement of Earth from the center of the universe, an intellectual revolution initiated by a Polish cleric in the sixteenth century. Nicolaus Copernicus was born in Torun, a mercantile town along the Vistula River. His training was in law and medicine, but his main interests were astronomy and mathematics. His great contribution to science was a critical reappraisal of the existing theories of planetary motion and the development of a new Sun-centered, or heliocentric, model of the solar system. Copernicus concluded that Earth is a planet and that all the planets circle the Sun. Only the Moon orbits Earth (Figure 2.17).

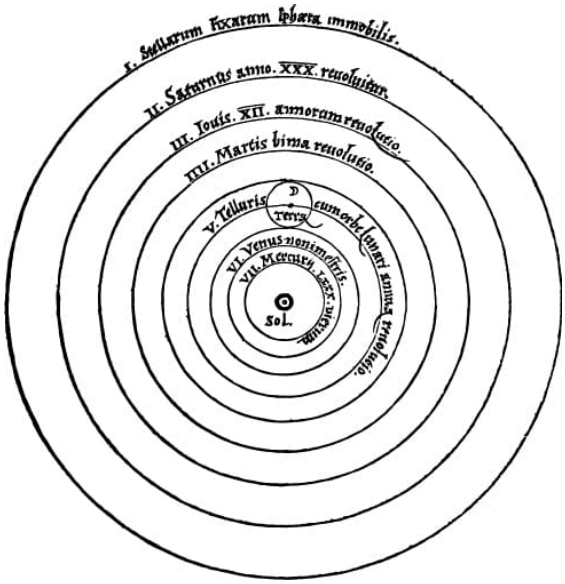


Figure 2.17 : Copernicus' System. Copernicus developed a heliocentric plan of the solar system. This system was published in the first edition of *De Revolutionibus Orbium Coelestium*. Notice the word *Sol* for "Sun" in the middle. (credit: Nicolai Copernici)

Copernicus described his ideas in detail in his book *De Revolutionibus Orbium Coelestium* (*On the Revolution of Celestial Orbs*), published in 1543, the year of his death. By this time, the old Ptolemaic system needed significant adjustments to predict the positions of the planets correctly. Copernicus wanted to develop an improved theory from which to calculate planetary positions, but in doing so, he was himself not free of all traditional prejudices.

He began with several assumptions that were common in his time, such as the idea that the motions of the heavenly bodies must be made up of combinations of uniform circular motions. But he did not assume (as most people did) that Earth had to be in the center of the universe, and he presented a defense of the heliocentric system that was elegant and persuasive. His ideas, although not widely accepted until more than a century after his death, were much discussed among scholars and, ultimately, had a profound influence on the course of world history.

One of the objections raised to the heliocentric theory was that if Earth were moving, we would all sense or feel this motion. Solid objects would be ripped from the surface, a ball dropped from a great height would not strike the ground directly below it, and so forth. But a moving person is not necessarily aware of that motion. We have all experienced seeing an adjacent train, bus, or ship appear to move, only to discover that it is we who are moving.

Copernicus argued that the apparent motion of the Sun about Earth during the course of a year could be represented equally well by a motion of Earth about the Sun. He also reasoned that the apparent rotation of the celestial sphere could be explained by assuming that Earth rotates while the celestial sphere is stationary. To the objection that if Earth rotated about an axis it would fly into pieces, Copernicus answered that if such motion would tear Earth apart, the still faster motion of the much larger celestial sphere required by the geocentric hypothesis would be even more devastating.

The Heliocentric Model

The most important idea in Copernicus' *De Revolutionibus* is that Earth is one of six (then-known) planets that revolve about the Sun. Using this concept, he was able to work out the correct general picture of the solar system. He placed the planets, starting nearest the Sun, in the correct order: Mercury, Venus, Earth, Mars, Jupiter, and Saturn. Further, he deduced that the nearer a planet is to the Sun, the greater its orbital speed. With his theory, he was able to explain the complex retrograde motions of the planets without epicycles and to work out a roughly correct scale for the solar system.

Copernicus could not prove that Earth revolves about the Sun. In fact, with some adjustments, the old Ptolemaic system could have accounted, as well, for the motions of the planets in the sky. But Copernicus pointed out that the Ptolemaic cosmology was clumsy and lacking the beauty and symmetry of its successor.

In Copernicus' time, in fact, few people thought there were ways to prove whether the heliocentric or the older geocentric system was correct. A long philosophical tradition, going back to the Greeks and defended by the Catholic Church, held that pure human

thought combined with divine revelation represented the path to truth. Nature, as revealed by our senses, was suspect. For example, Aristotle had reasoned that heavier objects (having more of the quality that made them heavy) must fall to Earth faster than lighter ones. This is absolutely incorrect, as any simple experiment dropping two balls of different weights shows. However, in Copernicus' day, experiments did not carry much weight (if you will pardon the expression); Aristotle's reasoning was more convincing.

In this environment, there was little motivation to carry out observations or experiments to distinguish between competing cosmological theories (or anything else). It should not surprise us, therefore, that the heliocentric idea was debated for more than half a century without any tests being applied to determine its validity. (In fact, in the North American colonies, the older geocentric system was still taught at Harvard University in the first years after it was founded in 1636.)

Contrast this with the situation today, when scientists rush to test each new hypothesis and do not accept any ideas until the results are in. For example, when two researchers at the University of Utah announced in 1989 that they had discovered a way to achieve nuclear fusion (the process that powers the stars) at room temperature, other scientists at more than 25 laboratories around the United States attempted to duplicate "cold fusion" within a few weeks—without success, as it turned out. The cold fusion theory soon went down in flames.

How would we look at Copernicus' model today? When a new hypothesis or theory is proposed in science, it must first be checked for consistency with what is already known. Copernicus' heliocentric idea passes this test, for it allows planetary positions to be calculated at least as well as does the geocentric theory. The next step is to determine which predictions the new hypothesis makes that differ from those of competing ideas. In the case of Copernicus, one example is the prediction that, if Venus circles the Sun, the planet should go through the full range of phases just as the Moon does, whereas if it circles Earth, it should not (Figure 2.18). Also, we should not be able to see the full phase of Venus from Earth because the Sun would then be between Venus and Earth. But in those days, before the telescope, no one imagined testing these predictions.

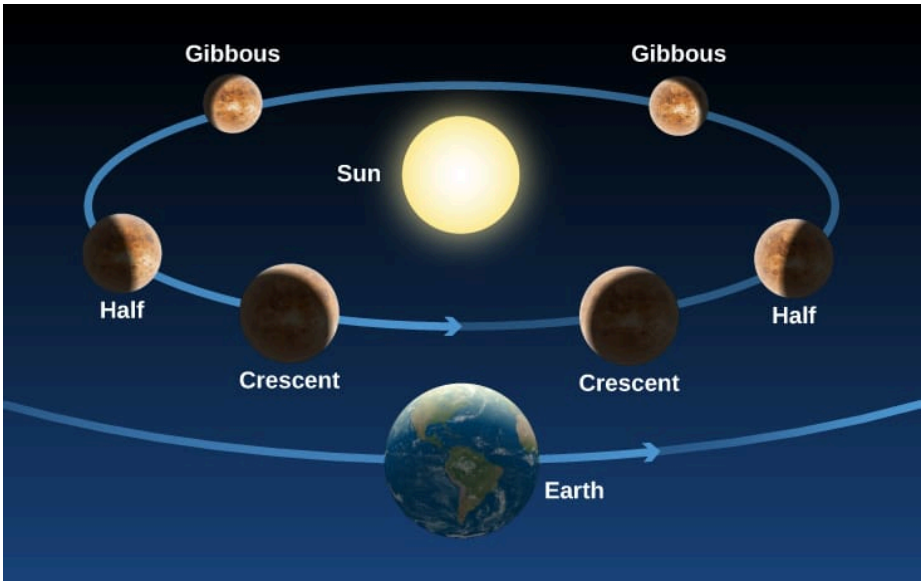


Figure 2.18 : Phases of Venus. As Venus moves around the Sun, we see changing illumination of its surface, just as we see the face of the Moon illuminated differently in the course of a month.

Link to Learning

This animation shows the phases of Venus. You can also see its distance from Earth as it orbits the Sun.

Galileo and the Beginning of Modern Science

Many of the modern scientific concepts of observation, experimentation, and the testing of hypotheses through careful quantitative measurements were pioneered by a man who lived nearly a century after Copernicus. Galileo Galilei (Figure 2.19), a contemporary of Shakespeare, was born in Pisa. Like Copernicus, he began training for a medical career, but he had little interest in the subject and later switched to mathematics. He held faculty positions at the University of Pisa and the University of Padua,

and eventually became mathematician to the Grand Duke of Tuscany in Florence.



Figure 2.19 : Galileo Galilei (1564–1642). Galileo advocated that we perform experiments or make observations to ask nature its ways. When Galileo turned the telescope to the sky, he found things were not the way philosophers had supposed.

Galileo's greatest contributions were in the field of mechanics, the study of motion and the actions of forces on bodies. It was familiar to all persons then, as it is to us now, that if something is at rest, it tends to remain at rest and requires some outside influence to start it in motion. Rest was thus generally regarded as the natural state of matter. Galileo showed, however, that rest is no more natural than motion.

If an object is slid along a rough horizontal floor, it soon comes to rest because friction between it and the floor acts as a retarding force. However, if the floor and the object are both highly polished, the object, given the same initial speed, will slide farther before stopping. On a smooth layer of ice, it will slide farther still. Galileo reasoned that if all resisting effects could be removed, the object would continue in a steady state of motion indefinitely. He argued that a force is required not only to start an object moving from rest but also to slow down, stop, speed up, or change the direction of a moving object. You will appreciate this if you have ever tried to stop a rolling car by leaning against it, or a moving boat by tugging on a line.

Galileo also studied the way objects accelerate—change their speed or direction of motion. Galileo watched objects as they fell freely or rolled down a ramp. He found that such objects accelerate uniformly; that is, in equal intervals of time they gain equal increments in speed. Galileo formulated these newly found laws in precise mathematical terms that enabled future experimenters to predict how far and how fast objects would move in various lengths of time.

Link to Learning

In theory, if Galileo is right, a feather and a hammer, dropped at the same time from a height, should land at the same moment. On Earth, this experiment is not possible because air resistance and air movements make the feather flutter, instead of falling straight down, accelerated only by the force of gravity. For generations, physics teachers had said that the place to try this experiment is somewhere where there is no air, such as the Moon. In 1971, *Apollo 15* astronaut David Scott took a hammer and feather to the Moon and tried it, to the delight of physics nerds everywhere. NASA provides the video of the hammer and feather as well as a brief explanation.

Sometime in the 1590s, Galileo adopted the Copernican hypothesis of a heliocentric solar system. In Roman Catholic Italy, this was not a popular philosophy, for Church authorities still upheld the ideas of Aristotle and Ptolemy, and they had powerful political and economic reasons for insisting that Earth was the center of creation. Galileo not only challenged this thinking but also had the audacity to write in Italian rather than scholarly Latin, and to lecture publicly on those topics. For him, there was no contradiction between the authority of the Church in matters of religion and morality, and the authority of nature (revealed by experiments) in matters of science. It was primarily because of Galileo and his "dangerous" opinions that, in 1616, the Church issued a prohibition decree stating that the Copernican doctrine was "false and absurd" and not to be held or defended.

Galileo's Astronomical Observations

It is not certain who first conceived of the idea of combining two or more pieces of glass to produce an instrument that enlarged images of distant objects, making them appear nearer. The first such "spyglasses" (now called *telescopes*) that attracted much notice were made in 1608 by the Dutch spectacle maker Hans Lippershey (1570–1619). Galileo heard of the discovery and, without ever having seen an assembled telescope, constructed one of his own with a three-power magnification (3×), which made distant objects appear three times nearer and larger (Figure 2.20).



Figure 2.20 : Telescope Used by Galileo. The telescope has a wooden tube covered with paper and a lens 26 millimeters across.

On August 25, 1609, Galileo demonstrated a telescope with a magnification of 9× to government officials of the city-state of Venice. By a magnification of 9×, we mean the linear dimensions of the objects being viewed appeared nine times larger or, alternatively, the objects appeared nine times closer than they really were. There were obvious military advantages associated with a device for seeing distant objects. For his invention, Galileo's salary was nearly doubled, and he was granted lifetime tenure as a professor. (His university colleagues were outraged, particularly because the invention was not even original.)

Others had used the telescope before Galileo to observe things on Earth. But in a flash of insight that changed the history of astronomy, Galileo realized that he could turn the power of the telescope toward the heavens. Before using his telescope for astronomical observations, Galileo had to devise a stable mount and improve the optics. He increased the magnification to 30×. Galileo also needed to acquire confidence in the telescope.

At that time, human eyes were believed to be the final arbiter of truth about size, shape, and color. Lenses, mirrors, and prisms were known to distort distant images by enlarging, reducing, or inverting them, or spreading the light into a spectrum (rainbow of colors). Galileo undertook repeated experiments to convince himself that what he saw through the telescope was identical to what he saw up close. Only then could he begin to believe that the miraculous phenomena the telescope revealed in the heavens were real.

Beginning his astronomical work late in 1609, Galileo found that many stars too faint to be seen with the unaided eye became visible with his telescope. In particular, he found that some nebulous blurs resolved into many stars, and that the Milky Way—the strip of whiteness across the night sky—was also made up of a multitude of individual stars.

Examining the planets, Galileo found four moons revolving about Jupiter in times ranging from just under 2 days to about 17 days. This discovery was particularly important because it showed that not everything has to revolve around Earth. Furthermore, it demonstrated that there could be centers of motion that are themselves in motion. Defenders of the geocentric view had argued that if Earth was in motion, then the Moon would be left behind because it could hardly keep up with a rapidly moving planet. Yet, here were Jupiter's moons doing exactly that. (To recognize this discovery and honor his work, NASA named a spacecraft that explored the Jupiter system Galileo.)

With his telescope, Galileo was able to carry out the test of the Copernican theory mentioned earlier, based on the phases of Venus. Within a few months, he had found that Venus goes through phases like the Moon, showing that it must revolve about the Sun, so that we see different parts of its daylight side at different times (see Figure 2.18.) These observations could not be reconciled with Ptolemy's model, in which Venus circled about Earth. In Ptolemy's model, Venus could also show phases, but they were the wrong phases in the wrong order from what Galileo observed.

Galileo also observed the Moon and saw craters, mountain ranges, valleys, and flat, dark areas that he thought might be water. These discoveries showed that the Moon might be not so dissimilar to Earth—suggesting that Earth, too, could belong to the realm of celestial bodies.

Link to Learning

For more information about the life and work of Galileo, see the Galileo Project at Rice University.

After Galileo's work, it became increasingly difficult to deny the Copernican view, and Earth was slowly dethroned from its central position in the universe and given its rightful place as one of the planets attending the Sun. Initially, however, Galileo met with a great deal of opposition. The Roman Catholic Church, still reeling from the Protestant Reformation, was looking to assert its authority and chose to make an example of Galileo. He had to appear before the Inquisition to answer charges that his work was heretical, and he was ultimately condemned to house arrest. His books were on the Church's forbidden list until 1836, although in countries where the Roman Catholic Church held less sway, they were widely read and discussed. Not until 1992 did the Catholic Church admit publicly that it had erred in the matter of censoring Galileo's ideas.

The new ideas of Copernicus and Galileo began a revolution in our conception of the cosmos. It eventually became evident that the universe is a vast place and that Earth's role in it is relatively unimportant. The idea that Earth moves around the Sun like the other planets raised the possibility that they might be worlds themselves, perhaps even supporting life. As Earth was demoted from its position at the center of the universe, so, too, was humanity. The universe, despite what we may wish, does not revolve around us.

Most of us take these things for granted today, but four centuries ago such concepts were frightening and heretical for some, immensely stimulating for others. The pioneers of the Renaissance started the European world along the path toward science and technology that we still tread today. For them, nature was rational and ultimately knowable, and experiments and observations provided the means to reveal its secrets.

Seeing for Yourself: Observing the Planets

At most any time of the night, and at any season, you can spot one or more bright planets in the sky. All five of the planets known to the ancients—Mercury, Venus, Mars, Jupiter, and Saturn—are more prominent than any but the brightest stars, and they can be seen even from urban locations if you know where and when to look. One way to tell planets from bright stars is that planets twinkle less.

Venus, which stays close to the Sun from our perspective, appears either as an "evening star" in the west after sunset or as a "morning star" in the east before sunrise. It is the brightest object in the sky after the Sun and Moon. It far outshines any real star, and under the most favorable circumstances, it can even cast a visible shadow. Some young military recruits have tried to shoot Venus down as an approaching enemy craft or UFO.

Mars, with its distinctive red color, can be nearly as bright as Venus is when close to Earth, but normally it remains much less conspicuous. Jupiter is most often the second-brightest planet, approximately equaling in brilliance the brightest stars. Saturn is dimmer, and it varies considerably in brightness, depending on whether its large rings are seen nearly edge-on (faint) or more widely opened (bright).

Mercury is quite bright, but few people ever notice it because it never moves very far from the Sun (it's never more than 28° away in the sky) and is always seen against bright twilight skies.

True to their name, the planets "wander" against the background of the "fixed" stars. Although their apparent motions are complex, they reflect an underlying order upon which the heliocentric model of the solar system, as described in this chapter, was based. The positions of the planets are often listed in newspapers (sometimes on the weather page), and clear maps and guides to their locations can be found each month in such magazines as *Sky & Telescope* and *Astronomy* (available at most libraries and online). There are also a number of computer programs and phone and tablet apps that allow you to display where the planets are on any night.

This page titled 3.3: The Birth of Modern Astronomy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 2.5: The Birth of Modern Astronomy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

3.4: The Laws of Planetary Motion

Learning Objectives

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

- Describe how Tycho Brahe and Johannes Kepler contributed to our understanding of how planets move around the Sun
- Explain Kepler's three laws of planetary motion

At about the time that Galileo was beginning his experiments with falling bodies, the efforts of two other scientists dramatically advanced our understanding of the motions of the planets. These two astronomers were the observer Tycho Brahe and the mathematician Johannes Kepler. Together, they placed the speculations of Copernicus on a sound mathematical basis and paved the way for the work of Isaac Newton in the next century.

Tycho Brahe's Observatory

Three years after the publication of Copernicus' *De Revolutionibus*, Tycho Brahe was born to a family of Danish nobility. He developed an early interest in astronomy and, as a young man, made significant astronomical observations. Among these was a careful study of what we now know was an exploding star that flared up to great brilliance in the night sky. His growing reputation gained him the patronage of the Danish King Frederick II, and at the age of 30, Brahe was able to establish a fine astronomical observatory on the North Sea island of Hven (Figure 3.2). Brahe was the last and greatest of the pre-telescopic observers in Europe.



(a) (b)

Figure 3.2 : Tycho Brahe (1546–1601) and Johannes Kepler (1571–1630). (a) A stylized engraving shows Tycho Brahe using his instruments to measure the altitude of celestial objects above the horizon. The large curved instrument in the foreground allowed him to measure precise angles in the sky. Note that the scene includes hints of the grandeur of Brahe's observatory at Hven. (b) Kepler was a German mathematician and astronomer. His discovery of the basic laws that describe planetary motion placed the heliocentric cosmology of Copernicus on a firm mathematical basis.

At Hven, Brahe made a continuous record of the positions of the Sun, Moon, and planets for almost 20 years. His extensive and precise observations enabled him to note that the positions of the planets varied from those given in published tables, which were based on the work of Ptolemy. These data were extremely valuable, but Brahe didn't have the ability to analyze them and develop a better model than what Ptolemy had published. He was further inhibited because he was an extravagant and cantankerous fellow, and he accumulated enemies among government officials. When his patron, Frederick II, died in 1597, Brahe lost his political base and decided to leave Denmark. He took up residence in Prague, where he became court astronomer to Emperor Rudolf of

Bohemia. There, in the year before his death, Brahe found a most able young mathematician, Johannes Kepler, to assist him in analyzing his extensive planetary data.

Johannes Kepler

Johannes Kepler was born into a poor family in the German province of Württemberg and lived much of his life amid the turmoil of the Thirty Years' War (see Figure 3.2). He attended university at Tubingen and studied for a theological career. There, he learned the principles of the Copernican system and became converted to the heliocentric hypothesis. Eventually, Kepler went to Prague to serve as an assistant to Brahe, who set him to work trying to find a satisfactory theory of planetary motion—one that was compatible with the long series of observations made at Hven. Brahe was reluctant to provide Kepler with much material at any one time for fear that Kepler would discover the secrets of the universal motion by himself, thereby robbing Brahe of some of the glory. Only after Brahe's death in 1601 did Kepler get full possession of the priceless records. Their study occupied most of Kepler's time for more than 20 years.

Through his analysis of the motions of the planets, Kepler developed a series of principles, now known as *Kepler's three laws,* which described the behavior of planets based on their paths through space. The first two laws of planetary motion were published in 1609 in *The New Astronomy*. Their discovery was a profound step in the development of modern science.

The First Two Laws of Planetary Motion

The path of an object through space is called its orbit. Kepler initially assumed that the orbits of planets were circles, but doing so did not allow him to find orbits that were consistent with Brahe's observations. Working with the data for Mars, he eventually discovered that the orbit of that planet had the shape of a somewhat flattened circle, or ellipse. Next to the circle, the ellipse is the simplest kind of closed curve, belonging to a family of curves known as *conic sections* (Figure 3.3).

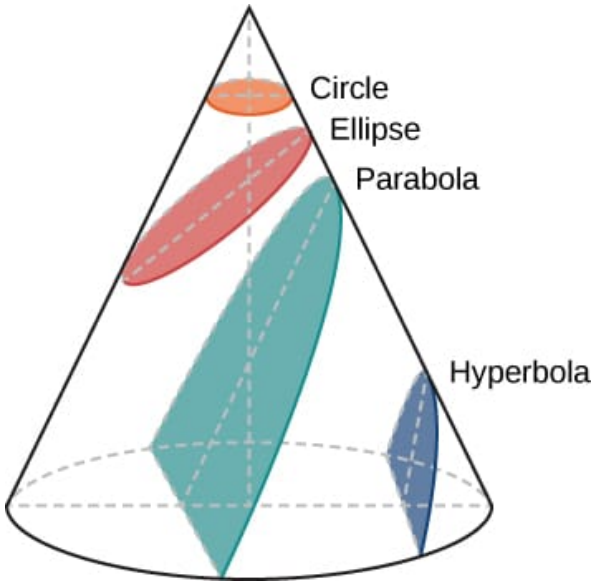


Figure 3.3 : Conic Sections. The circle, ellipse, parabola, and hyperbola are all formed by the intersection of a plane with a cone. This is why such curves are called conic sections.

You might recall from math classes that in a circle, the center is a special point. The distance from the center to anywhere on the circle is exactly the same. In an ellipse, the sum of the distance from two special points inside the ellipse to any point on the ellipse is always the same. These two points inside the ellipse are called its foci (singular: focus), a word invented for this purpose by Kepler.

This property suggests a simple way to draw an ellipse (Figure 3.4). We wrap the ends of a loop of string around two tacks pushed through a sheet of paper into a drawing board, so that the string is slack. If we push a pencil against the string, making the string taut, and then slide the pencil against the string all around the tacks, the curve that results is an ellipse. At any point where the pencil may be, the sum of the distances from the pencil to the two tacks is a constant length—the length of the string. The tacks are at the two foci of the ellipse.

The widest diameter of the ellipse is called its major axis. Half this distance—that is, the distance from the center of the ellipse to one end—is the semimajor axis, which is usually used to specify the size of the ellipse. For example, the semimajor axis of the orbit of Mars, which is also the planet's average distance from the Sun, is 228 million kilometers.

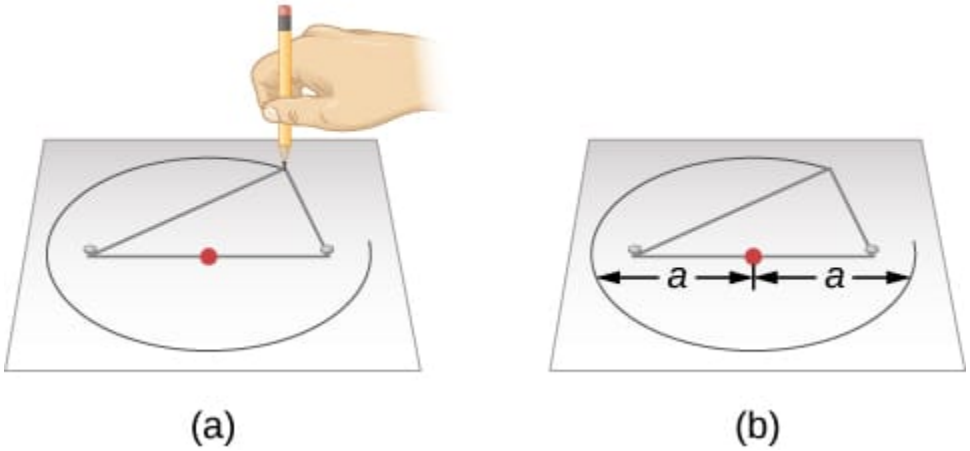


Figure 3.4 : Drawing an Ellipse. (a) We can construct an ellipse by pushing two tacks (the white objects) into a piece of paper on a drawing board, and then looping a string around the tacks. Each tack represents a focus of the ellipse, with one of the tacks being the Sun. Stretch the string tight using a pencil, and then move the pencil around the tacks. The length of the string remains the same, so that the sum of the distances from any point on the ellipse to the foci is always constant. (b) In this illustration, each semimajor axis is denoted by *a*. The distance *2a* is called the major axis of the ellipse.

The shape (roundness) of an ellipse depends on how close together the two foci are, compared with the major axis. The ratio of the distance between the foci to the length of the major axis is called the eccentricity of the ellipse.

If the foci (or tacks) are moved to the same location, then the distance between the foci would be zero. This means that the eccentricity is zero and the ellipse is just a circle; thus, a circle can be called an ellipse of zero eccentricity. In a circle, the semimajor axis would be the radius.

Next, we can make ellipses of various elongations (or extended lengths) by varying the spacing of the tacks (as long as they are not farther apart than the length of the string). The greater the eccentricity, the more elongated is the ellipse, up to a maximum eccentricity of 1.0, when the ellipse becomes "flat," the other extreme from a circle.

The size and shape of an ellipse are completely specified by its semimajor axis and its eccentricity. Using Brahe's data, Kepler found that Mars has an elliptical orbit, with the Sun at one focus (the other focus is empty). The eccentricity of the orbit of Mars is only about 0.1; its orbit, drawn to scale, would be practically indistinguishable from a circle, but the difference turned out to be critical for understanding planetary motions.

Kepler generalized this result in his first law and said that *the orbits of all the planets are ellipses*. Here was a decisive moment in the history of human thought: it was not necessary to have only circles in order to have an acceptable cosmos. The universe could be a bit more complex than the Greek philosophers had wanted it to be.

Kepler's second law deals with the speed with which each planet moves along its ellipse, also known as its orbital speed. Working with Brahe's observations of Mars, Kepler discovered that the planet speeds up as it comes closer to the Sun and slows down as it pulls away from the Sun. He expressed the precise form of this relationship by imagining that the Sun and Mars are connected by a straight, elastic line. When Mars is closer to the Sun (positions 1 and 2 in Figure 3.5), the elastic line is not stretched as much, and the planet moves rapidly. Farther from the Sun, as in positions 3 and 4, the line is stretched a lot, and the planet does not move so fast. As Mars travels in its elliptical orbit around the Sun, the elastic line sweeps out areas of the ellipse as it moves (the colored regions in our figure). Kepler found that in equal intervals of time (t), the areas swept out in space by this imaginary line are always equal; that is, the area of the region B from 1 to 2 is the same as that of region A from 3 to 4.

If a planet moves in a circular orbit, the elastic line is always stretched the same amount and the planet moves at a constant speed around its orbit. But, as Kepler discovered, in most orbits that speed of a planet orbiting its star (or moon orbiting its planet) tends to vary because the orbit is elliptical.

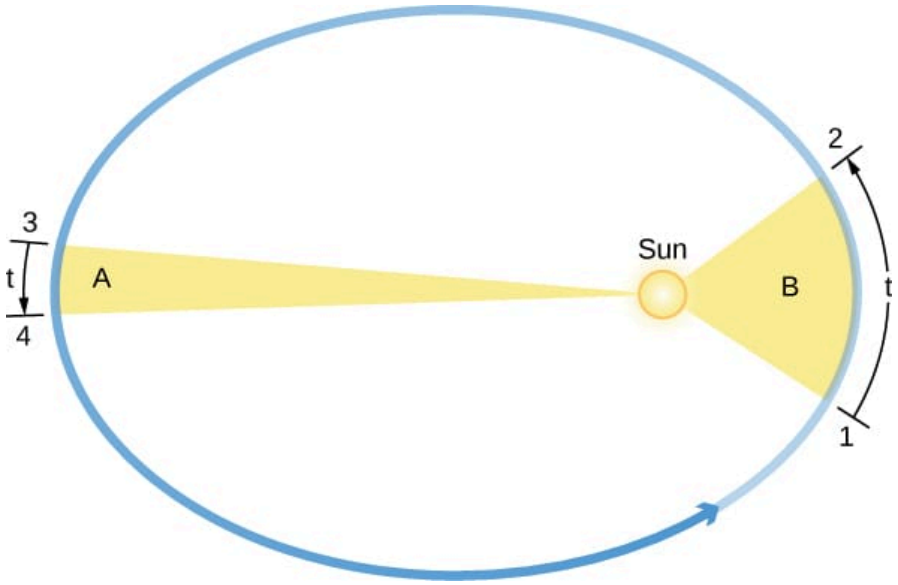


Figure 3.5 : Kepler's Second Law: The Law of Equal Areas. The orbital speed of a planet traveling around the Sun (the circular object inside the ellipse) varies in such a way that in equal intervals of time (t), a line between the Sun and a planet sweeps out equal areas (A and B). Note that the eccentricities of the planets' orbits in our solar system are substantially less than shown here.

Link to Learning

The Kepler's Second Law demonstrator from CCNY's ScienceSims project shows how an orbiting planet sweeps out the same area in the same time.

Kepler's Third Law

Kepler's first two laws of planetary motion describe the shape of a planet's orbit and allow us to calculate the speed of its motion at any point in the orbit. Kepler was pleased to have discovered such fundamental rules, but they did not satisfy his quest to fully understand planetary motions. He wanted to know why the orbits of the planets were spaced as they are and to find a mathematical pattern in their movements—a "harmony of the spheres" as he called it. For many years he worked to discover mathematical relationships governing planetary spacing and the time each planet took to go around the Sun.

In 1619, Kepler discovered a basic relationship to relate the planets' orbits to their relative distances from the Sun. We define a planet's orbital period, (***P***), as the time it takes a planet to travel once around the Sun. Also, recall that a planet's semimajor axis, *a,* is equal to its average distance from the Sun. The relationship, now known as *Kepler's third law*, says that a planet's orbital period squared is proportional to the semimajor axis of its orbit cubed, or

$$P^2 \propto a^3$$

When *P* (the orbital period) is measured in years, and *a* is expressed in a quantity known as an astronomical unit (AU), the two sides of the formula are not only proportional but equal. One AU is the average distance between Earth and the Sun and is approximately equal to 1.5×10^8 kilometers. In these units,

$$P^2 = a^3$$

Kepler's third law applies to all objects orbiting the Sun, including Earth, and provides a means for calculating their relative distances from the Sun from the time they take to orbit. Let's look at a specific example to illustrate how useful Kepler's third law is.

For instance, suppose you time how long Mars takes to go around the Sun (in Earth years). Kepler's third law can then be used to calculate Mars' average distance from the Sun. Mars' orbital period (1.88 Earth years) squared, or P^2, is $1.88^2 = 3.53$, and according to the equation for Kepler's third law, this equals the cube of its semimajor axis, or a^3. So what number must be cubed to give 3.53? The answer is $1.52 (\text{since} 1.52 \times 1.52 \times 1.52 = 3.53)$. Thus, Mars' semimajor axis in astronomical units must be 1.52 AU.

In other words, to go around the Sun in a little less than two years, Mars must be about 50% (half again) as far from the Sun as Earth is.

Example 3.1: Calculating Periods

Imagine an object is traveling around the Sun. What would be the orbital period of the object if its orbit has a semimajor axis of 50 AU?

Solution

From Kepler's third law, we know that (when we use units of years and AU)

$$P^2 = a^3$$

If the object's orbit has a semimajor axis of 50 AU (a = 50), we can cube 50 and then take the square root of the result to get P:

$$\begin{aligned} P &= \sqrt{a^3} \\ P &= \sqrt{50 \times 50 \times 50} = \sqrt{125{,}000} = 353.6 \text{ years} \end{aligned}$$

Therefore, the orbital period of the object is about 350 years. This would place our hypothetical object beyond the orbit of Pluto.

Exercise 3.4.1

What would be the orbital period of an asteroid (a rocky chunk between Mars and Jupiter) with a semimajor axis of 3 AU?

Answer available at the end of this section.

Kepler's three laws of planetary motion can be summarized as follows:

- Kepler's first law: Each planet moves around the Sun in an orbit that is an ellipse, with the Sun at one focus of the ellipse.
- Kepler's second law: The straight line joining a planet and the Sun sweeps out equal areas in space in equal intervals of time.
- Kepler's third law: The square of a planet's orbital period is directly proportional to the cube of the semimajor axis of its orbit.

Kepler's three laws provide a precise geometric description of planetary motion within the framework of the Copernican system. With these tools, it was possible to calculate planetary positions with greatly improved precision. Still, Kepler's laws are purely descriptive: they do not help us understand what forces of nature constrain the planets to follow this particular set of rules. That step was left to Isaac Newton.

Example 3.2: Applying Kepler's Third Law

Using the orbital periods and semimajor axes for Venus and Earth that are provided here, calculate p^2 and a^3, and verify that they obey Kepler's third law. Venus' orbital period is 0.62 year, and its semimajor axis is 0.72 AU. Earth's orbital period is 1.00 year, and its semimajor axis is 1.00 AU.

Solution

We can use the equation for Kepler's third law, $P^2 \propto a^3$. For Venus, $P^2 = 0.62 \times 0.62 = 0.38$ and $a^3 = 0.72 \times 0.72 \times 0.72 = .037$ (rounding numbers sometimes causes minor discrepancies like this). The square of the orbital period (0.38) approximates the cube of the semimajor axis (0.37). Therefore, Venus obeys Kepler's third law. For Earth, $P^2 = 1.00 \times 1.00 = 1.00$ and $a^3 = 1.00 \times 1.00 \times 1.00 = 1.00$. The square of the orbital period (1.00) approximates (in this case, equals) the cube of the semimajor axis (1.00). Therefore, Earth obeys Kepler's third law.

Exercise 3.4.1

Using the orbital periods and semimajor axes for Saturn and Jupiter that are provided here, calculate P^2 and a^3, and verify that they obey Kepler's third law. Saturn's orbital period is 29.46 years, and its semimajor axis is 9.54 AU. Jupiter's orbital period is 11.86 years, and its semimajor axis is 5.20 AU.

Answer available at the end of this section.

Link to Learning

In honor of the scientist who first devised the laws that govern the motions of planets, the team that built the first spacecraft to search for planets orbiting other stars decided to name the probe "Kepler." Visit NASA's Kepler website to learn more about Johannes Kepler's life and his laws of planetary motion. NASA's Kepler website and follow the links that interest you.

Answers

Exercise 3.4.1

$P = \sqrt{3 \times 3 \times 3} = \sqrt{27} = 5.2 \text{ years}$

Exercise 3.4.1

For Saturn, $P^2 = 29.46 \times 29.46 = 867.9$ and $a^3 = 9.54 \times 9.54 \times 9.54 = 868.3$. The square of the orbital period (867.9) approximates the cube of the semimajor axis (868.3). Therefore, Saturn obeys Kepler's third law.

This page titled 3.4: The Laws of Planetary Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 3.2: The Laws of Planetary Motion by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

3.5: The Nature of Science

The ultimate judge in science is always what nature itself reveals based on observations, experiments, models, and testing. Science is not merely a body of knowledge, but a *method* by which we attempt to understand nature and how it behaves. This method begins with many observations over a period of time. From the trends found through observations, scientists can *model* the particular phenomena we want to understand. Such models are always approximations of nature, subject to further testing.

As a concrete astronomical example, ancient astronomers constructed a model (partly from observations and partly from philosophical beliefs) that Earth was the center of the universe and everything moved around it in circular orbits. At first, our available observations of the Sun, Moon, and planets did fit this model; however, after further observations, the model had to be updated by adding circle after circle to represent the movements of the planets around Earth at the center. As the centuries passed and improved instruments were developed for keeping track of objects in the sky, the old model (even with a huge number of circles) could no longer explain all the observed facts. As we will see in the chapter on Observing the Sky: The Birth of Astronomy, a new model, with the Sun at the center, fit the experimental evidence better. After a period of philosophical struggle, it became accepted as our view of the universe.

When they are first proposed, new models or ideas are sometimes called *hypotheses*. You may think there can be no new hypotheses in a science such as astronomy—that everything important has already been learned. Nothing could be further from the truth. Throughout this textbook you will find discussions of recent, and occasionally still controversial, hypotheses in astronomy. For example, the significance that the huge chunks of rock and ice that hit Earth have for life on Earth itself is still debated. And while the evidence is strong that vast quantities of invisible “dark energy” make up the bulk of the universe, scientists have no convincing explanation for what the dark energy actually is. Resolving these issues will require difficult observations done at the forefront of our technology, and all such hypotheses need further testing before we incorporate them fully into our standard astronomical models.

This last point is crucial: a hypothesis must be a proposed explanation that can be *tested*. The most straightforward approach to such testing in science is to perform an experiment. If the experiment is conducted properly, its results either will agree with the predictions of the hypothesis or they will contradict it. If the experimental result is truly inconsistent with the hypothesis, a scientist must discard the hypothesis and try to develop an alternative. If the experimental result agrees with predictions, this does not necessarily prove that the hypothesis is absolutely correct; perhaps later experiments will contradict crucial parts of the hypothesis. But, the more experiments that agree with the hypothesis, the more likely we are to accept the hypothesis as a useful description of nature.

One way to think about this is to consider a scientist who was born and lives on an island where only black sheep live. Day after day the scientist encounters black sheep only, so he or she hypothesizes that all sheep are black. Although every observed sheep adds confidence to the hypothesis, the scientist only has to visit the mainland and observe one white sheep to prove the hypothesis wrong.

When you read about experiments, you probably have a mental picture of a scientist in a laboratory conducting tests or taking careful measurements. This is certainly the case for a biologist or a chemist, but what can astronomers do when our laboratory is the universe? It’s impossible to put a group of stars into a test tube or to order another comet from a scientific supply company.

As a result, astronomy is sometimes called an *observational* science; we often make our tests by observing many samples of the kind of object we want to study and noting carefully how different samples vary. New instruments and technology can let us look at astronomical objects from new perspectives and in greater detail. Our hypotheses are then judged in the light of this new information, and they pass or fail in the same way we would evaluate the result of a laboratory experiment.

Much of astronomy is also a *historical* science—meaning that what we observe has already happened in the universe and we can do nothing to change it. In the same way, a geologist cannot alter what has happened to our planet, and a paleontologist cannot bring an ancient animal back to life. While this can make astronomy challenging, it also gives us fascinating opportunities to discover the secrets of our cosmic past.

You might compare an astronomer to a detective trying to solve a crime that occurred before the detective arrived at the scene. There is lots of evidence, but both the detective and the scientist must sift through and organize the evidence to test various hypotheses about what actually happened. And there is another way in which the scientist is like a detective: they both must prove their case. The detective must convince the district attorney, the judge, and perhaps ultimately the jury that his hypothesis

is correct. Similarly, the scientist must convince colleagues, editors of journals, and ultimately a broad cross-section of other scientists that her hypothesis is provisionally correct. In both cases, one can only ask for evidence “beyond a reasonable doubt.” And sometimes new evidence will force both the detective and the scientist to revise their last hypothesis.

This self-correcting aspect of science sets it off from most human activities. Scientists spend a great deal of time questioning and challenging one another, which is why applications for project funding—as well as reports for publication in academic journals—go through an extensive process of *peer review*, which is a careful examination by other scientists in the same field. In science (after formal education and training), everyone is encouraged to improve upon experiments and to challenge any and all hypotheses. New scientists know that one of the best ways to advance their careers is to find a weakness in our current understanding of something and to correct it with a new or modified hypothesis.

This is one of the reasons science has made such dramatic progress. An undergraduate science major today knows more about science and math than did Sir Isaac Newton, one of the most renowned scientists who ever lived. Even in this introductory astronomy course, you will learn about objects and processes that no one a few generations ago even dreamed existed.

This page titled 3.5: The Nature of Science is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 1.3: The Nature of Science by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

4: MAKING SENSE OF THE UNIVERSE

4.1: Newton’s Great Synthesis
4.2: Newton’s Universal Law of Gravitation

This page titled 4: Making Sense of the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.1: Newton’s Great Synthesis

Learning Objectives

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

- Describe Newton’s three laws of motion
- Explain how Newton’s three laws of motion relate to momentum
- Define mass, volume, and density and how they differ
- Define angular momentum

It was the genius of Isaac Newton that found a conceptual framework that completely explained the observations and rules assembled by Galileo, Brahe, Kepler, and others. Newton was born in Lincolnshire, England, in the year after Galileo’s death (Figure 3.6). Against the advice of his mother, who wanted him to stay home and help with the family farm, he entered Trinity College at Cambridge in 1661 and eight years later was appointed professor of mathematics. Among Newton’s contemporaries in England were architect Christopher Wren, authors Aphra Behn and Daniel Defoe, and composer G. F. Handel.

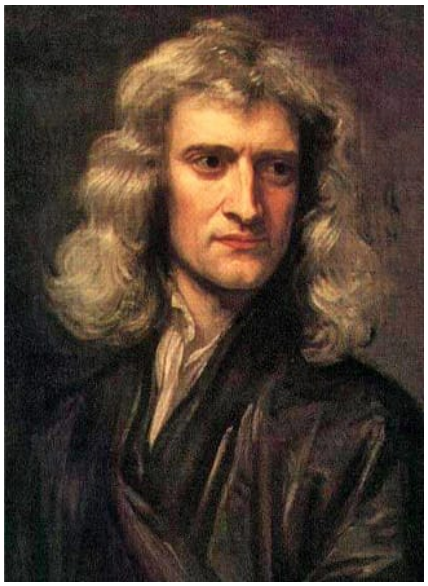


Figure 3.6 : Isaac Newton (1643–1727), 1689 Portrait by Sir Godfrey Kneller. Isaac Newton’s work on the laws of motion, gravity, optics, and mathematics laid the foundations for much of physical science.

Newton’s Laws of Motion

As a young man in college, Newton became interested in natural philosophy, as science was then called. He worked out some of his first ideas on machines and optics during the plague years of 1665 and 1666, when students were sent home from college. Newton, a moody and often difficult man, continued to work on his ideas in private, even inventing new mathematical tools to help him deal with the complexities involved. Eventually, his friend Edmund Halley (profiled in Comets and Asteroids: Debris of the Solar System) prevailed on him to collect and publish the results of his remarkable investigations on motion and gravity. The result was a volume that set out the underlying system of the physical world, *Philosophiae Naturalis Principia Mathematica*. The *Principia*, as the book is generally known, was published at Halley’s expense in 1687.

At the very beginning of the *Principia*, Newton proposes three laws that would govern the motions of all objects:

- Newton’s first law: Every object will continue to be in a state of rest or move at a constant speed in a straight line unless it is compelled to change by an outside force.
- Newton’s second law: The change of motion of a body is proportional to and in the direction of the force acting on it.

- Newton's third law: For every action there is an equal and opposite reaction (*or:* the mutual actions of two bodies upon each other are always equal and act in opposite directions).

In the original Latin, the three laws contain only 59 words, but those few words set the stage for modern science. Let us examine them more carefully.

Interpretation of Newton's Laws

Newton's first law is a restatement of one of Galileo's discoveries, called the *conservation of momentum*. The law states that in the absence of any outside influence, there is a measure of a body's motion, called its momentum, that remains unchanged. You may have heard the term momentum used in everyday expressions, such as "This bill in Congress has a lot of momentum; it's going to be hard to stop."

Newton's first law is sometimes called the *law of inertia*, where inertia is the tendency of objects (and legislatures) to keep doing what they are already doing. In other words, a stationary object stays put, and a moving object keeps moving unless some force intervenes.

Let's define the precise meaning of momentum—it depends on three factors: (1) speed—how fast a body moves (zero if it is stationary), (2) the direction of its motion, and (3) its mass—a measure of the amount of matter in a body, which we will discuss later. Scientists use the term velocity to describe the speed and direction of motion. For example, 20 kilometers per hour due south is velocity, whereas 20 kilometers per hour just by itself is speed. Momentum then can be defined as an object's mass times its velocity.

It's not so easy to see this rule in action in the everyday world because of the many forces acting on a body at any one time. One important force is friction, which generally slows things down. If you roll a ball along the sidewalk, it eventually comes to a stop because the sidewalk exerts a rubbing force on the ball. But in the space between the stars, where there is so little matter that friction is insignificant, objects can in fact continue to move (to coast) indefinitely.

The momentum of a body can change only under the action of an outside influence. Newton's second law expresses *force* in terms of its ability to change momentum with time. A force (a push or a pull) has both size and direction. When a force is applied to a body, the momentum changes in the direction of the applied force. This means that a force is required to change either the speed or the direction of a body, or both—that is, to start it moving, to speed it up, to slow it down, to stop it, or to change its direction.

As you learned in Observing the Sky: The Birth of Astronomy, the rate of change in an object's velocity is called *acceleration*. Newton showed that the acceleration of a body was proportional to the force being applied to it. Suppose that after a long period of reading, you push an astronomy book away from you on a long, smooth table. (We use a smooth table so we can ignore friction.) If you push the book steadily, it will continue to speed up as long as you are pushing it. The harder you push the book, the larger its acceleration will be. How much a force will accelerate an object is also determined by the object's mass. If you kept pushing a pen with the same force with which you pushed the textbook, the pen—having less mass—would be accelerated to a greater speed.

Newton's third law is perhaps the most profound of the rules he discovered. Basically, it is a generalization of the first law, but it also gives us a way to define mass. If we consider a system of two or more objects isolated from outside influences, Newton's first law says that the total momentum of the objects should remain constant. Therefore, any change of momentum within the system must be balanced by another change that is equal and opposite so that the momentum of the entire system is not changed.

This means that forces in nature do not occur alone: we find that in each situation there is always a *pair* of forces that are equal to and opposite each other. If a force is exerted on an object, it must be exerted by something else, and the object will exert an equal and opposite force back on that something. We can look at a simple example to demonstrate this.

Suppose that a daredevil astronomy student—and avid skateboarder—wants to jump from his second-story dorm window onto his board below (we don't recommend trying this!). The force pulling him down after jumping (as we will see in the next section) is the force of gravity between him and Earth. Both he and Earth must experience the same total change of momentum because of the influence of these mutual forces. So, both the student and Earth are accelerated by each other's pull. However, the student does much more of the moving. Because Earth has enormously greater mass, it can experience the same change of momentum by accelerating only a very small amount. Things fall toward Earth all the time, but the acceleration of our planet as a result is far too small to be measured.

A more obvious example of the mutual nature of forces between objects is familiar to all who have batted a baseball. The recoil you feel as you swing your bat shows that the ball exerts a force on it during the impact, just as the bat does on the ball. Similarly, when a rifle you are bracing on your shoulder is discharged, the force pushing the bullet out of the muzzle is equal to the force pushing backward upon the gun and your shoulder.

This is the principle behind jet engines and rockets: the force that discharges the exhaust gases from the rear of the rocket is accompanied by the force that pushes the rocket forward. The exhaust gases need not push against air or Earth; a rocket actually operates best in a vacuum (Figure 3.7).



Figure 3.7 : Demonstrating Newton's Third Law. The U.S. Space Shuttle (here launching *Discovery*), powered by three fuel engines burning liquid oxygen and liquid hydrogen, with two solid fuel boosters, demonstrates Newton's third law. (credit: modification of work by NASA)

Link to Learning

For more about Isaac Newton's life and work, check out this timeline page with snapshots from his career, produced by the British Broadcasting Corporation (BBC).

Mass, Volume, and Density

Before we go on to discuss Newton's other work, we want to take a brief look at some terms that will be important to sort out clearly. We begin with mass, which is a measure of the amount of material within an object.

The *volume* of an object is the measure of the physical space it occupies. Volume is measured in cubic units, such as cubic centimeters or liters. The volume is the "size" of an object. A penny and an inflated balloon may both have the same mass, but they have very different volumes. The reason is that they also have very different *densities*, which is a measure of how much mass there is per unit volume. Specifically, density is the mass divided by the volume. Note that in everyday language we often use "heavy" and "light" as indications of density (rather than weight) as, for instance, when we say that iron is heavy or that whipped cream is light.

The units of density that will be used in this book are grams per cubic centimeter (g/cm^3).[1] If a block of some material has a mass of 300 grams and a volume of 100 cm^3, its density is 3 g/cm^3. Familiar materials span a considerable range in density, from artificial materials such as plastic insulating foam (less than 0.1 g/cm^3) to gold (19.3 g/cm^3). Table 3.1 gives the densities of some familiar materials. In the astronomical universe, much more remarkable densities can be found, all the way from a comet's tail (10^{-16} g/cm^3) to a collapsed "star corpse" called a neutron star (10^{15} g/cm^3).

Table 3.1: Densities of Common Materials

Material	Density (g/cm^3)
Gold	19.3
Lead	11.3
Iron	7.9
Earth (bulk)	5.5
Rock (typical)	2.5
Water	1
Wood (typical)	0.8
Insulating foam	0.1
Silica gel	0.02

To sum up, mass is *how much*, volume is *how big*, and density is *how tightly packed*.

Link to Learning

You can play with a simple animation demonstrating the relationship between the concepts of density, mass, and volume, and find out why objects like wood float in water.

Angular Momentum

A concept that is a bit more complex, but important for understanding many astronomical objects, is angular momentum, which is a measure of the rotation of a body as it revolves around some fixed point (an example is a planet orbiting the Sun). The angular momentum of an object is defined as the product of its mass, its velocity, and its distance from the fixed point around which it revolves.

If these three quantities remain constant—that is, if the motion of a particular object takes place at a constant velocity at a fixed distance from the spin center—then the angular momentum is also a constant. Kepler's second law is a consequence of the *conservation of angular momentum*. As a planet approaches the Sun on its elliptical orbit and the distance to the spin center decreases, the planet speeds up to conserve the angular momentum. Similarly, when the planet is farther from the Sun, it moves more slowly.

The conservation of angular momentum is illustrated by figure skaters, who bring their arms and legs in to spin more rapidly, and extend their arms and legs to slow down (Figure 3.8). You can duplicate this yourself on a well-oiled swivel stool by starting yourself spinning slowly with your arms extended and then pulling your arms in. Another example of the conservation of angular momentum is a shrinking cloud of dust or a star collapsing on itself (both are situations that you will learn about as you read on). As material moves to a lesser distance from the spin center, the speed of the material increases to conserve angular momentum.

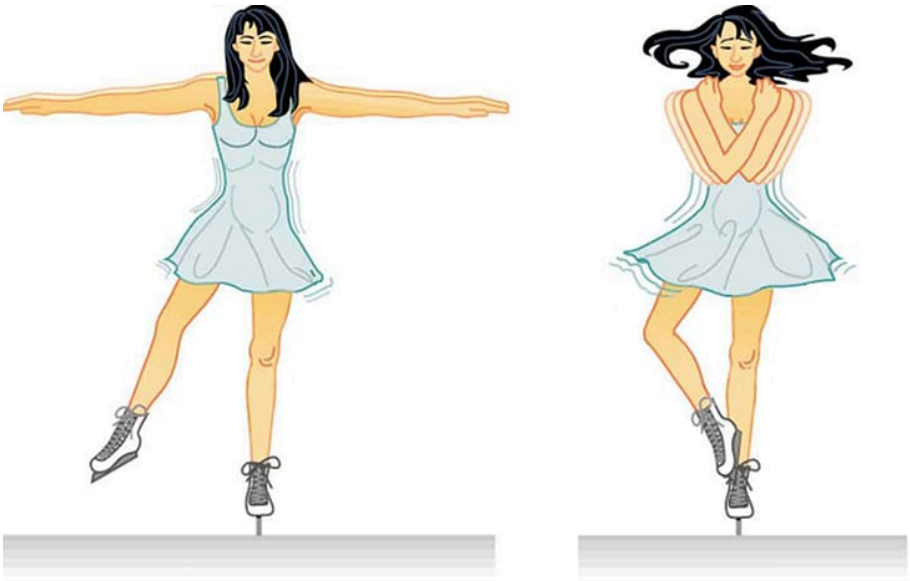


Figure 3.8 : Conservation of Angular Momentum. When a spinning figure skater brings in her arms, their distance from her spin center is smaller, so her speed increases. When her arms are out, their distance from the spin center is greater, so she slows down.

Footnotes

- 1Generally we use standard metric (or SI) units in this book. The proper metric unit of density in that system is kg/m^3. But to most people, g/cm^3 provides a more meaningful unit because the density of water is exactly 1 g/cm^3, and this is useful information for comparison. Density expressed in g/cm^3 is sometimes called specific density or specific weight.

This page titled 4.1: Newton's Great Synthesis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 3.3: Newton's Great Synthesis by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

4.2: Newton's Universal Law of Gravitation

Learning Objectives

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

- Explain what determines the strength of gravity
- Describe how Newton's universal law of gravitation extends our understanding of Kepler's laws

Newton's laws of motion show that objects at rest will stay at rest and those in motion will continue moving uniformly in a straight line unless acted upon by a force. Thus, it is the *straight line* that defines the most natural state of motion. But the planets move in ellipses, not straight lines; therefore, some force must be bending their paths. That force, Newton proposed, was gravity.

In Newton's time, gravity was something associated with Earth alone. Everyday experience shows us that Earth exerts a gravitational force upon objects at its surface. If you drop something, it accelerates toward Earth as it falls. Newton's insight was that Earth's gravity might extend as far as the Moon and produce the force required to curve the Moon's path from a straight line and keep it in its orbit. He further hypothesized that gravity is not limited to Earth, but that there is a general force of attraction between all material bodies. If so, the attractive force between the Sun and each of the planets could keep them in their orbits. (This may seem part of our everyday thinking today, but it was a remarkable insight in Newton's time.)

Once Newton boldly hypothesized that there was a universal attraction among all bodies everywhere in space, he had to determine the exact nature of the attraction. The precise mathematical description of that gravitational force had to dictate that the planets move exactly as Kepler had described them to (as expressed in Kepler's three laws). Also, that gravitational force had to predict the correct behavior of falling bodies on Earth, as observed by Galileo. How must the force of gravity depend on distance in order for these conditions to be met?

The answer to this question required mathematical tools that had not yet been developed, but this did not deter Isaac Newton, who invented what we today call calculus to deal with this problem. Eventually he was able to conclude that the magnitude of the force of gravity must decrease with increasing distance between the Sun and a planet (or between any two objects) in proportion to the inverse square of their separation. In other words, if a planet were twice as far from the Sun, the force would be $\left(\frac{1}{2}\right)^2$, or $\frac{1}{4}$ as large. Put the planet three times farther away, and the force is $\left(\frac{1}{3}\right)^2$, or $\frac{1}{9}$ as large.

Newton also concluded that the gravitational attraction between two bodies must be proportional to their masses. The more mass an object has, the stronger the pull of its gravitational force. The gravitational attraction between any two objects is therefore given by one of the most famous equations in all of science:

$$F_{\text{gravity}} =$$

where F_{gravity} is the gravitational force between two objects, M_1 and M_2 are the masses of the two objects, and R is their separation. G is a constant number known as the *universal gravitational constant*, and the equation itself symbolically summarizes Newton's *universal law of gravitation*. With such a force and the laws of motion, Newton was able to show mathematically that the only orbits permitted were exactly those described by Kepler's laws.

Newton's universal law of gravitation works for the planets, but is it really universal? The gravitational theory should also predict the observed acceleration of the Moon toward Earth as it orbits Earth, as well as of any object (say, an apple) dropped near Earth's surface. The falling of an apple is something we can measure quite easily, but can we use it to predict the motions of the Moon?

Recall that according to Newton's second law, forces cause acceleration. Newton's universal law of gravitation says that the force acting upon (and therefore the acceleration of) an object toward Earth should be inversely proportional to the square of its distance from the center of Earth. Objects like apples at the surface of Earth, at a distance of one Earth-radius from the center of Earth, are observed to accelerate downward at 9.8 meters per second per second $(9.8\text{m}/\text{s}^2)$.

It is this force of gravity on the surface of Earth that gives us our sense of *weight*. Unlike your mass, which would remain the same on any planet or moon, your weight depends on the local force of gravity. So you would weigh less on Mars and the Moon than on Earth, even though there is no change in your mass. (Which means you would still have to go easy on the desserts in the college

cafeteria when you got back!)

The Moon is 60 Earth radii away from the center of Earth. If gravity (and the acceleration it causes) gets weaker with distance squared, the acceleration the Moon experiences should be a lot less than for the apple. The acceleration should be $(1/60)^2 = 1/3600$ (or 3600 times less—about $0.00272\mathrm{m/s^2}$). This is precisely the observed acceleration of the Moon in its orbit. (As we shall see, the Moon does not fall *to* Earth with this acceleration, but falls *around* Earth.) Imagine the thrill Newton must have felt to realize he had discovered, and verified, a law that holds for Earth, apples, the Moon, and, as far as he knew, everything in the universe.

Example 3.3: Calculating Weight

By what factor would a person's weight at the surface of Earth change if Earth had its present mass but eight times its present volume?

Solution

With eight times the volume, Earth's radius would double. This means the gravitational force at the surface would reduce by a factor of $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$, so a person would weigh only one-fourth as much.

Exercise 4.2.1

By what factor would a person's weight at the surface of Earth change if Earth had its present size but only one-third its present mass?

Answer available at the end of this section.

Gravity is a "built-in" property of mass. Whenever there are masses in the universe, they will interact via the force of gravitational attraction. The more mass there is, the greater the force of attraction. Here on Earth, the largest concentration of mass is, of course, the planet we stand on, and its pull dominates the gravitational interactions we experience. But everything with mass attracts everything else with mass anywhere in the universe.

Newton's law also implies that gravity never becomes zero. It quickly gets weaker with distance, but it continues to act to some degree no matter how far away you get. The pull of the Sun is stronger at Mercury than at Pluto, but it can be felt far beyond Pluto, where astronomers have good evidence that it continuously makes enormous numbers of smaller icy bodies move around huge orbits. And the Sun's gravitational pull joins with the pull of billions of others stars to create the gravitational pull of our Milky Way Galaxy. That force, in turn, can make other smaller galaxies orbit around the Milky Way, and so on.

Why is it then, you may ask, that the astronauts aboard the Space Shuttle appear to have no gravitational forces acting on them when we see images on television of the astronauts and objects floating in the spacecraft? After all, the astronauts in the shuttle are only a few hundred kilometers above the surface of Earth, which is not a significant distance compared to the size of Earth, so gravity is certainly not a great deal weaker that much farther away. The astronauts feel "weightless" (meaning that they don't feel the gravitational force acting on them) for the same reason that passengers in an elevator whose cable has broken or in an airplane whose engines no longer work feel weightless: they are falling (Figure 3.9).[2]



Figure 3.9 : Astronauts in Free Fall. While in space, astronauts are falling freely, so they experience “weightlessness.” Clockwise from top left: Tracy Caldwell Dyson (NASA), Naoko Yamazaki (JAXA), Dorothy Metcalf-Lindenburger (NASA), and Stephanie Wilson (NASA). (credit: NASA)

When *falling*, they are in free fall and accelerate at the same rate as everything around them, including their spacecraft or a camera with which they are taking photographs of Earth. When doing so, astronauts experience no additional forces and therefore feel “weightless.” Unlike the falling elevator passengers, however, the astronauts are falling *around* Earth, not *to* Earth; as a result they will continue to fall and are said to be “in orbit” around Earth (see the next section for more about orbits).

Orbital Motion and Mass

Kepler’s laws describe the orbits of the objects whose motions are described by Newton’s laws of motion and the law of gravity. Knowing that gravity is the force that attracts planets toward the Sun, however, allowed Newton to rethink Kepler’s third law. Recall that Kepler had found a relationship between the orbital period of a planet’s revolution and its distance from the Sun. But Newton’s formulation introduces the additional factor of the masses of the Sun (M_1) and the planet (M_2), both expressed in units of the Sun’s mass. Newton’s universal law of gravitation can be used to show mathematically that this relationship is actually

$$a^3 = (M_1 + M_2) \times P^2$$

where *a* is the semimajor axis and *P* is the orbital period.

How did Kepler miss this factor? In units of the Sun’s mass, the mass of the Sun is 1, and in units of the Sun’s mass, the mass of a typical planet is a negligibly small factor. This means that the sum of the Sun’s mass and a planet’s mass, $(M_1 + M_2)$, is very, very close to 1. This makes Newton’s formula appear almost the same as Kepler’s; the tiny mass of the planets compared to the Sun is the reason that Kepler did not realize that both masses had to be included in the calculation. There are many situations in astronomy, however, in which we *do* need to include the two mass terms—for example, when two stars or two galaxies orbit each other.

Including the mass term allows us to use this formula in a new way. If we can measure the motions (distances and orbital periods) of objects acting under their mutual gravity, then the formula will permit us to deduce their masses. For example, we can calculate the mass of the Sun by using the distances and orbital periods of the planets, or the mass of Jupiter by noting the motions of its moons.

Indeed, Newton’s reformulation of Kepler’s third law is one of the most powerful concepts in astronomy. Our ability to deduce the masses of objects from their motions is key to understanding the nature and evolution of many astronomical bodies. We will use this law repeatedly throughout this text in calculations that range from the orbits of comets to the interactions of galaxies.

Example 3.4: Calculating the Effects of Gravity

A planet like Earth is found orbiting its star at a distance of 1 AU in 0.71 Earth-year. Can you use Newton's version of Kepler's third law to find the mass of the star? (Remember that compared to the mass of a star, the mass of an earthlike planet can be considered negligible.)

Solution

In the formula $a^3 = (M_1 + M_2) \times P^2$, the factor $M_1 + M_2$ would now be approximately equal to M_1 (the mass of the star), since the planet's mass is so small by comparison. Then the formula becomes $a^3 = M_1 \times P^2$, and we can solve for M[1]:

$$M_1 = \frac{a^3}{P^2}$$

Since $a = 1, a^3 = 1$, so

$$M_1 = \frac{1}{P^2} = \frac{1}{0.71^2} = \frac{1}{0.5} = 2$$

So the mass of the star is twice the mass of our Sun. (Remember that this way of expressing the law has units in terms of Earth and the Sun, so masses are expressed in units of the mass of our Sun.)

Exercise 4.2.1

Suppose a star with twice the mass of our Sun had an earthlike planet that took 4 years to orbit the star. At what distance (semimajor axis) would this planet orbit its star?

Answer available at the end of this section.

Link to Learning

You might like to try a simulation that lets you move the Sun, Earth, Moon, and space station to see the effects of changing their distances on their gravitational forces and orbital paths. You can even turn off gravity and see what happens.

Footnotes

- 2In the film *Apollo 13*, the scenes in which the astronauts were "weightless" were actually filmed in a falling airplane. As you might imagine, the plane fell for only short periods before the engines engaged again.

Answers

Exercise 4.2.1

With one-third its present mass, the gravitational force at the surface would reduce by a factor of $\frac{1}{3}$, so a person would weight only one-third as much.

Exercise 4.2.1

Again, we can neglect the mass of the planet. So $M_1 = 2$ and $P = 4$ years. The formula is $a^3 = M_1 \times P^2$, so $a^3 = 2 \times 4^2 = 2 \times 16 = 32$. So *a* is the cube root of 32. To find this, you can just ask Google, "What is the cube root of 32?" and get the answer 3.2 AU.

This page titled 4.2: Newton's Universal Law of Gravitation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 3.4: Newton's Universal Law of Gravitation by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5: LIGHT- MESSENGER OF THE COSMOS

5.1: The Behavior of Light

5.2: The Electromagnetic Spectrum

5.3: Spectroscopy in Astronomy

5.4: The Structure of the Atom

5.5: Formation of Spectral Lines

5.6: The Doppler Effect

5.7: Telescopes

5.8: Telescopes Today

5.9: Visible-Light Detectors and Instruments

5.10: Radio Telescopes

5.11: Observations outside Earth's Atmosphere

5.12: The Future of Large Telescopes

This page titled 5: Light- Messenger of the Cosmos is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.1: The Behavior of Light

Learning Objectives

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

- Explain the evidence for Maxwell's electromagnetic model of light
- Describe the relationship between wavelength, frequency, and speed of light
- Discuss the particle model of light and the definition of photon
- Explain how and why the amount of light we see from an object depends upon its distance

Coded into the light and other kinds of radiation that reach us from objects in the universe is a wide range of information about what those objects are like and how they work. If we can decipher this code and read the messages it contains, we can learn an enormous amount about the cosmos without ever having to leave Earth or its immediate environment.

The visible light and other radiation we receive from the stars and planets is generated by processes at the atomic level—by changes in the way the parts of an atom interact and move. Thus, to appreciate how light is generated, we must explore how atoms work. There is a bit of irony in the fact that in order to understand some of the largest structures in the universe, we must become acquainted with some of the smallest.

Notice that we have twice used the phrase "light and other radiation." One of the key ideas explored in this chapter is that visible light is not unique; it is merely the most familiar example of a much larger family of radiation that can carry information to us.

The word "radiation" will be used frequently in this book, so it is important to understand what it means. In everyday language, "radiation" is often used to describe certain kinds of energetic subatomic particles released by radioactive materials in our environment. (An example is the kind of radiation used to treat some cancers.) But this is not what we mean when we use the word "radiation" in an astronomy text. *Radiation*, as used in this book, is a general term for waves (including light waves) that *radiate* outward from a source.

As we saw in Orbits and Gravity, Newton's theory of gravity accounts for the motions of planets as well as objects on Earth. Application of this theory to a variety of problems dominated the work of scientists for nearly two centuries. In the nineteenth century, many physicists turned to the study of electricity and magnetism, which are intimately connected with the production of light.

The scientist who played a role in this field comparable to Newton's role in the study of gravity was physicist James Clerk Maxwell, born and educated in Scotland (Figure 5.2). Inspired by a number of ingenious experiments that showed an intimate relationship between electricity and magnetism, Maxwell developed a theory that describes both electricity and magnetism with only a small number of elegant equations. It is this theory that gives us important insights into the nature and behavior of light.



Figure 5.2 : James Clerk Maxwell (1831–1879). Maxwell unified the rules governing electricity and magnetism into a coherent theory.

Maxwell's Theory of Electromagnetism

We will look at the structure of the atom in more detail later, but we begin by noting that the typical atom consists of several types of particles, a number of which have not only mass but an additional property called electric charge. In the nucleus (central part) of every atom are *protons*, which are positively charged; outside the nucleus are electrons, which have a negative charge.

Maxwell's theory deals with these electric charges and their effects, especially when they are moving. In the vicinity of an electron charge, another charge feels a force of attraction or repulsion: opposite charges attract; like charges repel. When charges are not in motion, we observe only this electric attraction or repulsion. If charges are in motion, however (as they are inside every atom and in a wire carrying a current), then we measure another force called *magnetism*.

Magnetism was well known for much of recorded human history, but its cause was not understood until the nineteenth century. Experiments with electric charges demonstrated that magnetism was the result of moving charged particles. Sometimes, the motion is clear, as in the coils of heavy wire that make an industrial electromagnet. Other times, it is more subtle, as in the kind of magnet you buy in a hardware store, in which many of the electrons inside the atoms are spinning in roughly the same direction; it is the alignment of their motion that causes the material to become magnetic.

Physicists use the word *field* to describe the action of forces that one object exerts on other distant objects. For example, we say the Sun produces a *gravitational field* that controls Earth's orbit, even though the Sun and Earth do not come directly into contact. Using this terminology, we can say that stationary electric charges produce *electric fields*, and moving electric charges also produce *magnetic fields*.

Actually, the relationship between electric and magnetic phenomena is even more profound. Experiments showed that changing magnetic fields could produce electric currents (and thus changing electric fields), and changing electric currents could in turn produce changing magnetic fields. So once begun, electric and magnetic field changes could continue to trigger each other.

Maxwell analyzed what would happen if electric charges were oscillating (moving constantly back and forth) and found that the resulting pattern of electric and magnetic fields would spread out and travel rapidly through space. Something similar happens when a raindrop strikes the surface of water or a frog jumps into a pond. The disturbance moves outward and creates a pattern we call a *wave* in the water (Figure 5.3). You might, at first, think that there must be very few situations in nature where electric charges oscillate, but this is not at all the case. As we shall see, atoms and molecules (which consist of charged particles) oscillate back and forth all the time. The resulting electromagnetic disturbances are among the most common phenomena in the universe.



Figure 5.3 : Making Waves. An oscillation in a pool of water creates an expanding disturbance called a wave. (credit: modification of work by "vastateparksstaff"/Flickr)

Maxwell was able to calculate the speed at which an electromagnetic disturbance moves through space; he found that it is equal to the speed of light, which had been measured experimentally. On that basis, he speculated that light was one form of a family of possible electromagnetic disturbances called electromagnetic radiation, a conclusion that was again confirmed in laboratory experiments. When light (reflected from the pages of an astronomy textbook, for example) enters a human eye, its changing electric and magnetic fields stimulate nerve endings, which then transmit the information contained in these changing fields to the brain. The science of astronomy is primarily about analyzing radiation from distant objects to understand what they are and how they work.

The Wave-Like Characteristics of Light

The changing electric and magnetic fields in light are similar to the waves that can be set up in a quiet pool of water. In both cases, the disturbance travels rapidly outward from the point of origin and can use its energy to disturb other things farther away. (For example, in water, the expanding ripples moving away from our frog could disturb the peace of a dragonfly resting on a leaf in the same pool.) In the case of electromagnetic waves, the radiation generated by a transmitting antenna full of charged particles and moving electrons at your local radio station can, sometime later, disturb a group of electrons in your car radio antenna and bring you the news and weather while you are driving to class or work in the morning.

The waves generated by charged particles differ from water waves in some profound ways, however. Water waves require water to travel in. The sound waves we hear, to give another example, are pressure disturbances that require air to travel though. But electromagnetic waves do not require water or air: the fields generate each other and so can move through a vacuum (such as outer space). This was such a disturbing idea to nineteenth-century scientists that they actually made up a substance to fill all of space—one for which there was not a single shred of evidence—just so light waves could have something to travel through: they called it the *aether*. Today, we know that there is no aether and that electromagnetic waves have no trouble at all moving through empty space (as all the starlight visible on a clear night must surely be doing).

The other difference is that *all* electromagnetic waves move at the same speed in empty space (the speed of light—approximately 300,000 kilometers per second, or 300,000,000 meters per second, which can also be written as $3 \times 10^8 \mathrm{m}/$), which turns out to be the fastest possible speed in the universe. No matter where electromagnetic waves are generated from and no matter what other properties they have, when they are moving (and not interacting with matter), they move at the speed of light. Yet you know from everyday experience that there are different kinds of light. For example, we perceive that light waves differ from one another in a property we call color. Let's see how we can denote the differences among the whole broad family of electromagnetic waves.

The nice thing about a wave is that it is a repeating phenomenon. Whether it is the up-and-down motion of a water wave or the changing electric and magnetic fields in a wave of light, the pattern of disturbance repeats in a cyclical way. Thus, any wave motion can be characterized by a series of crests and troughs (Figure 5.4). Moving from one crest through a trough to the next crest completes one cycle. The horizontal length covered by one cycle is called the wavelength. Ocean waves provide an analogy:

the wavelength is the distance that separates successive wave crests.

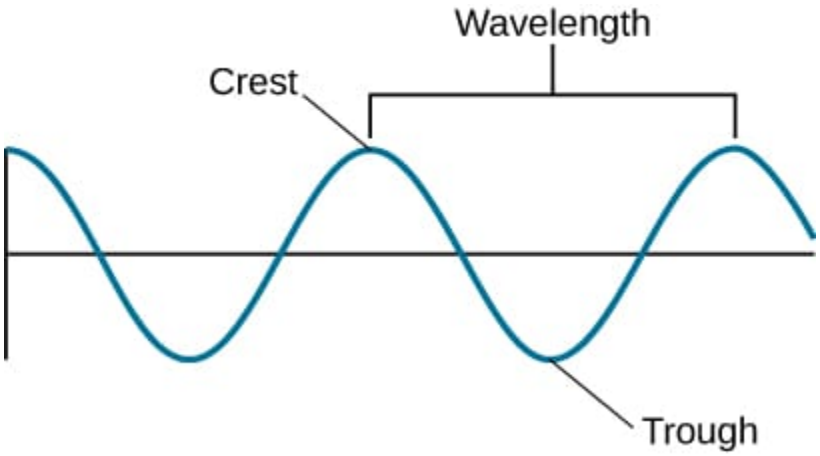


Figure 5.4 : Characterizing Waves. Electromagnetic radiation has wave-like characteristics. The wavelength (λ) is the distance between crests, the frequency (*f*) is the number of cycles per second, and the speed (*c*) is the distance the wave covers during a specified period of time (e.g., kilometers per second).

For visible light, our eyes perceive different wavelengths as different colors: red, for example, is the longest visible wavelength, and violet is the shortest. The main colors of visible light from longest to shortest wavelength can be remembered using the mnemonic ROY G BIV—for Red, Orange, Yellow, Green, Blue, Indigo, and Violet. Other invisible forms of electromagnetic radiation have different wavelengths, as we will see in the next section.

We can also characterize different waves by their frequency, the number of wave cycles that pass by per second. If you count 10 crests moving by each second, for example, then the frequency is 10 cycles per second (cps). In honor of Heinrich Hertz, the physicist who—inspired by Maxwell's work—discovered radio waves, a cps is also called a *hertz* (Hz). Take a look at your radio, for example, and you will see the channel assigned to each radio station is characterized by its frequency, usually in units of KHz (kilohertz, or thousands of hertz) or MHz (megahertz, or millions of hertz).

Wavelength (λ) and frequency (*f*) are related because all electromagnetic waves travel at the same speed. To see how this works, imagine a parade in which everyone is forced by prevailing traffic conditions to move at exactly the same speed. You stand on a corner and watch the waves of marchers come by. First you see row after row of miniature ponies. Because they are not very large and, therefore, have a shorter wavelength, a good number of the ponies can move past you each minute; we can say they have a high frequency. Next, however, come several rows of circus elephants. The elephants are large and marching at the same speed as the ponies, so far fewer of them can march past you per minute: Because they have a wider spacing (longer wavelength), they represent a lower frequency.

The formula for this relationship can be expressed as follows: for any wave motion, the speed at which a wave moves equals the frequency times the wavelength. Waves with longer wavelengths have lower frequencies. Mathematically, we can express this as

$$c = \lambda f$$

where the Greek letter for "l"—lambda, λ—is used to denote wavelength and *c* is the scientific symbol for the speed of light. Solving for the wavelength, this is expressed as:

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

Example 5.1: Deriving and Using the Wave Equation

The equation for the relationship between the speed and other characteristics of a wave can be derived from our basic understanding of motion. The average speed of anything that is moving is:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}$$

(So, for example, a car on the highway traveling at a speed of 100 km/h covers 100 km during the time of 1 h.) For an electromagnetic wave to travel the distance of one of its wavelengths, λ, at the speed of light, *c*, we have $c = \frac{\lambda}{t}$. The frequency of a wave is the number of cycles per second. If a wave has a frequency of a million cycles per second, then the time for each cycle to go by is a millionth of a second. So, in general, $t = \frac{1}{f}$. Substituting into our wave equation, we get $c = \lambda \times f$. Now let's use this to calculate an example. What is the wavelength of visible light that has a frequency of $5.66 \times 10^{14}\text{Hz}$?

Solution

Solving the wave equation for wavelength, we find:

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

Substituting our values gives:

$$\lambda = \frac{3.00 \times 10^{8}\,\text{m/s}}{5.66 \times 10^{14}\,\text{Hz}} = 5.30 \times 10^{-7}\,\text{m}$$

This answer can also be written as 530 nm, which is in the yellow-green part of the visible spectrum (nm stands for nanometers, where the term "nano" means "billionths").

Exercise 5.1.1

"Tidal waves," or tsunamis, are waves caused by earthquakes that travel rapidly through the ocean. If a tsunami travels at the speed of 600 km/h and approaches a shore at a rate of one wave crest every 15 min (4 waves/h), what would be the distance between those wave crests at sea?

Answer available at the end of this section.

Light as a Photon

The electromagnetic wave model of light (as formulated by Maxwell) was one of the great triumphs of nineteenth-century science. In 1887, when Heinrich Hertz actually made invisible electromagnetic waves (what today are called radio waves) on one side of a room and detected them on the other side, it ushered in a new era that led to the modern age of telecommunications. His experiment ultimately led to the technologies of television, cell phones, and today's wireless networks around the globe.

However, by the beginning of the twentieth century, more sophisticated experiments had revealed that light behaves in certain ways that cannot be explained by the wave model. Reluctantly, physicists had to accept that sometimes light behaves more like a "particle"—or at least a self-contained packet of energy—than a wave. We call such a packet of electromagnetic energy a photon.

The fact that light behaves like a wave in certain experiments and like a particle in others was a very surprising and unlikely idea. After all, our common sense says that waves and particles are opposite concepts. On one hand, a wave is a repeating disturbance that, by its very nature, is not in only one place, but spreads out. A particle, on the other hand, is something that can be in only one place at any given time. Strange as it sounds, though, countless experiments now confirm that electromagnetic radiation can sometimes behave like a wave and at other times like a particle.

Then, again, perhaps we shouldn't be surprised that something that always travels at the "speed limit" of the universe and doesn't need a medium to travel through might not obey our everyday common sense ideas. The confusion that this wave-particle duality of light caused in physics was eventually resolved by the introduction of a more complicated theory of waves and particles, now called quantum mechanics. (This is one of the most interesting fields of modern science, but it is mostly beyond the scope of our book. If you are interested in it, see some of the suggested resources at the end of this chapter.)

In any case, you should now be prepared when scientists (or the authors of this book) sometimes discuss electromagnetic radiation as if it consisted of waves and at other times refer to it as a stream of photons. A photon (being a packet of energy) carries a specific amount of energy. We can use the idea of energy to connect the photon and wave models. How much energy a photon has depends on its frequency when you think about it as a wave. A low-energy radio wave has a low frequency as a wave, while a high-energy X-ray at your dentist's office is a high-frequency wave. Among the colors of visible light, violet-light photons

have the highest energy and red-light photons have the lowest.

Test whether the connection between photons and waves is clear to you. In the above example, which photon would have the longer wavelength as a wave: the radio wave or the X-ray? If you answered the radio wave, you are correct. Radio waves have a lower frequency, so the wave cycles are longer (they are elephants, not miniature ponies).

Propagation of Light

Let's think for a moment about how light from a lightbulb moves through space. As waves expand, they travel away from the bulb, not just toward your eyes but in all directions. They must therefore cover an ever-widening space. Yet the total amount of light available can't change once the light has left the bulb. This means that, as the same expanding shell of light covers a larger and larger area, there must be less and less of it in any given place. Light (and all other electromagnetic radiation) gets weaker and weaker as it gets farther from its source.

The increase in the area that the light must cover is proportional to the square of the distance that the light has traveled (Figure 5.5). If we stand twice as far from the source, our eyes will intercept two-squared (2 × 2), or four times less light. If we stand 10 times farther from the source, we get 10-squared, or 100 times less light. You can see how this weakening means trouble for sources of light at astronomical distances. One of the nearest stars, Alpha Centauri A, emits about the same total energy as the Sun. But it is about 270,000 times farther away, and so it appears about 73 billion times fainter. No wonder the stars, which close-up would look more or less like the Sun, look like faint pinpoints of light from far away.

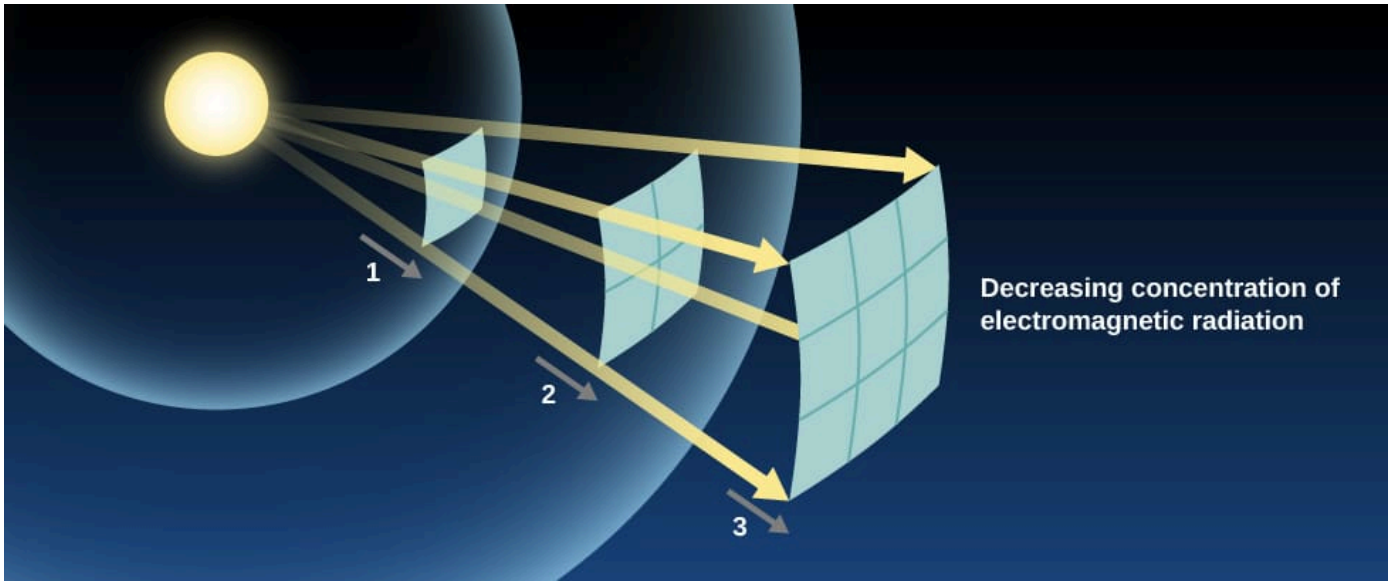


Figure 5.5 : Inverse Square Law for Light. As light radiates away from its source, it spreads out in such a way that the energy per unit area (the amount of energy passing through one of the small squares) decreases as the square of the distance from its source.

This idea—that the apparent brightness of a source (how bright it looks to us) gets weaker with distance in the way we have described—is known as the inverse square law for light propagation. In this respect, the propagation of light is similar to the effects of gravity. Remember that the force of gravity between two attracting masses is also inversely proportional to the square of their separation.

Example 5.2: The Inverse Square Law for Light

The intensity of a 120-W lightbulb observed from a distance 2 m away is 2.4 W/m^2. What would be the intensity if this distance was doubled?

Solution

If we move twice as far away, then the answer will change according to the inverse square of the distance, so the new intensity will be $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$ of the original intensity, or 0.6 W/m^2.

Exercise 5.1.1

How many times brighter or fainter would a star appear if it were moved to:

a. twice its present distance?
b. ten times its present distance?
c. half its present distance?

Answer available at the end of this section.

Answers

Exercise 5.1.1

$\lambda = \frac{600 \text{ km/h}}{4 \text{ waves/h}} = 150 \text{ km}$

Exercise 5.1.1

a. $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$; b. $\left(\frac{1}{10}\right)^2 = \frac{1}{100}$; c. $\left(\frac{1}{1/2}\right)^2 = 4$

This page titled 5.1: The Behavior of Light is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.2: The Behavior of Light by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.2: The Electromagnetic Spectrum

Learning Objectives

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

- Understand the bands of the electromagnetic spectrum and how they differ from one another
- Understand how each part of the spectrum interacts with Earth's atmosphere
- Explain how and why the light emitted by an object depends on its temperature

Objects in the universe send out an enormous range of electromagnetic radiation. Scientists call this range the electromagnetic spectrum, which they have divided into a number of categories. The spectrum is shown in Figure 5.6, with some information about the waves in each part or band.

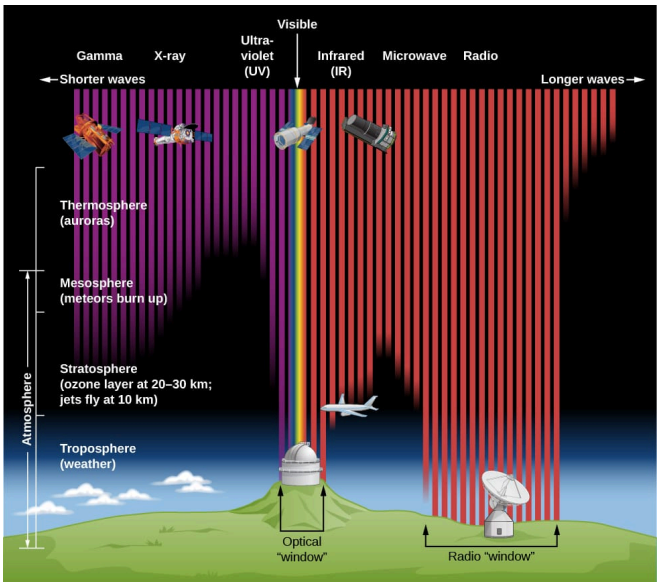


Figure 5.6 : Radiation and Earth's Atmosphere. This figure shows the bands of the electromagnetic spectrum and how well Earth's atmosphere transmits them. Note that high-frequency waves from space do not make it to the surface and must therefore be observed from space. Some infrared and microwaves are absorbed by water and thus are best observed from high altitudes. Low-frequency radio waves are blocked by Earth's ionosphere. (credit: modification of work by STScI/JHU/NASA)

Types of Electromagnetic Radiation

Electromagnetic radiation with the shortest wavelengths, no longer than 0.01 nanometer, is categorized as gamma rays (1 nanometer = 10^{-9} meters; see Appendix D). The name *gamma* comes from the third letter of the Greek alphabet: gamma rays were the third kind of radiation discovered coming from radioactive atoms when physicists first investigated their behavior. Because gamma rays carry a lot of energy, they can be dangerous for living tissues. Gamma radiation is generated deep in the interior of stars, as well as by some of the most violent phenomena in the universe, such as the deaths of stars and the merging of stellar corpses. Gamma rays coming to Earth are absorbed by our atmosphere before they reach the ground (which is a good thing for our health); thus, they can only be studied using instruments in space.

Electromagnetic radiation with wavelengths between 0.01 nanometer and 20 nanometers is referred to as X-rays. Being more energetic than visible light, X-rays are able to penetrate soft tissues but not bones, and so allow us to make images of the shadows of the bones inside us. While X-rays can penetrate a short length of human flesh, they are stopped by the large numbers of atoms in Earth's atmosphere with which they interact. Thus, X-ray astronomy (like gamma-ray astronomy) could not develop until we invented ways of sending instruments above our atmosphere (Figure 5.7).

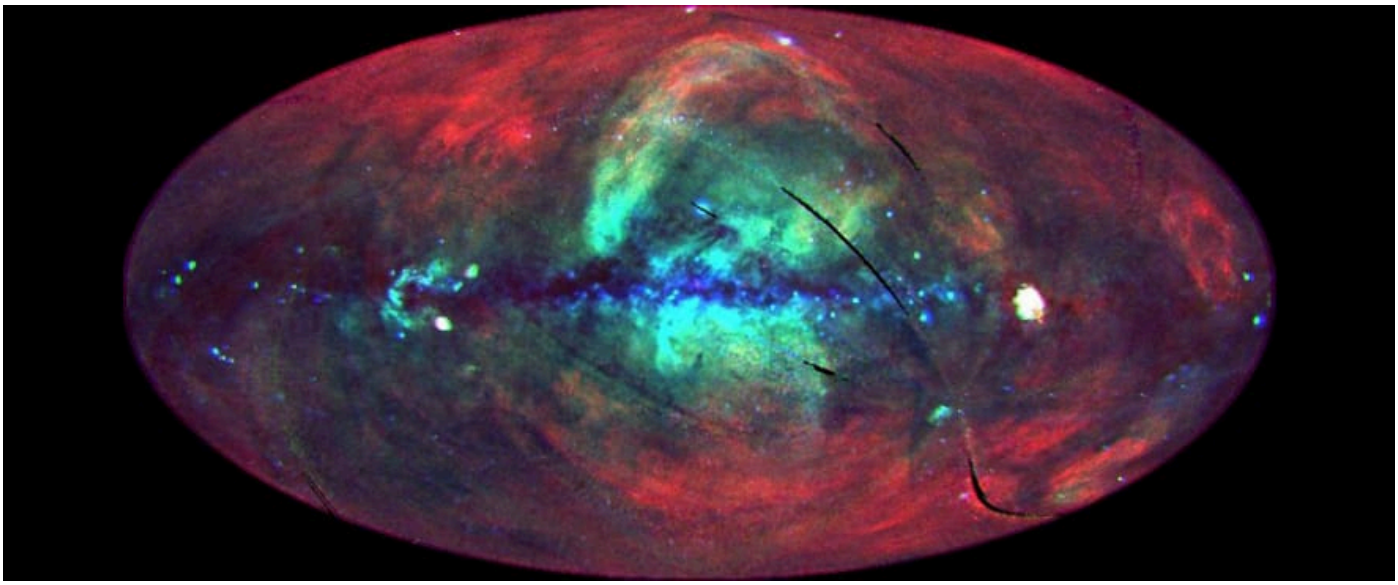


Figure 5.7 : X-Ray Sky. This is a map of the sky tuned to certain types of X-rays (seen from above Earth's atmosphere). The map tilts the sky so that the disk of our Milky Way Galaxy runs across its center. It was constructed and artificially colored from data gathered by the European ROSAT satellite. Each color (red, yellow, and blue) shows X-rays of different frequencies or energies. For example, red outlines the glow from a hot local bubble of gas all around us, blown by one or more exploding stars in our cosmic vicinity. Yellow and blue show more distant sources of X-rays, such as remnants of other exploded stars or the active center of our Galaxy (in the middle of the picture). (credit: modification of work by NASA)

Radiation intermediate between X-rays and visible light is ultraviolet (meaning higher energy than violet). Outside the world of science, ultraviolet light is sometimes called "black light" because our eyes cannot see it. Ultraviolet radiation is mostly blocked by the ozone layer of Earth's atmosphere, but a small fraction of ultraviolet rays from our Sun do penetrate to cause sunburn or, in extreme cases of overexposure, skin cancer in human beings. Ultraviolet astronomy is also best done from space.

Electromagnetic radiation with wavelengths between roughly 400 and 700 nm is called visible light because these are the waves that human vision can perceive. This is also the band of the electromagnetic spectrum that most readily reaches Earth's surface. These two observations are not coincidental: human eyes evolved to see the kinds of waves that arrive from the Sun most effectively. Visible light penetrates Earth's atmosphere effectively, except when it is temporarily blocked by clouds.

Between visible light and radio waves are the wavelengths of infrared or heat radiation. Astronomer William Herschel first discovered infrared in 1800 while trying to measure the temperatures of different colors of sunlight spread out into a spectrum. He noticed that when he accidently positioned his thermometer beyond the reddest color, it still registered heating due to some invisible energy coming from the Sun. This was the first hint about the existence of the other (invisible) bands of the electromagnetic spectrum, although it would take many decades for our full understanding to develop.

A heat lamp radiates mostly infrared radiation, and the nerve endings in our skin are sensitive to this band of the electromagnetic spectrum. Infrared waves are absorbed by water and carbon dioxide molecules, which are more concentrated low in Earth's atmosphere. For this reason, infrared astronomy is best done from high mountaintops, high-flying airplanes, and spacecraft.

After infrared comes the familiar microwave, used in short-wave communication and microwave ovens. (Wavelengths vary from 1 millimeter to 1 meter and are absorbed by water vapor, which makes them effective in heating foods.) The "micro-" prefix refers to the fact that microwaves are small in comparison to radio waves, the next on the spectrum. You may remember that tea—which is full of water—heats up quickly in your microwave oven, while a ceramic cup—from which water has been removed by baking—stays cool in comparison.

All electromagnetic waves longer than microwaves are called radio waves, but this is so broad a category that we generally divide it into several subsections. Among the most familiar of these are radar waves, which are used in radar guns by traffic officers to determine vehicle speeds, and AM radio waves, which were the first to be developed for broadcasting. The wavelengths of these different categories range from over a meter to hundreds of meters, and other radio radiation can have wavelengths as long as several kilometers.

With such a wide range of wavelengths, not all radio waves interact with Earth's atmosphere in the same way. FM and TV waves

are not absorbed and can travel easily through our atmosphere. AM radio waves are absorbed or reflected by a layer in Earth's atmosphere called the ionosphere (the ionosphere is a layer of charged particles at the top of our atmosphere, produced by interactions with sunlight and charged particles that are ejected from the Sun).

We hope this brief survey has left you with one strong impression: although visible light is what most people associate with astronomy, the light that our eyes can see is only a tiny fraction of the broad range of waves generated in the universe. Today, we understand that judging some astronomical phenomenon by using only the light we can see is like hiding under the table at a big dinner party and judging all the guests by nothing but their shoes. There's a lot more to each person than meets our eye under the table. It is very important for those who study astronomy today to avoid being "visible light chauvinists"—to respect only the information seen by their eyes while ignoring the information gathered by instruments sensitive to other bands of the electromagnetic spectrum.

Table 5.1 summarizes the bands of the electromagnetic spectrum and indicates the temperatures and typical astronomical objects that emit each kind of electromagnetic radiation. While at first, some of the types of radiation listed in the table may seem unfamiliar, you will get to know them better as your astronomy course continues. You can return to this table as you learn more about the types of objects astronomers study.

Table 5.1: Types of Electromagnetic Radiation

Type of Radiation	Wavelength Range (nm)	Radiated by Objects at This Temperature	Typical Sources
Gamma rays	Less than 0.01	More than 10^8 K	Produced in nuclear reactions; require very high-energy processes
X-rays	0.01–20	10^6–10^8 K	Gas in clusters of galaxies, supernova remnants, solar corona
Ultraviolet	20–400	10^4–10^6 K	Supernova remnants, very hot stars
Visible	400–700	10^3–10^4 K	Stars
Infrared	10^3–10^6	10–10^3 K	Cool clouds of dust and gas, planets, moons
Microwave	10^6–10^9	Less than 10 K	Active galaxies, pulsars, cosmic background radiation
Radio	More than 10^9	Less than 10 K	Supernova remnants, pulsars, cold gas

Radiation and Temperature

Some astronomical objects emit mostly infrared radiation, others mostly visible light, and still others mostly ultraviolet radiation. What determines the type of electromagnetic radiation emitted by the Sun, stars, and other dense astronomical objects? The answer often turns out to be their *temperature*.

At the microscopic level, everything in nature is in motion. A solid is composed of molecules and atoms in continuous vibration: they move back and forth in place, but their motion is much too small for our eyes to make out. A gas consists of atoms and/or molecules that are flying about freely at high speed, continually bumping into one another and bombarding the surrounding matter. The hotter the solid or gas, the more rapid the motion of its molecules or atoms. The temperature of something is thus a measure of the average motion energy of the particles that make it up.

This motion at the microscopic level is responsible for much of the electromagnetic radiation on Earth and in the universe. As atoms and molecules move about and collide, or vibrate in place, their electrons give off electromagnetic radiation. The characteristics of this radiation are determined by the temperature of those atoms and molecules. In a hot material, for example, the individual particles vibrate in place or move rapidly from collisions, so the emitted waves are, on average, more energetic. And recall that higher energy waves have a higher frequency. In very cool material, the particles have low-energy atomic and molecular motions and thus generate lower-energy waves.

Link to Learning

Check out the NASA briefing or NASA's 5-minute introductory video to learn more about the electromagnetic spectrum.

Radiation Laws

To understand, in more quantitative detail, the relationship between temperature and electromagnetic radiation, we imagine an idealized object called a blackbody. Such an object (unlike your sweater or your astronomy instructor's head) does not reflect or scatter any radiation, but absorbs all the electromagnetic energy that falls onto it. The energy that is absorbed causes the atoms and molecules in it to vibrate or move around at increasing speeds. As it gets hotter, this object will radiate electromagnetic waves until absorption and radiation are in balance. We want to discuss such an idealized object because, as you will see, stars behave in very nearly the same way.

The radiation from a blackbody has several characteristics, as illustrated in Figure 5.8. The graph shows the power emitted at each wavelength by objects of different temperatures. In science, the word *power* means the energy coming off per second (and it is typically measured in *watts*, which you are probably familiar with from buying lightbulbs).

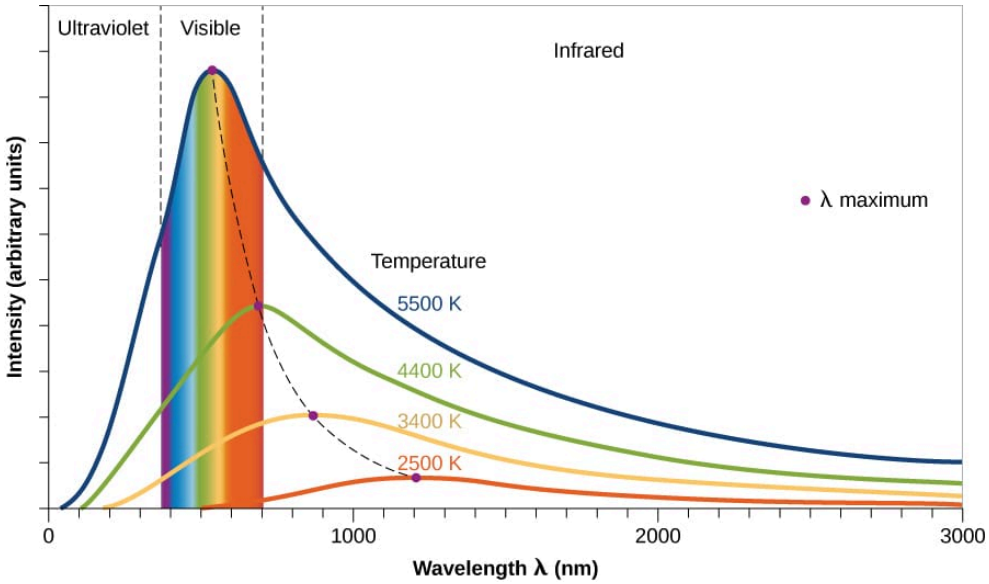


Figure 5.8 : Radiation Laws Illustrated. This graph shows in arbitrary units how many photons are given off at each wavelength for objects at four different temperatures. The wavelengths corresponding to visible light are shown by the colored bands. Note that at hotter temperatures, more energy (in the form of photons) is emitted at all wavelengths. The higher the temperature, the shorter the wavelength at which the peak amount of energy is radiated (this is known as Wien's law).

First of all, notice that the curves show that, at each temperature, our blackbody object emits radiation (photons) at all wavelengths (all colors). This is because in any solid or denser gas, some molecules or atoms vibrate or move between collisions slower than average and some move faster than average. So when we look at the electromagnetic waves emitted, we find a broad range, or spectrum, of energies and wavelengths. More energy is emitted at the average vibration or motion rate (the highest part of each curve), but if we have a large number of atoms or molecules, some energy will be detected at each wavelength.

Second, note that an object at a higher temperature emits more power at all wavelengths than does a cooler one. In a hot gas (the taller curves in Figure 5.8), for example, the atoms have more collisions and give off more energy. In the real world of stars, this means that hotter stars give off more energy at every wavelength than do cooler stars.

Third, the graph shows us that the higher the temperature, the shorter the wavelength at which the maximum power is emitted. Remember that a shorter wavelength means a higher frequency and energy. It makes sense, then, that hot objects give off a larger fraction of their energy at shorter wavelengths (higher energies) than do cool objects. You may have observed examples of this rule in everyday life. When a burner on an electric stove is turned on low, it emits only heat, which is infrared radiation, but does not glow with visible light. If the burner is set to a higher temperature, it starts to glow a dull red. At a still-higher setting, it glows a brighter orange-red (shorter wavelength). At even higher temperatures, which cannot be reached with ordinary stoves,

metal can appear brilliant yellow or even blue-white.

We can use these ideas to come up with a rough sort of “thermometer” for measuring the temperatures of stars. Because many stars give off most of their energy in visible light, the color of light that dominates a star’s appearance is a rough indicator of its temperature. If one star looks red and another looks blue, which one has the higher temperature? Because blue is the shorter-wavelength color, it is the sign of a hotter star. (Note that the temperatures we associate with different colors in science are not the same as the ones artists use. In art, red is often called a “hot” color and blue a “cool” color. Likewise, we commonly see red on faucet or air conditioning controls to indicate hot temperatures and blue to indicate cold temperatures. Although these are common uses to us in daily life, in nature, it’s the other way around.)

We can develop a more precise star thermometer by measuring how much energy a star gives off at each wavelength and by constructing diagrams like Figure 5.8. The location of the peak (or maximum) in the power curve of each star can tell us its temperature. The average temperature at the surface of the Sun, which is where the radiation that we see is emitted, turns out to be 5800 K. (Throughout this text, we use the kelvin or absolute temperature scale. On this scale, water freezes at 273 K and boils at 373 K. All molecular motion ceases at 0 K. The various temperature scales are described in Appendix D.) There are stars cooler than the Sun and stars hotter than the Sun.

The wavelength at which maximum power is emitted can be calculated according to the equation

$$\lambda_{\text{max}} = \frac{3 \times 10^6}{T}$$

where the wavelength is in nanometers (one billionth of a meter) and the temperature is in K (the constant Math input error

Example 5.3: Calculating the Temperature of a Blackbody

We can use Wien’s law to calculate the temperature of a star provided we know the wavelength of peak intensity for its spectrum. If the emitted radiation from a red dwarf star has a wavelength of maximum power at 1200 nm, what is the temperature of this star, assuming it is a blackbody?

Solution

Solving Wien’s law for temperature gives:

$$T = \frac{3 \times 10^6 \text{ nm K}}{\lambda_{\text{max}}} = \frac{3 \times 10^6 \text{ nm K}}{1200 \text{ nm}} = 2500 \text{ K}$$

Exercise 5.2.1

What is the temperature of a star whose maximum light is emitted at a much shorter wavelength of 290 nm?

Answer available at the end of this section.

Since this star has a peak wavelength that is at a shorter wavelength (in the ultraviolet part of the spectrum) than that of our Sun (in the visible part of the spectrum), it should come as no surprise that its surface temperature is much hotter than our Sun’s.

We can also describe our observation that hotter objects radiate more power at all wavelengths in a mathematical form. If we sum up the contributions from all parts of the electromagnetic spectrum, we obtain the total energy emitted by a blackbody. What we usually measure from a large object like a star is the energy flux, the power emitted per square meter. The word *flux* means “flow” here: we are interested in the flow of power into an area (like the area of a telescope mirror). It turns out that the energy flux from a blackbody at temperature *T* is proportional to the fourth power of its absolute temperature. This relationship is known as the Stefan-Boltzmann law and can be written in the form of an equation as

$$F = \sigma T^4$$

where *F* stands for the energy flux (in units of watts per square meter), T is given in Kelvins, and σ (Greek letter sigma) is a constant number $\left(5.67 \times 10^{-8}\right)$.

Notice how impressive this result is. Increasing the temperature of a star would have a tremendous effect on the power it radiates. If the Sun, for example, were twice as hot—that is, if it had a temperature of 11,600 K—it would radiate 2^4, or 16 times more power than it does now. Tripling the temperature would raise the power output 81 times. Hot stars really shine away a tremendous amount of energy.

Example 5.4: Calculating the Power of a Star

While energy flux tells us how much power a star emits per square meter, we would often like to know how much total power is emitted by the star. We can determine that by multiplying the energy flux by the number of square meters on the surface of the star. Stars are mostly spherical, so we can use the formula $4\pi R^2$ for the surface area, where *R* is the radius of the star. The total power emitted by the star (which we call the star's "absolute luminosity") can be found by multiplying the formula for energy flux and the formula for the surface area:

$$L = 4\pi R^2 \sigma T^4$$

Two stars have the same size and are the same distance from us. Star A has a surface temperature of 6000 K, and star B has a surface temperature twice as high, 12,000 K. How much more luminous is star B compared to star A?

Solution

$$L_\mathrm{A} = 4\pi {R_\mathrm{A}}^2 \sigma {T_\mathrm{A}}^4 \text{ and } L_\mathrm{B} = 4\pi {R_\mathrm{B}}^2 \sigma {T_\mathrm{B}}^4$$

Take the ratio of the luminosity of Star A to Star B:

$$\frac{L_\mathrm{B}}{L_\mathrm{A}} = \frac{4\pi {R_\mathrm{B}}^2 \sigma {T_\mathrm{B}}^4}{4\pi {R_\mathrm{A}}^2 \sigma {T_\mathrm{A}}^4} = \frac{{R_\mathrm{B}}^2 {T_\mathrm{B}}^4}{{R_\mathrm{A}}^2 {T_\mathrm{A}}^4}$$

Because the two stars are the same size, $R_\mathrm{A} = R_\mathrm{B}$, leaving

$$\frac{{T_\mathrm{B}}^4}{{T_\mathrm{A}}^4} = \frac{(12{,}000\ \mathrm{K})^4}{(6000\ \mathrm{K})^4} = 2^4 = 16$$

Exercise 5.2.1

Two stars with identical diameters are the same distance away. One has a temperature of 8700 K and the other has a temperature of 2900 K. Which is brighter? How much brighter is it?

Answer available at the end of this section.

Link to Learning

You can use the Blackbody Spectrum simulator to explore the relationship between an object's temperature, peak wavelength, overall brightness, and apparent color.

Answers

Exercise 5.2.1

$T = \frac{3 \times 10^6\ \mathrm{nm\ K}}{\lambda_\mathrm{max}} = \frac{3 \times 10^6\ \mathrm{nm\ K}}{290\ \mathrm{nm}} = 10{,}300\ \mathrm{K}$

Exercise 5.2.1

The 8700 K star has triple the temperature, so it is 3^4 = 81 times brighter.

This page titled 5.2: The Electromagnetic Spectrum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.3: The Electromagnetic Spectrum by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.3: Spectroscopy in Astronomy

Learning Objectives

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

- Describe the properties of light
- Explain how astronomers learn the composition of a gas by examining its spectral lines
- Discuss the various types of spectra

Electromagnetic radiation carries a lot of information about the nature of stars and other astronomical objects. To extract this information, however, astronomers must be able to study the amounts of energy we receive at different wavelengths of light in fine detail. Let's examine how we can do this and what we can learn.

Properties of Light

Light exhibits certain behaviors that are important to the design of telescopes and other instruments. For example, light can be *reflected* from a surface. If the surface is smooth and shiny, as with a mirror, the direction of the reflected light beam can be calculated accurately from knowledge of the shape of the reflecting surface. Light is also bent, or *refracted*, when it passes from one kind of transparent material into another—say, from the air into a glass lens.

Reflection and refraction of light are the basic properties that make possible all *optical* instruments (devices that help us to see things better)—from eyeglasses to giant astronomical telescopes. Such instruments are generally combinations of glass lenses, which bend light according to the principles of refraction, and curved mirrors, which depend on the properties of reflection. Small optical devices, such as eyeglasses or binoculars, generally use lenses, whereas large telescopes depend almost entirely on mirrors for their main optical elements. We will discuss astronomical instruments and their uses more fully in Astronomical Instruments. For now, we turn to another behavior of light, one that is essential for the decoding of light.

In 1672, in the first paper that he submitted to the Royal Society, Sir Isaac Newton described an experiment in which he permitted sunlight to pass through a small hole and then through a prism. Newton found that sunlight, which looks white to us, is actually made up of a mixture of all the colors of the rainbow (Figure 5.9).

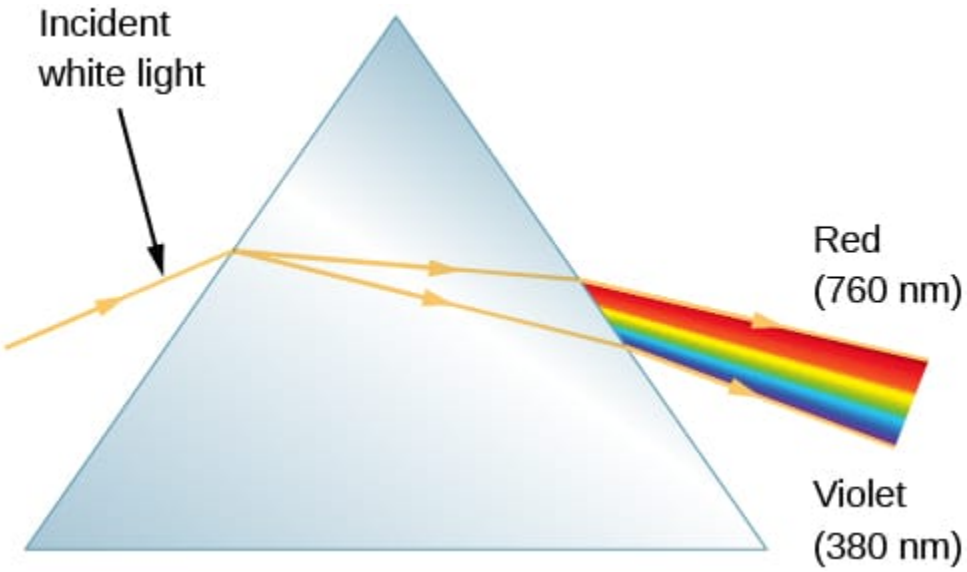


Figure 5.9 : Action of a Prism. When we pass a beam of white sunlight through a prism, we see a rainbow-colored band of light that we call a continuous spectrum.

Figure 5.9 shows how light is separated into different colors with a prism—a piece of glass in the shape of a triangle with refracting surfaces. Upon entering one face of the prism, the path of the light is refracted (bent), but not all of the colors are bent by the same amount. The bending of the beam depends on the wavelength of the light as well as the properties of the material, and as a result, different wavelengths (or colors of light) are bent by different amounts and therefore follow slightly different paths through the prism. The violet light is bent more than the red. This phenomenon is called dispersion and explains Newton's rainbow experiment.

Upon leaving the opposite face of the prism, the light is bent again and further dispersed. If the light leaving the prism is focused on a screen, the different wavelengths or colors that make up white light are lined up side by side just like a rainbow (Figure 5.10). (In fact, a rainbow is formed by the dispersion of light though raindrops; see The Rainbow feature box.) Because this array of colors is a spectrum of light, the instrument used to disperse the light and form the spectrum is called a spectrometer.

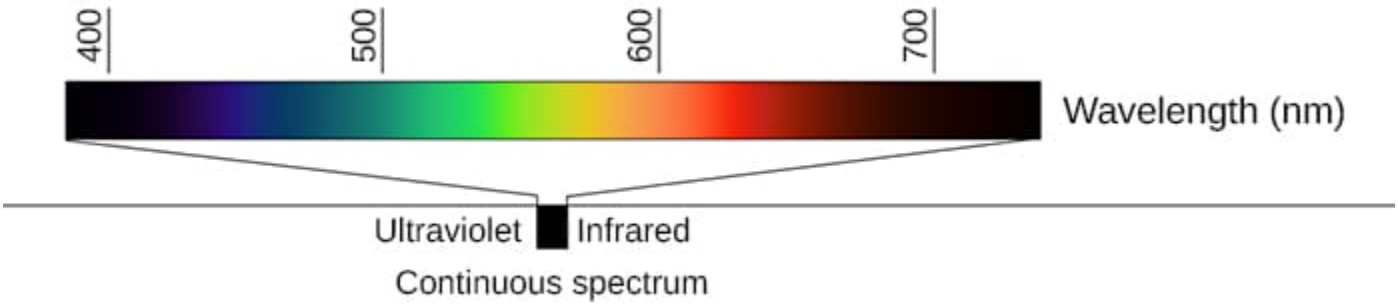


Figure 5.10 : Continuous Spectrum. When white light passes through a prism, it is dispersed and forms a continuous spectrum of all the colors. Although it is hard to see in this printed version, in a well-dispersed spectrum, many subtle gradations in color are visible as your eye scans from one end (violet) to the other (red).

The Value of Stellar Spectra

When Newton described the laws of refraction and dispersion in optics, and observed the solar spectrum, all he could see was a continuous band of colors. If the spectrum of the white light from the Sun and stars were simply a continuous rainbow of colors, astronomers would have little interest in the detailed study of a star's spectrum once they had learned its average surface temperature. In 1802, however, William Wollaston built an improved spectrometer that included a lens to focus the Sun's spectrum on a screen. With this device, Wollaston saw that the colors were not spread out uniformly, but instead, some ranges of color were missing, appearing as dark bands in the solar spectrum. He mistakenly attributed these lines to natural boundaries between the colors. In 1815, German physicist Joseph Fraunhofer, upon a more careful examination of the solar spectrum, found about 600 such dark lines (missing colors), which led scientists to rule out the boundary hypothesis (Figure 5.11).

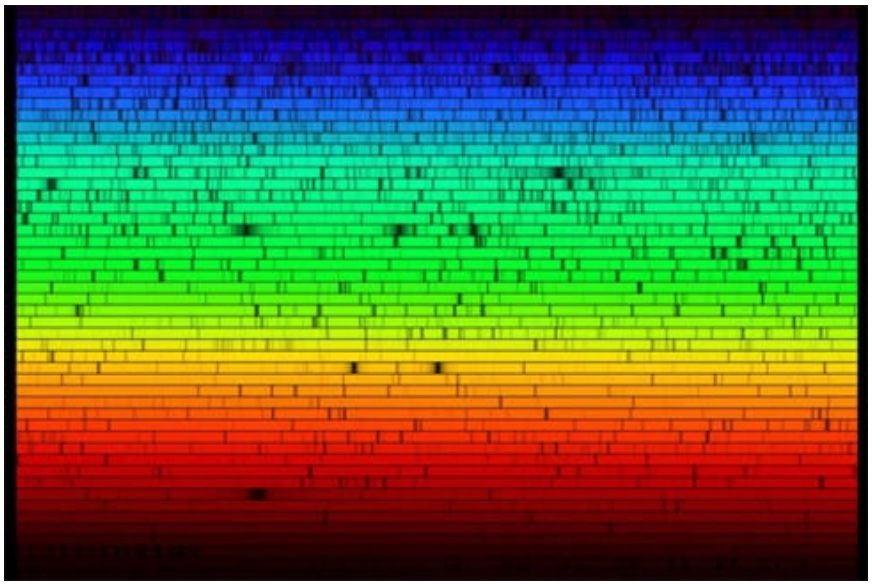


Figure 5.11 : Visible Spectrum of the Sun. Our star's spectrum is crossed by dark lines produced by atoms in the solar atmosphere that absorb light at certain wavelengths. (credit: modification of work by Nigel Sharp, NOAO/National Solar Observatory at Kitt Peak/AURA, and the National Science Foundation)

Later, researchers found that similar dark lines could be produced in the spectra ("spectra" is the plural of "spectrum") of artificial light sources. They did this by passing their light through various apparently transparent substances—usually containers with just a bit of thin gas in them.

These gases turned out not to be transparent at *all* colors: they were quite opaque at a few sharply defined wavelengths. Something in each gas had to be absorbing just a few colors of light and no others. All gases did this, but each different element absorbed a different set of colors and thus showed different dark lines. If the gas in a container consisted of two elements, then light passing through it was missing the colors (showing dark lines) for both of the elements. So it became clear that certain lines in the spectrum "go with" certain elements. This discovery was one of the most important steps forward in the history of astronomy.

What would happen if there were no continuous spectrum for our gases to remove light from? What if, instead, we heated the same thin gases until they were hot enough to glow with their own light? When the gases were heated, a spectrometer revealed no continuous spectrum, but several separate bright lines. That is, these hot gases emitted light only at certain specific wavelengths or colors.

When the gas was pure hydrogen, it would emit one pattern of colors; when it was pure sodium, it would emit a different pattern. A mixture of hydrogen and sodium emitted both sets of spectral lines. The colors the gases emitted when they were heated were the very same colors as those they had absorbed when a continuous source of light was behind them. From such experiments, scientists began to see that different substances showed distinctive *spectral signature*s by which their presence could be detected (Figure 5.12). Just as your signature allows the bank to identify you, the unique pattern of colors for each type of atom (its spectrum) can help us identify which element or elements are in a gas.

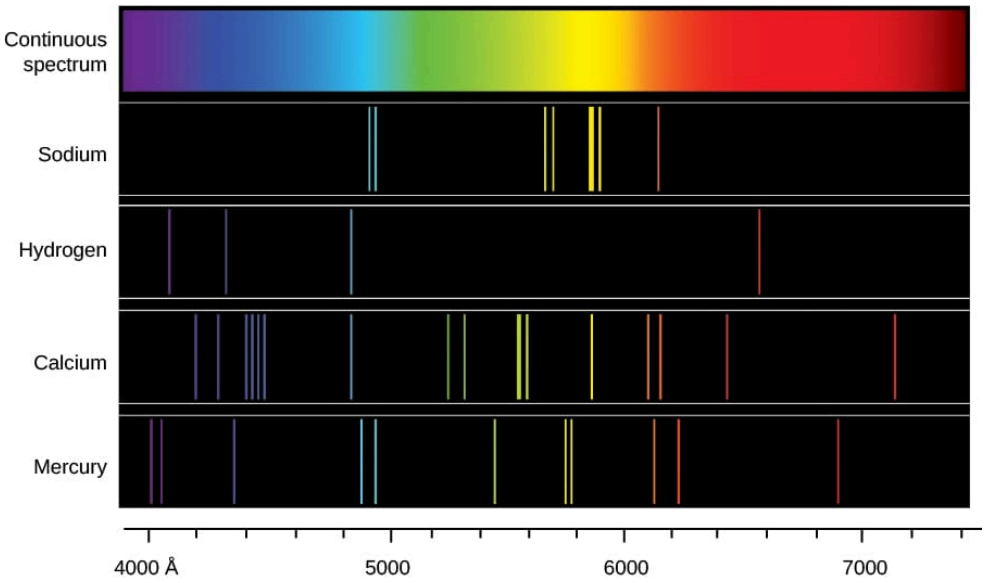


Figure 5.12 : Continuous Spectrum and Line Spectra from Different Elements. Each type of glowing gas (each element) produces its own unique pattern of lines, so the composition of a gas can be identified by its spectrum. The spectra of sodium, hydrogen, calcium, and mercury gases are shown here.

Types of Spectra

In these experiments, then, there were three different types of spectra. A continuous spectrum (formed when a solid or very dense gas gives off radiation) is an array of all wavelengths or colors of the rainbow. A continuous spectrum can serve as a backdrop from which the atoms of much less dense gas can absorb light. A dark line, or absorption spectrum, consists of a series or pattern of dark lines—missing colors—superimposed upon the continuous spectrum of a source. A bright line, or emission spectrum, appears as a pattern or series of bright lines; it consists of light in which only certain discrete wavelengths are present. (Figure 5.11 shows an absorption spectrum, whereas Figure 5.12 shows the emission spectrum of a number of common elements along with an example of a continuous spectrum.)

When we have a hot, thin gas, each particular chemical element or compound produces its own characteristic pattern of spectral lines—its spectral signature. No two types of atoms or molecules give the same patterns. In other words, each particular gas can absorb or emit only certain wavelengths of the light peculiar to that gas. In contrast, absorption spectra occur when passing white light through a cool, thin gas. The temperature and other conditions determine whether the lines are bright or dark (whether light is absorbed or emitted), but the wavelengths of the lines for any element are the same in either case. It is the precise pattern of wavelengths that makes the signature of each element unique. Liquids and solids can also generate spectral lines or bands, but they are broader and less well defined—and hence, more difficult to interpret. Spectral analysis, however, can be quite useful. It can, for example, be applied to light reflected off the surface of a nearby asteroid as well as to light from a distant galaxy.

The dark lines in the solar spectrum thus give evidence of certain chemical elements between us and the Sun absorbing those wavelengths of sunlight. Because the space between us and the Sun is pretty empty, astronomers realized that the atoms doing the absorbing must be in a thin atmosphere of cooler gas around the Sun. This outer atmosphere is not all that different from the rest of the Sun, just thinner and cooler. Thus, we can use what we learn about its composition as an indicator of what the whole Sun is made of. Similarly, we can use the presence of absorption and emission lines to analyze the composition of other stars and clouds of gas in space.

Such analysis of spectra is the key to modern astronomy. Only in this way can we “sample” the stars, which are too far away for us to visit. Encoded in the electromagnetic radiation from celestial objects is clear information about the chemical makeup of these objects. Only by understanding what the stars were made of could astronomers begin to form theories about what made them shine and how they evolved.

In 1860, German physicist Gustav Kirchhoff became the first person to use spectroscopy to identify an element in the Sun when he found the spectral signature of sodium gas. In the years that followed, astronomers found many other chemical elements in the Sun and stars. In fact, the element helium was found first in the Sun from its spectrum and only later identified on Earth. (The word “helium” comes from *helios*, the Greek name for the Sun.)

Why are there specific lines for each element? The answer to that question was not found until the twentieth century; it required the development of a model for the atom. We therefore turn next to a closer examination of the atoms that make up all matter.

Link to Learning

You can use the Spectrum Constructor to explore the three types of spectra and how they appear through a diffraction grating when the brightness of each wavelength is measured with a spectrometer.

Making Connections: The Rainbow

Rainbows are an excellent illustration of the dispersion of sunlight. You have a good chance of seeing a rainbow any time you are between the Sun and a rain shower, as illustrated in Figure 5.13. The raindrops act like little prisms and break white light into the spectrum of colors. Suppose a ray of sunlight encounters a raindrop and passes into it. The light changes direction—is refracted—when it passes from air to water; the blue and violet light are refracted more than the red. Some of the light is then reflected at the backside of the drop and reemerges from the front, where it is again refracted. As a result, the white light is spread out into a rainbow of colors.



Figure 5.13 : Rainbow Refraction. (a) This diagram shows how light from the Sun, which is located behind the observer, can be refracted by raindrops to produce (b) a rainbow. (c) Refraction separates white light into its component colors.

Note that violet light lies above the red light after it emerges from the raindrop. When you look at a rainbow, however, the red light is higher in the sky. Why? Look again at Figure 5.13. If the observer looks at a raindrop that is high in the sky, the violet light passes over her head and the red light enters her eye. Similarly, if the observer looks at a raindrop that is low in the sky, the violet light reaches her eye and the drop appears violet, whereas the red light from that same drop strikes the ground and is not seen. Colors of intermediate wavelengths are refracted to the eye by drops that are intermediate in altitude between the drops that appear violet and the ones that appear red. Thus, a single rainbow always has red on the outside and violet on the inside.

This page titled 5.3: Spectroscopy in Astronomy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.4: Spectroscopy in Astronomy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.4: The Structure of the Atom

Learning Objectives

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

- Describe the structure of atoms and the components of nuclei
- The idea that matter is composed of tiny particles called atoms is at least 25 centuries old. It took until the twentieth century, however, for scientists to invent instruments that permitted them to probe inside an atom and find that it is not, as had been thought, hard and indivisible. Instead, the atom is a complex structure composed of still smaller particles.

Probing the Atom

The first of these smaller particles was discovered by British physicist James (J. J.) Thomson in 1897. Named the *electron*, this particle is negatively charged. (It is the flow of these particles that produces currents of electricity, whether in lightning bolts or in the wires leading to your lamp.) Because an atom in its normal state is electrically neutral, each electron in an atom must be balanced by the same amount of positive charge.

The next step was to determine where in the atom the positive and negative charges are located. In 1911, British physicist Ernest Rutherford devised an experiment that provided part of the answer to this question. He bombarded an extremely thin piece of gold foil, only about 400 atoms thick, with a beam of alpha particles (Figure 5.14). *Alpha particles* (α particles) are helium atoms that have lost their electrons and thus are positively charged. Most of these particles passed though the gold foil just as if it and the atoms in it were nearly empty space. About 1 in 8000 of the alpha particles, however, completely reversed direction and bounced backward from the foil. Rutherford wrote, “It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.”

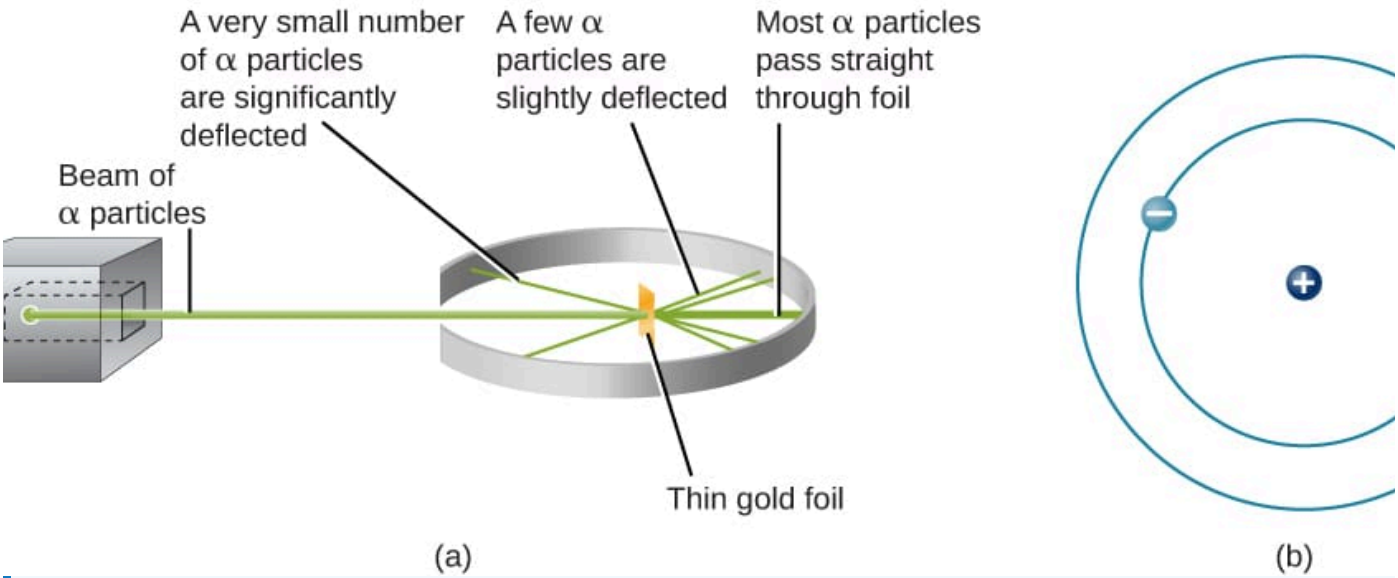


Figure 5.14 : Rutherford’s Experiment. (a) When Rutherford allowed α particles from a radioactive source to strike a target of gold foil, he found that, although most of them went straight through, some rebounded back in the direction from which they came. (b) From this experiment, he concluded that the atom must be constructed like a miniature solar system, with the positive charge concentrated in the nucleus and the negative charge orbiting in the large volume around the nucleus. Note that this drawing is not to scale; the electron orbits are much larger relative to the size of the nucleus.

The only way to account for the particles that reversed direction when they hit the gold foil was to assume that nearly all of the mass, as well as all of the positive charge in each individual gold atom, is concentrated in a tiny center or nucleus. When a positively charged alpha particle strikes a nucleus, it reverses direction, much as a cue ball reverses direction when it strikes another billiard ball. Rutherford's model placed the other type of charge—the negative electrons—in orbit around this nucleus.

Rutherford's model required that the electrons be in motion. Positive and negative charges attract each other, so stationary electrons would fall into the positive nucleus. Also, because both the electrons and the nucleus are extremely small, most of the atom is empty, which is why nearly all of Rutherford's particles were able to pass right through the gold foil without colliding with anything. Rutherford's model was a very successful explanation of the experiments he conducted, although eventually scientists would discover that even the nucleus itself has structure.

The Atomic Nucleus

The simplest possible atom (and the most common one in the Sun and stars) is hydrogen. The nucleus of ordinary hydrogen contains a single proton. Moving around this proton is a single electron. The mass of an electron is nearly 2000 times smaller than the mass of a proton; the electron carries an amount of charge exactly equal to that of the proton but opposite in sign (Figure 5.15). Opposite charges attract each other, so it is an electromagnetic force that holds the proton and electron together, just as gravity is the force that keeps planets in orbit around the Sun.

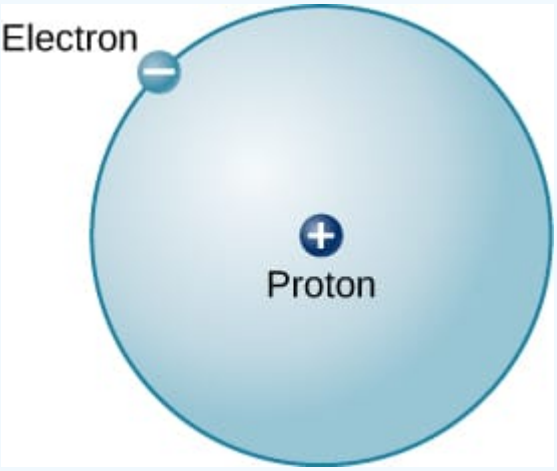


Figure 5.15 : Hydrogen Atom. This is a schematic diagram of a hydrogen atom in its lowest energy state, also called the ground state. The proton and electron have equal but opposite charges, which exert an electromagnetic force that binds the hydrogen atom together. In the illustration, the size of the particles is exaggerated so that you can see them; they are not to scale. They are also shown much closer than they would actually be as it would take more than an entire page to show their actual distance to scale.

There are many other types of atoms in nature. Helium, for example, is the second-most abundant element in the Sun. Helium has two protons in its nucleus instead of the single proton that characterizes hydrogen. In addition, the helium nucleus contains two neutrons, particles with a mass comparable to that of the proton but with no electric charge. Moving around this nucleus are two electrons, so the total net charge of the helium atom is also zero (Figure 5.16).

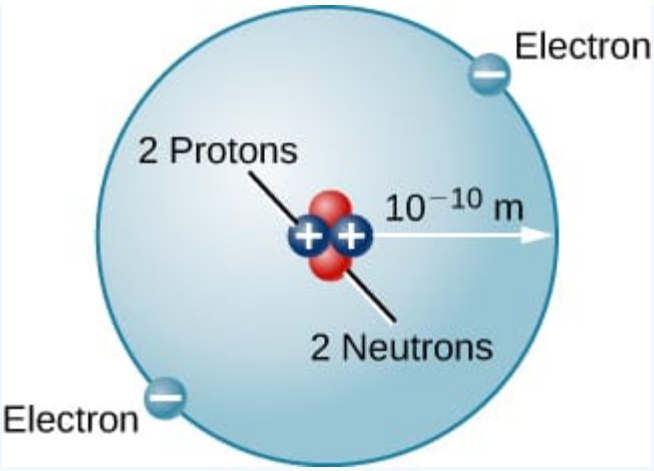


Figure 5.16 : Helium Atom. Here we see a schematic diagram of a helium atom in its lowest energy state. Two protons are present in the nucleus of all helium atoms. In the most common variety of helium, the nucleus also contains two neutrons, which have nearly the same mass as the proton but carry no charge. Two electrons orbit the nucleus.

From this description of hydrogen and helium, perhaps you have guessed the pattern for building up all the elements (different types of atoms) that we find in the universe. The type of element is determined by the number of protons in the nucleus of the atom. For example, any atom with six protons is the element carbon, with eight protons is oxygen, with 26 is iron, and with 92 is uranium. On Earth, a typical atom has the same number of electrons as protons, and these electrons follow complex orbital patterns around the nucleus. Deep inside stars, however, it is so hot that the electrons get loose from the nucleus and (as we shall see) lead separate yet productive lives.

The ratio of neutrons to protons increases as the number of protons increases, but each element is unique. The number of neutrons is not necessarily the same for all atoms of a given element. For example, most hydrogen atoms contain no neutrons at all. There are, however, hydrogen atoms that contain one proton and one neutron, and others that contain one proton and two neutrons. The various types of hydrogen nuclei with different numbers of neutrons are called isotopes of hydrogen (Figure 5.17), and all other elements have isotopes as well. You can think of isotopes as siblings in the same element "family"—closely related but with different characteristics and behaviors.

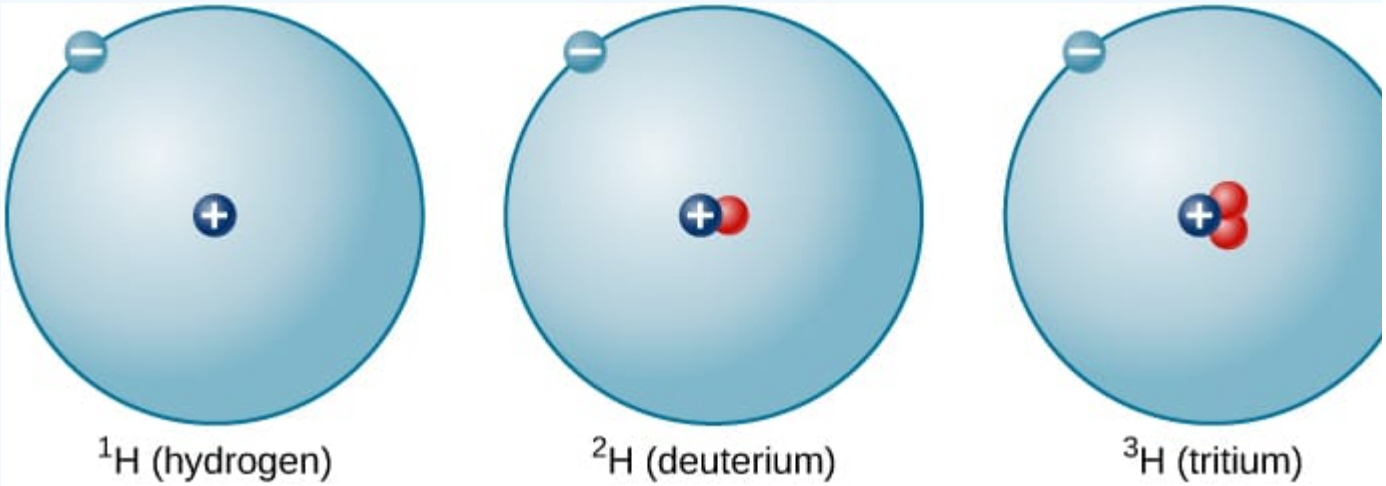


Figure 5.17 : Isotopes of Hydrogen. A single proton in the nucleus defines the atom to be hydrogen, but there may be zero, one, or two neutrons. The most common isotope of hydrogen is the one with only a single proton and no neutrons.

Link to Learning

To explore the structure of atoms, go to the PhET Build an Atom website where you can add protons, neutrons, or electrons to a model and the name of the element you have created will appear. You can also see the net charge, the mass number, whether it is stable or unstable, and whether it is an ion or a neutral atom.

The Bohr Atom

Rutherford's model for atoms has one serious problem. Maxwell's theory of electromagnetic radiation says that when electrons change either speed or the direction of motion, they must emit energy. Orbiting electrons constantly change their direction of motion, so they should emit a constant stream of energy. Applying Maxwell's theory to Rutherford's model, all electrons should spiral into the nucleus of the atom as they lose energy, and this collapse should happen very quickly—in about 10^{-16} seconds.

It was Danish physicist Niels Bohr (1885–1962) who solved the mystery of how electrons remain in orbit. He was trying to develop a model of the atom that would also explain certain regularities observed in the spectrum of hydrogen. He suggested that the spectrum of hydrogen can be understood if we assume that orbits of only certain sizes are possible for the electron. Bohr further assumed that as long as the electron moves in only one of these allowed orbits, it radiates no energy: its energy would change only if it moved from one orbit to another.

This suggestion, in the words of science historian Abraham Pais, was "one of the most audacious hypotheses ever introduced in physics." If something equivalent were at work in the everyday world, you might find that, as you went for a walk after astronomy class, nature permitted you to walk two steps per minute, five steps per minute, and 12 steps per minute, but no speeds in between. No matter how you tried to move your legs, only certain walking speeds would be permitted. To make things more bizarre, it would take no effort to walk at any one of the allowed speeds, but it would be difficult to change from one speed to another. Luckily, no such rules apply at the level of human behavior. But at the microscopic level of the atom, experiment after experiment has confirmed the validity of Bohr's strange idea. Bohr's suggestions became one of the foundations of the new (and much more sophisticated) model of the subatomic world called quantum mechanics.

In Bohr's model, if the electron moves from one orbit to another closer to the atomic nucleus, it must give up some energy in the form of electromagnetic radiation. If the electron goes from an inner orbit to one farther from the nucleus, however, it requires some additional energy. One way to obtain the necessary energy is to absorb electromagnetic radiation that may be streaming past the atom from an outside source.

A key feature of Bohr's model is that each of the permitted electron orbits around a given atom has a certain energy value; we therefore can think of each orbit as an energy level. To move from one orbit to another (which will have its own specific energy value) requires a change in the electron's energy—a change determined by the difference between the two energy values. If the electron goes to a lower level, the energy difference will be given off; if the electron goes to a higher level, the energy difference must be obtained from somewhere else. Each jump (or transition) to a different level has a fixed and definite energy change associated with it.

A crude analogy for this situation might be life in a tower of luxury apartments where the rent is determined by the quality of the view. Such a building has certain, definite numbered levels or floors on which apartments are located. No one can live on floor 5.37 or 22.5. In addition, the rent gets higher as you go up to higher floors. If you want to exchange an apartment on the twentieth floor for one on the second floor, you will not owe as much rent. However, if you want to move from the third floor to the twenty-fifth floor, your rent will increase. In an atom, too, the "cheapest" place for an electron to live is the lowest possible level, and energy is required to move to a higher level.

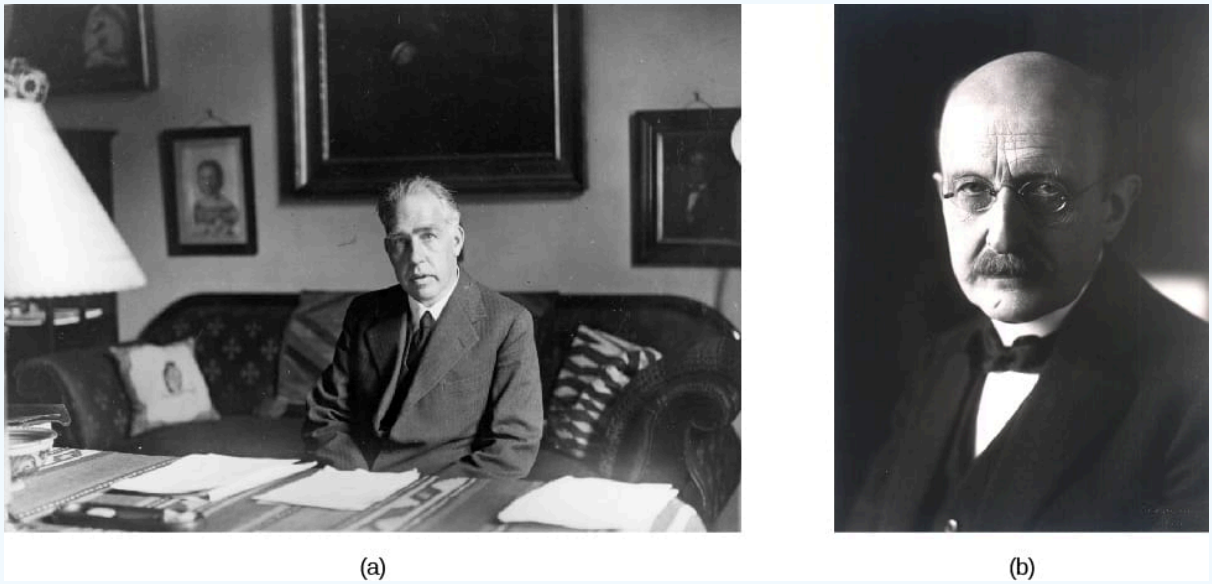
Here we have one of the situations where it is easier to think of electromagnetic radiation as particles (photons) rather than as waves. As electrons move from one level to another, they give off or absorb little packets of energy. When an electron moves to a higher level, it absorbs a photon of just the right energy (provided one is available). When it moves to a lower level, it emits a photon with the exact amount of energy it no longer needs in its "lower-cost living situation."

The photon and wave perspectives must be equivalent: light is light, no matter how we look at it. Thus, each photon carries a certain amount of energy that is proportional to the frequency (*f*) of the wave it represents. The value of its energy (*E*) is given by the formula

$$E = hf$$

where the constant of proportionality, *h*, is called Planck's constant.

The constant is named for Max Planck, the German physicist who was one of the originators of the quantum theory (Figure 5.18). If metric units are used (that is, if energy is measured in joules and frequency in hertz), then Planck's constant has the value $h = 6.626 \times 10^{-34}$ joule-seconds (J-s). Higher-energy photons correspond to higher-frequency waves (which have a shorter wavelength); lower-energy photons are waves of lower frequency.



(a) (b)

Figure 5.18 : Niels Bohr (1885–1962) and Max Planck (1858–1947). (a) Bohr, shown at his desk in this 1935 photograph, and (b) Planck helped us understand the energy behavior of photons.

To take a specific example, consider a calcium atom inside the Sun's atmosphere in which an electron jumps from a lower level to a higher level. To do this, it needs about 5×10^{-19} joules of energy, which it can conveniently obtain by absorbing a passing photon of that energy coming from deeper inside the Sun. This photon is equivalent to a wave of light whose frequency is about 7.5×10^{14} hertz and whose wavelength is about 3.9×10^{-7} meters (393 nanometers), in the deep violet part of the visible light spectrum. Although it may seem strange at first to switch from picturing light as a photon (or energy packet) to picturing it as a wave, such switching has become second nature to astronomers and can be a handy tool for doing calculations about spectra.

Example 5.5: The Energy of a Photon

Now that we know how to calculate the wavelength and frequency of a photon, we can use this information, along with Planck's constant, to determine how much energy each photon carries. How much energy does a red photon of wavelength 630 nm have?

Solution

First, as we learned earlier, we can find the frequency of the photon:

$$f = \frac{c}{\lambda} = \frac{3 \times 10^{8}\,\mathrm{m/s}}{630 \times 10^{-9}\,\mathrm{m}} = 4.8 \times 10^{14}\,\mathrm{Hz}$$

Next, we can use Planck's constant to determine the energy (remember that a Hz is the same as 1/s):

$$E = hf = \left(6.626 \times 10^{-34}\ \text{J-s}\right)\left(4.8 \times 10^{14}\ \text{Hz}(1/\text{s})\right) = 3.2 \times 10^{-19}\ \text{J}$$

Exercise 5.4.1

What is the energy of a yellow photon with a frequency of 5.5×10^{14} Hz?

Answer available at the end of this section.

Answers

Exercise 5.4.1

$E = hf = \left(6.626 \times 10^{-34}\text{J-s}\right)\left(5.5 \times 10^{14}\text{Hz}\right) = 3.6 \times 10^{-19}\ \text{J}$

This page titled 5.4: The Structure of the Atom is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.5: The Structure of the Atom by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.5: Formation of Spectral Lines

Learning Objectives

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

- Explain how emission line spectra and absorption line spectra are formed
- Describe what ions are and how they are formed
- Explain how spectral lines and ionization levels in a gas can help us determine its temperature

We can use Bohr's model of the atom to understand how spectral lines are formed. The concept of energy levels for the electron orbits in an atom leads naturally to an explanation of why atoms absorb or emit only specific energies or wavelengths of light.

The Hydrogen Spectrum

Let's look at the hydrogen atom from the perspective of the Bohr model. Suppose a beam of white light (which consists of photons of all visible wavelengths) shines through a gas of atomic hydrogen. A photon of wavelength 656 nanometers has just the right energy to raise an electron in a hydrogen atom from the second to the third orbit. Thus, as all the photons of different energies (or wavelengths or colors) stream by the hydrogen atoms, photons with *this* particular wavelength can be absorbed by those atoms whose electrons are orbiting on the second level. When they are absorbed, the electrons on the second level will move to the third level, and a number of the photons of this wavelength and energy will be missing from the general stream of white light.

Other photons will have the right energies to raise electrons from the second to the fourth orbit, or from the first to the fifth orbit, and so on. Only photons with these exact energies can be absorbed. All of the other photons will stream past the atoms untouched. Thus, hydrogen atoms absorb light at only certain wavelengths and produce dark lines at those wavelengths in the spectrum we see.

Suppose we have a container of hydrogen gas through which a whole series of photons is passing, allowing many electrons to move up to higher levels. When we turn off the light source, these electrons "fall" back down from larger to smaller orbits and emit photons of light—but, again, only light of those energies or wavelengths that correspond to the energy difference between permissible orbits. The orbital changes of hydrogen electrons that give rise to some spectral lines are shown in Figure 5.19.

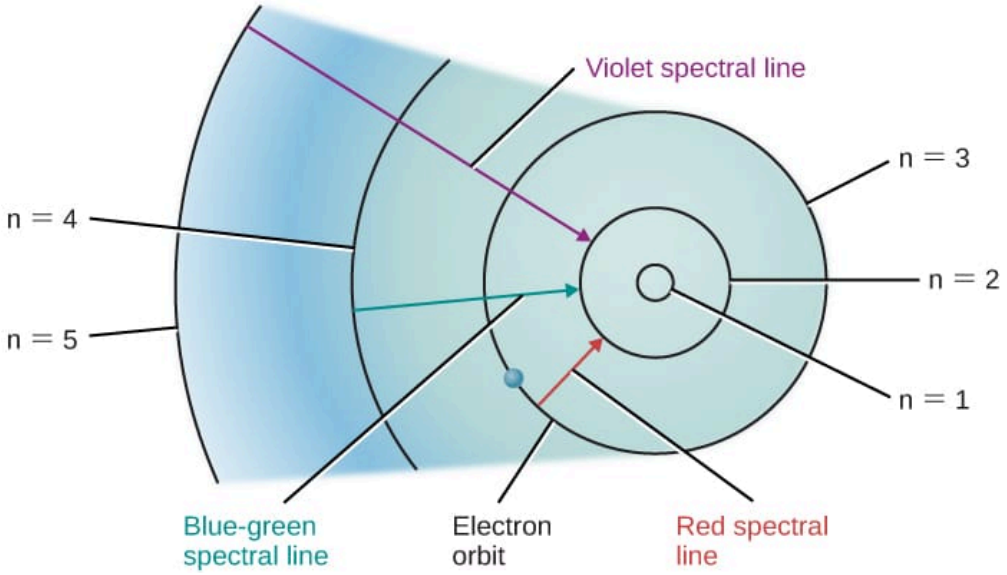


Figure 5.19 : Bohr Model for Hydrogen. In this simplified model of a hydrogen atom, the concentric circles shown represent permitted orbits or energy levels. An electron in a hydrogen atom can only exist in one of these energy levels (or states). The closer the electron is to the nucleus, the more tightly bound the electron is to the nucleus. By absorbing energy, the electron can move to energy levels farther from the nucleus (and even escape if enough energy is absorbed).

Similar pictures can be drawn for atoms other than hydrogen. However, because these other atoms ordinarily have more than one electron each, the orbits of their electrons are much more complicated, and the spectra are more complex as well. For our purposes, the key conclusion is this: *each type of atom has its own unique pattern of electron orbits, and no two sets of orbits are exactly alike*. This means that each type of atom shows its own unique set of spectral lines, produced by electrons moving between its unique set of orbits.

Astronomers and physicists have worked hard to learn the lines that go with each element by studying the way atoms absorb and emit light in laboratories here on Earth. Then they can use this knowledge to identify the elements in celestial bodies. In this way, we now know the chemical makeup of not just any star, but even galaxies of stars so distant that their light started on its way to us long before Earth had even formed.

Energy Levels and Excitation

Bohr's model of the hydrogen atom was a great step forward in our understanding of the atom. However, we know today that atoms cannot be represented by quite so simple a picture. For example, the concept of sharply defined electron orbits is not really correct; however, at the level of this introductory course, the notion that only certain discrete energies are allowable for an atom is very useful. The energy levels we have been discussing can be thought of as representing certain average distances of the electron's possible orbits from the atomic nucleus.

Ordinarily, an atom is in the state of lowest possible energy, its ground state. In the Bohr model of the hydrogen atom, the ground state corresponds to the electron being in the innermost orbit. An atom can absorb energy, which raises it to a higher energy level (corresponding, in the simple Bohr picture, to an electron's movement to a larger orbit)—this is referred to as excitation. The atom is then said to be in an *excited state*. Generally, an atom remains excited for only a very brief time. After a short interval, typically a hundred-millionth of a second or so, it drops back spontaneously to its ground state, with the simultaneous emission of light. The atom may return to its lowest state in one jump, or it may make the transition in steps of two or more jumps, stopping at intermediate levels on the way down. With each jump, it emits a photon of the wavelength that corresponds to the energy difference between the levels at the beginning and end of that jump.

An energy-level diagram for a hydrogen atom and several possible atomic transitions are shown in Figure 5.20. When we measure the energies involved as the atom jumps between levels, we find that the transitions to or from the ground state, called the *Lyman series* of lines, result in the emission or absorption of ultraviolet photons. But the transitions to or from the first excited state (labeled n = 2 in part (a) of Figure 5.20), called the Balmer series, produce emission or absorption in visible light. In fact, it was to explain this Balmer series that Bohr first suggested his model of the atom.

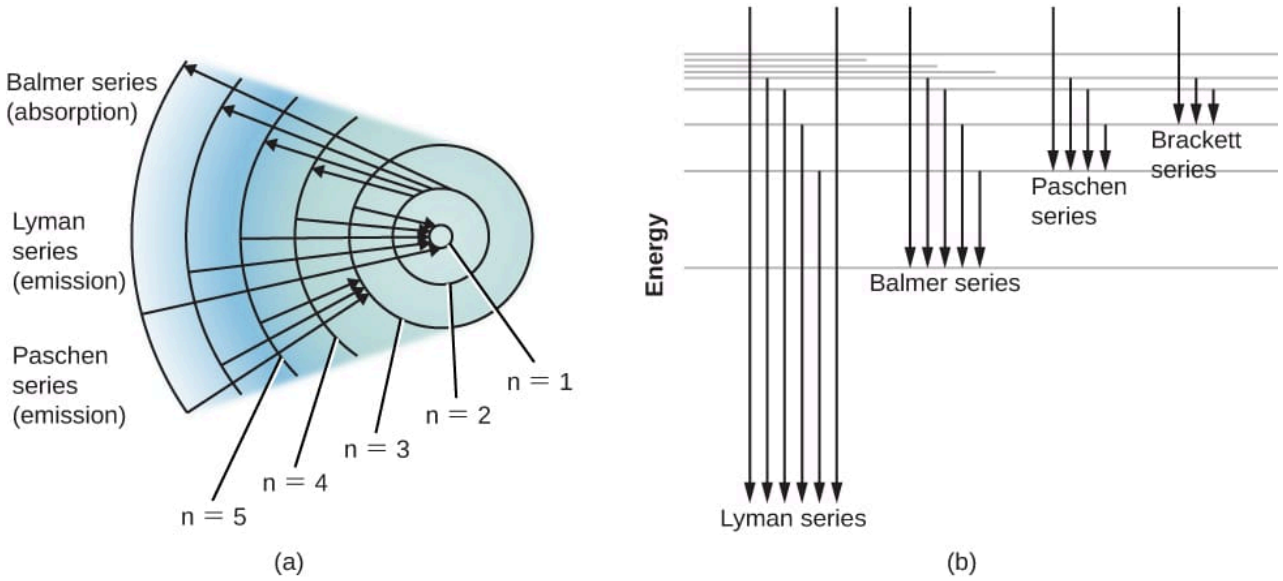


Figure 5.20 : Energy-Level Diagrams for Hydrogen. (a) Here we follow the emission or absorption of photons by a hydrogen atom according to the Bohr model. Several different series of spectral lines are shown, corresponding to transitions of electrons from or to certain allowed orbits. Each series of lines that terminates on a specific inner orbit is named for the physicist who studied it. At the top, for example, you see the Balmer series, and arrows show electrons jumping from the second orbit (n = 2) to the third, fourth, fifth, and sixth orbits. Each time a "poor" electron from a lower level wants to rise to a higher position in life, it must absorb energy to do so. It can absorb the energy it needs from passing waves (or photons) of light. The next set of arrows (Lyman series) show electrons falling down to the first orbit from different (higher) levels. Each time a "rich" electron goes downward toward the nucleus, it can afford to give off (emit) some energy it no longer needs. (In this idealized diagram, the energy levels are shown equally spaced; in real life, they are not.) (b) At higher and higher energy levels, the levels become more and more crowded together, approaching a limit. The region above the top line represents energies at which the atom is ionized (the electron is no longer attached to the atom). Each series of arrows represents electrons falling from higher levels to lower ones, releasing photons or waves of energy in the process.

Atoms that have absorbed specific photons from a passing beam of white light and have thus become excited generally de-excite themselves and emit that light again in a very short time. You might wonder, then, why *dark* spectral lines are ever produced. In other words, why doesn't this reemitted light quickly "fill in" the darker absorption lines?

Imagine a beam of white light coming toward you through some cooler gas. Some of the reemitted light *is* actually returned to the beam of white light you see, but this fills in the absorption lines only to a slight extent. The reason is that the atoms in the gas reemit light *in all directions*, and only a small fraction of the reemitted light is in the direction of the original beam (toward you). In a star, much of the reemitted light actually goes in directions leading back into the star, which does observers outside the star no good whatsoever.

Figure 5.21 summarizes the different kinds of spectra we have discussed. An incandescent lightbulb produces a continuous spectrum. When that continuous spectrum is viewed through a thinner cloud of gas, an absorption line spectrum can be seen superimposed on the continuous spectrum. If we look only at a cloud of excited gas atoms (with no continuous source seen behind it), we see that the excited atoms give off an emission line spectrum.

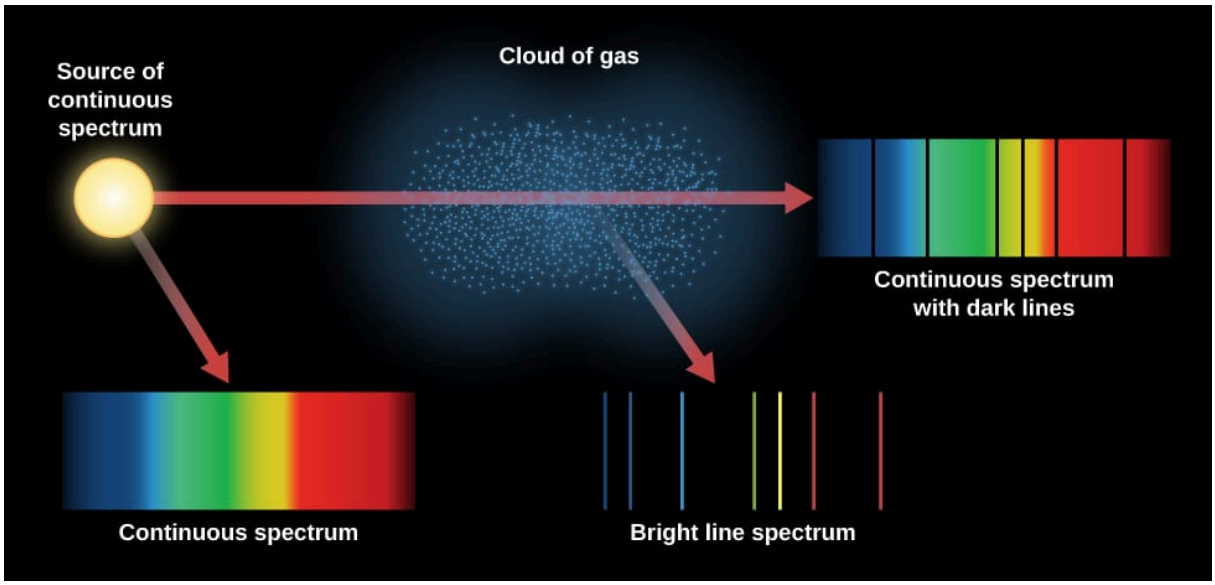


Figure 5.21 : Three Kinds of Spectra. When we see a lightbulb or other source of continuous radiation, all the colors are present. When the continuous spectrum is seen through a thinner gas cloud, the cloud's atoms produce absorption lines in the continuous spectrum. When the excited cloud is seen without the continuous source behind it, its atoms produce emission lines. We can learn which types of atoms are in the gas cloud from the pattern of absorption or emission lines.

Atoms in a hot gas are moving at high speeds and continually colliding with one another and with any loose electrons. They can be excited (electrons moving to a higher level) and de-excited (electrons moving to a lower level) by these collisions as well as by absorbing and emitting light. The speed of atoms in a gas depends on the temperature. When the temperature is higher, so are the speed and energy of the collisions. The hotter the gas, therefore, the more likely that electrons will occupy the outermost orbits, which correspond to the highest energy levels. This means that the level where electrons *start* their upward jumps in a gas can serve as an indicator of how hot that gas is. In this way, the absorption lines in a spectrum give astronomers information about the temperature of the regions where the lines originate.

Link to Learning

Use this simulation to play with a hydrogen atom and see what happens when electrons move to higher levels and then give off photons as they go to a lower level.

Ionization

We have described how certain discrete amounts of energy can be absorbed by an atom, raising it to an excited state and moving one of its electrons farther from its nucleus. If enough energy is absorbed, the electron can be completely removed from the atom—this is called ionization. The atom is then said to be ionized. The minimum amount of energy required to remove one electron from an atom in its ground state is called its ionization energy.

Still-greater amounts of energy must be absorbed by the now-ionized atom (called an ion) to remove an additional electron deeper in the structure of the atom. Successively greater energies are needed to remove the third, fourth, fifth—and so on—electrons from the atom. If enough energy is available, an atom can become completely ionized, losing all of its electrons. A hydrogen atom, having only one electron to lose, can be ionized only once; a helium atom can be ionized twice; and an oxygen atom up to eight times. When we examine regions of the cosmos where there is a great deal of energetic radiation, such as the neighborhoods where hot young stars have recently formed, we see a lot of ionization going on.

An atom that has become positively ionized has lost a negative charge—the missing electron—and thus is left with a net positive charge. It therefore exerts a strong attraction on any free electron. Eventually, one or more electrons will be captured and the atom will become neutral (or ionized to one less degree) again. During the electron-capture process, the atom emits one or more photons. Which photons are emitted depends on whether the electron is captured at once to the lowest energy level of the atom or stops at one or more intermediate levels on its way to the lowest available level.

Just as the excitation of an atom can result from a collision with another atom, ion, or electron (collisions with electrons are usually most important), so can ionization. The rate at which such collisional ionizations occur depends on the speeds of the atoms and hence on the temperature of the gas—the hotter the gas, the more of its atoms will be ionized.

The rate at which ions and electrons recombine also depends on their relative speeds—that is, on the temperature. In addition, it depends on the density of the gas: the higher the density, the greater the chance for recapture, because the different kinds of particles are crowded more closely together. From a knowledge of the temperature and density of a gas, it is possible to calculate the fraction of atoms that have been ionized once, twice, and so on. In the Sun, for example, we find that most of the hydrogen and helium atoms in its atmosphere are neutral, whereas most of the calcium atoms, as well as many other heavier atoms, are ionized once.

The energy levels of an ionized atom are entirely different from those of the same atom when it is neutral. Each time an electron is removed from the atom, the energy levels of the ion, and thus the wavelengths of the spectral lines it can produce, change. This helps astronomers differentiate the ions of a given element. Ionized hydrogen, having no electron, can produce no absorption lines.

This page titled 5.5: Formation of Spectral Lines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.6: Formation of Spectral Lines by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.6: The Doppler Effect

Learning Objectives

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

- Explain why the spectral lines of photons we observe from an object will change as a result of the object's motion toward or away from us
- Describe how we can use the Doppler effect to deduce how fast astronomical objects are moving through space

The last two sections introduced you to many new concepts, and we hope that through those, you have seen one major idea emerge. Astronomers can learn about the elements in stars and galaxies by decoding the information in their spectral lines. There is a complicating factor in learning how to decode the message of starlight, however. If a star is moving toward or away from us, its lines will be in a slightly different place in the spectrum from where they would be in a star at rest. And most objects in the universe do have some motion relative to the Sun.

Motion Affects Waves

In 1842, Christian Doppler first measured the effect of motion on waves by hiring a group of musicians to play on an open railroad car as it was moving along the track. He then applied what he learned to all waves, including light, and pointed out that if a light source is approaching or receding from the observer, the light waves will be, respectively, crowded more closely together or spread out. The general principle, now known as the Doppler effect, is illustrated in Figure 5.22.

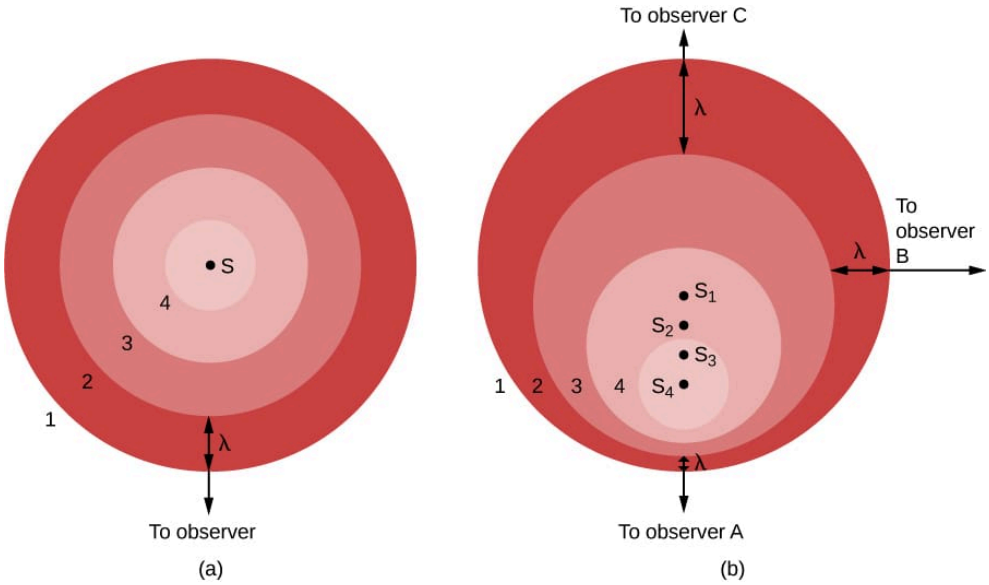


Figure 5.22 : Doppler Effect. (a) A source, S, makes waves whose numbered crests (1, 2, 3, and 4) wash over a stationary observer. (b) The source S now moves toward observer *A* and away from observer *C*. Wave crest 1 was emitted when the source was at position S1, crest 2 at position S2, and so forth. Observer *A* sees waves compressed by this motion and sees a blueshift (if the waves are light). Observer *C* sees the waves stretched out by the motion and sees a redshift. Observer *B*, whose line of sight is perpendicular to the source's motion, sees no change in the waves (and feels left out).

In part (a) of the figure, the light source (S) is at rest with respect to the observer. The source gives off a series of waves, whose crests we have labeled 1, 2, 3, and 4. The light waves spread out evenly in all directions, like the ripples from a splash in a pond. The crests are separated by a distance, λ, where λ is the wavelength. The observer, who happens to be located in the direction of the bottom of the image, sees the light waves coming nice and evenly, one wavelength apart. Observers located anywhere else would see the same thing.

On the other hand, if the source of light is moving with respect to the observer, as seen in part (b), the situation is more complicated. Between the time one crest is emitted and the next one is ready to come out, the source has moved a bit, toward the bottom of the page. From the point of view of observer *A*, this motion of the source has decreased the distance between crests—it's squeezing the crests together, this observer might say.

In part (b), we show the situation from the perspective of three observers. The source is seen in four positions, S_1, S_2, S_3, and S_4, each corresponding to the emission of one wave crest. To observer *A*, the waves seem to follow one another more closely, at a decreased wavelength and thus increased frequency. (Remember, all light waves travel at the speed of light through empty space, no matter what. This means that motion cannot affect the speed, but only the wavelength and the frequency. As the wavelength decreases, the frequency must increase. If the waves are shorter, more will be able to move by during each second.)

The situation is not the same for other observers. Let's look at the situation from the point of view of observer *C*, located opposite observer *A* in the figure. For her, the source is moving away from her location. As a result, the waves are not squeezed together but instead are spread out by the motion of the source. The crests arrive with an increased wavelength and decreased frequency. To observer *B*, in a direction at right angles to the motion of the source, no effect is observed. The wavelength and frequency remain the same as they were in part (a) of the figure.

We can see from this illustration that the Doppler effect is produced only by a motion toward or away from the observer, a motion called radial velocity. Sideways motion does not produce such an effect. Observers between *A* and *B* would observe some shortening of the light waves for that part of the motion of the source that is along their line of sight. Observers between *B* and *C* would observe lengthening of the light waves that are along their line of sight.

You may have heard the Doppler effect with sound waves. When a train whistle or police siren approaches you and then moves away, you will notice a decrease in the pitch (which is how human senses interpret sound wave frequency) of the sound waves. Compared to the waves at rest, they have changed from slightly more frequent when coming toward you, to slightly less frequent when moving away from you.

Link to Learning

A nice example of this change in the sound of a train whistle can be heard at the end of the classic Beach Boys song "Caroline, No" on their album *Pet Sounds*. To hear this sound, go to this YouTube version of the song. The sound of the train begins at approximately 2:20.

Color Shifts

When the source of waves moves toward you, the wavelength decreases a bit. If the waves involved are visible light, then the colors of the light change slightly. As wavelength decreases, they shift toward the blue end of the spectrum: astronomers call this a *blueshift* (since the end of the spectrum is really violet, the term should probably be *violetshift*, but blue is a more common color). When the source moves away from you and the wavelength gets longer, we call the change in colors a *redshift*. Because the Doppler effect was first used with visible light in astronomy, the terms "blueshift" and "redshift" became well established. Today, astronomers use these words to describe changes in the wavelengths of radio waves or X-rays as comfortably as they use them to describe changes in visible light.

The greater the motion toward or away from us, the greater the Doppler shift. If the relative motion is entirely along the line of sight, the formula for the Doppler shift of light is

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

where λ is the wavelength emitted by the source, $\Delta\lambda$ is the difference between λ and the wavelength measured by the observer, *c* is the speed of light, and *v* is the relative speed of the observer and the source in the line of sight. The variable *v* is counted as positive if the velocity is one of recession, and negative if it is one of approach. Solving this equation for the velocity, we find $v = c \times \Delta\lambda/\lambda$.

Link to Learning

Click-and-drag the object emitting waves or the object receiving waves in this simulator to experiment with the Doppler effect yourself. The plots along the top show how the detected wavelengths change if the objects are approaching each other or moving further apart.

If a star approaches or recedes from us, the wavelengths of light in its continuous spectrum appear shortened or lengthened, respectively, as do those of the dark lines. However, unless its speed is tens of thousands of kilometers per second, the star does not appear noticeably bluer or redder than normal. The Doppler shift is thus not easily detected in a continuous spectrum and cannot be measured accurately in such a spectrum. The wavelengths of the absorption lines can be measured accurately, however, and their Doppler shift is relatively simple to detect.

Example 5.6: The Doppler Effect

We can use the Doppler effect equation to calculate the radial velocity of an object if we know three things: the speed of light, the original (unshifted) wavelength of the light emitted, and the difference between the wavelength of the emitted light and the wavelength we observe. For particular absorption or emission lines, we usually know exactly what wavelength the line has in our laboratories on Earth, where the source of light is not moving. We can measure the new wavelength with our instruments at the telescope, and so we know the difference in wavelength due to Doppler shifting. Since the speed of light is a universal constant, we can then calculate the radial velocity of the star.

A particular emission line of hydrogen is originally emitted with a wavelength of 656.3 nm from a gas cloud. At our telescope, we observe the wavelength of the emission line to be 656.6 nm. How fast is this gas cloud moving toward or away from Earth?

Solution

Because the light is shifted to a longer wavelength (redshifted), we know this gas cloud is moving away from us. The speed can be calculated using the Doppler shift formula:

$$\begin{aligned} \nu &= c \times \tfrac{\Delta\lambda}{\lambda} = \left(3.0 \times 10^8\ \mathrm{m/s}\right)\left(\tfrac{0.3\ \mathrm{nm}}{656.3\ \mathrm{nm}}\right) = \left(3.0 \times 10^8\ \mathrm{m/s}\right)\left(\tfrac{0.3 \times 10^{-9}\ \mathrm{m}}{656.3 \times 10^{-9}\ \mathrm{m}}\right) \\ &= 140{,}000\ \mathrm{m/s} = 140\ \mathrm{km/s} \end{aligned}$$

Exercise 5.6.1

Suppose a spectral line of hydrogen, normally at 500 nm, is observed in the spectrum of a star to be at 500.1 nm. How fast is the star moving toward or away from Earth?

Answer available at the end of this section.

You may now be asking: if all the stars are moving and motion changes the wavelength of each spectral line, won't this be a disaster for astronomers trying to figure out what elements are present in the stars? After all, it is the precise wavelength (or color) that tells astronomers which lines belong to which element. And we first measure these wavelengths in containers of gas in our laboratories, which are not moving. If every line in a star's spectrum is now shifted by its motion to a different wavelength (color), how can we be sure which lines and which elements we are looking at in a star whose speed we do not know?

Take heart. This situation sounds worse than it really is. Astronomers rarely judge the presence of an element in an astronomical object by a single line. It is the *pattern* of lines unique to hydrogen or calcium that enables us to determine that those elements are part of the star or galaxy we are observing. The Doppler effect does not change the pattern of lines from a given element—it only shifts the whole pattern slightly toward redder or bluer wavelengths. The shifted pattern is still quite easy to recognize. Best of all, when we do recognize a familiar element's pattern, we get a bonus: the amount the pattern is shifted can enable us to determine the speed of the objects in our line of sight.

The training of astronomers includes much work on learning to decode light (and other electromagnetic radiation). A skillful "decoder" can learn the temperature of a star, what elements are in it, and even its speed in a direction toward us or away from

us. That's really an impressive amount of information for stars that are light-years away.

Answers

Exercise 5.6.1

Because the light is shifted to a longer wavelength, the star is moving away from us:
$\nu = c \times \frac{\Delta\lambda}{\lambda} = (3.0 \times 10^8\ \mathrm{m/s})\left(\frac{0.1\ \mathrm{nm}}{500\ \mathrm{nm}}\right) = (3.0 \times 10^8\ \mathrm{m/s})\left(\frac{0.1 \times 10^{-9}\ \mathrm{m}}{500 \times 10^{-9}\ \mathrm{m}}\right) = 60{,}000\ \mathrm{m/s}$. Its speed is 60,000 m/s.

This page titled 5.6: The Doppler Effect is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.7: The Doppler Effect by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.7: Telescopes

Learning Objectives

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

- Describe the three basic components of a modern system for measuring astronomical sources
- Describe the main functions of a telescope
- Describe the two basic types of visible-light telescopes and how they form images

Systems for Measuring Radiation

There are three basic components of a modern system for measuring radiation from astronomical sources. First, there is a telescope, which serves as a "bucket" for collecting visible light (or radiation at other wavelengths, as shown in (Figure 6.2). Just as you can catch more rain with a garbage can than with a coffee cup, large telescopes gather much more light than your eye can. Second, there is an instrument attached to the telescope that sorts the incoming radiation by wavelength. Sometimes the sorting is fairly crude. For example, we might simply want to separate blue light from red light so that we can determine the temperature of a star. But at other times, we want to see individual spectral lines to determine what an object is made of, or to measure its speed (as explained in the Radiation and Spectra chapter). Third, we need some type of detector, a device that senses the radiation in the wavelength regions we have chosen and permanently records the observations.

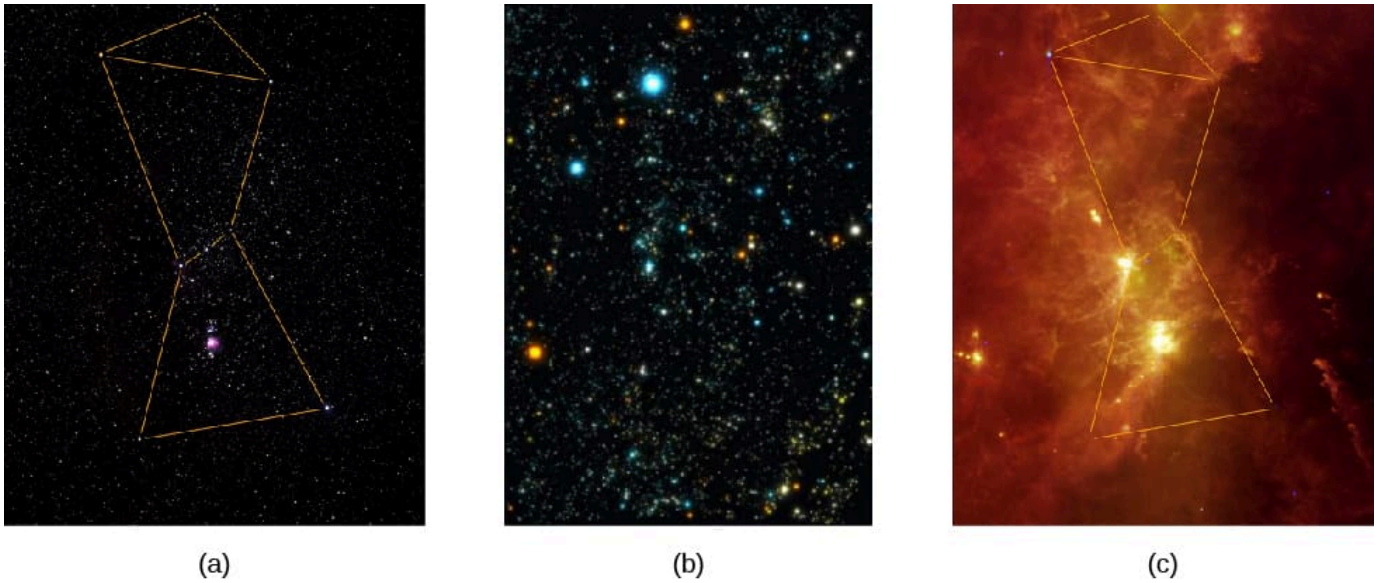


Figure 6.2 : Orion Region at Different Wavelengths. The same part of the sky looks different when observed with instruments that are sensitive to different bands of the spectrum. (a) Visible light: this shows part of the Orion region as the human eye sees it, with dotted lines added to show the figure of the mythical hunter, Orion. (b) X-rays: here, the view emphasizes the point-like X-ray sources nearby. The colors are artificial, changing from yellow to white to blue with increasing energy of the X-rays. The bright, hot stars in Orion are still seen in this image, but so are many other objects located at very different distances, including other stars, star corpses, and galaxies at the edge of the observable universe. (c) Infrared radiation: here, we mainly see the glowing dust in this region. (credit a: modification of work by Howard McCallon/NASA/IRAS; credit b: modification of work by Howard McCallon/NASA/IRAS; credit c: modification of work by Michael F. Corcoran)

The history of the development of astronomical telescopes is about how new technologies have been applied to improve the efficiency of these three basic components: the telescopes, the wavelength-sorting device, and the detectors. Let's first look at the development of the telescope.

Many ancient cultures built special sites for observing the sky (Figure 6.3). At these ancient *observatories*, they could measure the positions of celestial objects, mostly to keep track of time and date. Many of these ancient observatories had religious and

ritual functions as well. The eye was the only device available to gather light, all of the colors in the light were observed at once, and the only permanent record of the observations was made by human beings writing down or sketching what they saw.



(a)



(b)

Figure 6.3 : Two Pre-Telescopic Observatories. (a) Machu Picchu is a fifteenth century Incan site located in Peru. (b) Stonehenge, a prehistoric site (3000–2000 BCE), is located in England. (credit a: modification of work by Allard Schmidt)

While Hans Lippershey, Zaccharias Janssen, and Jacob Metius are all credited with the invention of the telescope around 1608—applying for patents within weeks of each other—it was Galileo who, in 1610, used this simple tube with lenses (which he called a spyglass) to observe the sky and gather more light than his eyes alone could. Even his small telescope—used over many nights—revolutionized ideas about the nature of the planets and the position of Earth.

How Telescopes Work

Telescopes have come a long way since Galileo's time. Now they tend to be huge devices; the most expensive cost hundreds of millions to billions of dollars. (To provide some reference point, however, keep in mind that just renovating college football stadiums typically costs hundreds of millions of dollars—with the most expensive recent renovation, at Texas A&M University's Kyle Field, costing $450 million.) The reason astronomers keep building bigger and bigger telescopes is that celestial objects—such as planets, stars, and galaxies—send much more light to Earth than any human eye (with its tiny opening) can catch, and bigger telescopes can detect fainter objects. If you have ever watched the stars with a group of friends, you know that there's plenty of starlight to go around; each of you can see each of the stars. If a thousand more people were watching, each of them would also catch a bit of each star's light. Yet, as far as you are concerned, the light not shining into your eye is wasted. It would be great if some of this "wasted" light could also be captured and brought to your eye. This is precisely what a telescope does.

The most important functions of a telescope are (1) to *collect* the faint light from an astronomical source and (2) to *focus* all the light into a point or an image. Most objects of interest to astronomers are extremely faint: the more light we can collect, the better we can study such objects. (And remember, even though we are focusing on visible light first, there are many telescopes that collect other kinds of electromagnetic radiation.)

Telescopes that collect visible radiation use a lens or mirror to gather the light. Other types of telescopes may use collecting devices that look very different from the lenses and mirrors with which we are familiar, but they serve the same function. In all types of telescopes, the light-gathering ability is determined by the area of the device acting as the light-gathering "bucket." Since most telescopes have mirrors or lenses, we can compare their light-gathering power by comparing the apertures, or diameters, of the opening through which light travels or reflects.

The amount of light a telescope can collect increases with the size of the aperture. A telescope with a mirror that is 4 meters in diameter can collect 16 times as much light as a telescope that is 1 meter in diameter. (The diameter is squared because the area of a circle equals $\frac{\pi d^2}{4}$, where d is the diameter of the circle.)

Example 6.1: Calculating the Light-Collecting Area

What is the area of a 1-m diameter telescope? A 4-m diameter one?

Solution

Using the equation for the area of a circle,

$$A =$$

the area of a 1-m telescope is

$$\frac{\pi d^2}{4} = \frac{\pi(1\,\mathrm{m})^2}{4} = 0.79\,\mathrm{m}^2$$

and the area of a 4-m telescope is

$$\frac{\pi d^2}{4} = \frac{\pi(4\,\mathrm{m})^2}{4} = 12.6\,\mathrm{m}^2$$

Exercise 5.7.1

Show that the ratio of the two areas is 16:1.

Answer available at the end of this section.

After the telescope forms an image, we need some way to detect and record it so that we can measure, reproduce, and analyze the image in various ways. Before the nineteenth century, astronomers simply viewed images with their eyes and wrote descriptions of what they saw. This was very inefficient and did not lead to a very reliable long-term record; you know from crime shows on television that eyewitness accounts are often inaccurate.

In the nineteenth century, the use of photography became widespread. In those days, photographs were a chemical record of an image on a specially treated glass plate. Today, the image is generally detected with sensors similar to those in digital cameras, recorded electronically, and stored in computers. This permanent record can then be used for detailed and quantitative studies. Professional astronomers rarely look through the large telescopes that they use for their research.

Formation of an Image by a Lens or a Mirror

Whether or not you wear glasses, you see the world through lenses; they are key elements of your eyes. A lens is a transparent piece of material that bends the rays of light passing through it. If the light rays are parallel as they enter, the lens brings them together in one place to form an image (Figure 6.4). If the curvatures of the lens surfaces are just right, all parallel rays of light (say, from a star) are bent, or *refracted*, in such a way that they converge toward a point, called the focus of the lens. At the focus, an image of the light source appears. In the case of parallel light rays, the distance from the lens to the location where the light rays focus, or image, behind the lens is called the *focal length* of the lens.

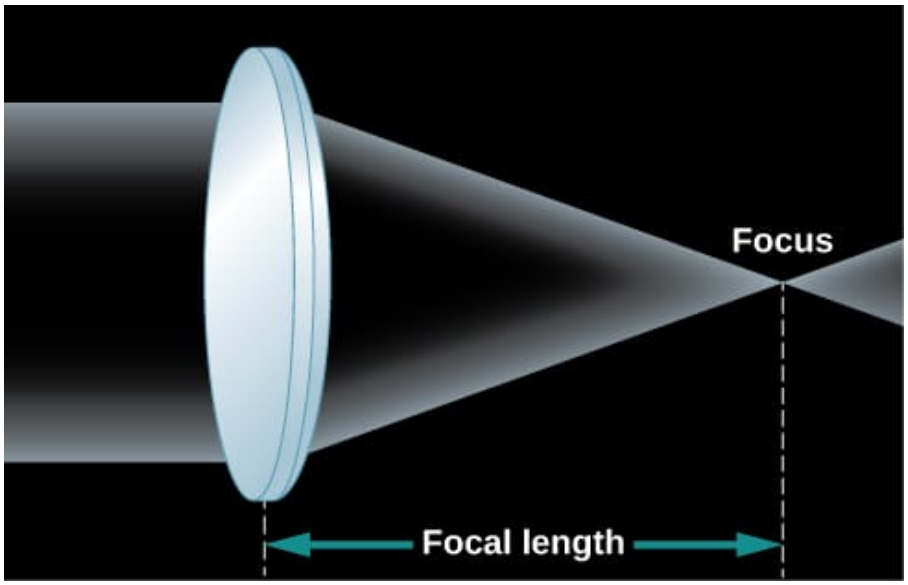


Figure 6.4 : Formation of an Image by a Simple Lens. Parallel rays from a distant source are bent by the convex lens so that they all come together in a single place (the focus) to form an image.

As you look at Figure 6.4, you may ask why two rays of light from the same star would be parallel to each other. After all, if you draw a picture of star shining in all directions, the rays of light coming from the star don't look parallel at all. But remember that the stars (and other astronomical objects) are all extremely far away. By the time the few rays of light pointed toward us actually arrive at Earth, they are, for all practical purposes, parallel to each other. Put another way, any rays that were *not* parallel to the ones pointed at Earth are now heading in some very different direction in the universe.

To view the image formed by the lens in a telescope, we use an additional lens called an eyepiece. The eyepiece focuses the image at a distance that is either directly viewable by a human or at a convenient place for a detector. Using different eyepieces, we can change the *magnification* (or size) of the image and also redirect the light to a more accessible location. Stars look like points of light, and magnifying them makes little difference, but the image of a planet or a galaxy, which has structure, can often benefit from being magnified.

Many people, when thinking of a telescope, picture a long tube with a large glass lens at one end. This design, which uses a lens as its main optical element to form an image, as we have been discussing, is known as a *refractor* (Figure 6.5), and a telescope based on this design is called a refracting telescope. Galileo's telescopes were refractors, as are today's binoculars and field glasses. However, there is a limit to the size of a refracting telescope. The largest one ever built was a 49-inch refractor built for the Paris 1900 Exposition, and it was dismantled after the Exposition. Currently, the largest refracting telescope is the 40-inch refractor at Yerkes Observatory in Wisconsin.

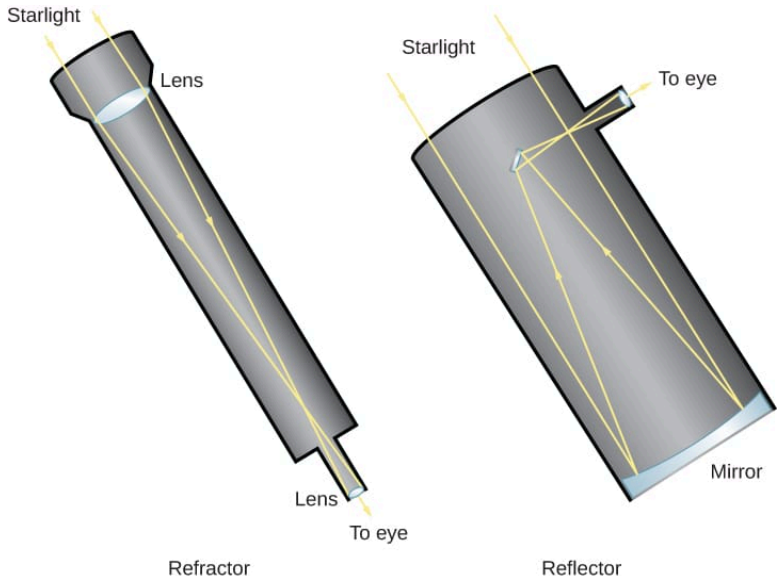


Figure 6.5 : Refracting and Reflecting Telescopes. Light enters a refracting telescope through a lens at the upper end, which focuses the light near the bottom of the telescope. An eyepiece then magnifies the image so that it can be viewed by the eye, or a detector like a photographic plate can be placed at the focus. The upper end of a reflecting telescope is open, and the light passes through to the mirror located at the bottom of the telescope. The mirror then focuses the light at the top end, where it can be detected. Alternatively, as in this sketch, a second mirror may reflect the light to a position outside the telescope structure, where an observer can have easier access to it. Professional astronomers' telescopes are more complicated than this, but they follow the same principles of reflection and refraction.

One problem with a refracting telescope is that the light must pass *through* the lens of a refractor. That means the glass must be perfect all the way through, and it has proven very difficult to make large pieces of glass without flaws and bubbles in them. Also, optical properties of transparent materials change a little bit with the wavelengths (or colors) of light, so there is some additional distortion, known as chromatic aberration. Each wavelength focuses at a slightly different spot, causing the image to appear blurry.

In addition, since the light must pass through the lens, the lens can only be supported around its edges (just like the frames of our eyeglasses). The force of gravity will cause a large lens to sag and distort the path of the light rays as they pass through it. Finally, because the light passes through it, both sides of the lens must be manufactured to precisely the right shape in order to produce a sharp image.

A different type of telescope uses a concave *primary mirror* as its main optical element. The mirror is curved like the inner surface of a sphere, and it reflects light in order to form an image (Figure 6.5). Telescope mirrors are coated with a shiny metal, usually silver, aluminum, or, occasionally, gold, to make them highly reflective. If the mirror has the correct shape, all parallel rays are reflected back to the same point, the focus of the mirror. Thus, images are produced by a mirror exactly as they are by a lens.

Telescopes designed with mirrors avoid the problems of refracting telescopes. Because the light is reflected from the front surface only, flaws and bubbles within the glass do not affect the path of the light. In a telescope designed with mirrors, only the front surface has to be manufactured to a precise shape, and the mirror can be supported from the back. For these reasons, most astronomical telescopes today (both amateur and professional) use a mirror rather than a lens to form an image; this type of telescope is called a reflecting telescope. The first successful reflecting telescope was built by Isaac Newton in 1668.

In a reflecting telescope, the concave mirror is placed at the bottom of a tube or open framework. The mirror reflects the light back up the tube to form an image near the front end at a location called the prime focus. The image can be observed at the prime focus, or additional mirrors can intercept the light and redirect it to a position where the observer can view it more easily (Figure 6.6). Since an astronomer at the prime focus can block much of the light coming to the main mirror, the use of a small *secondary mirror* allows more light to get through the system.

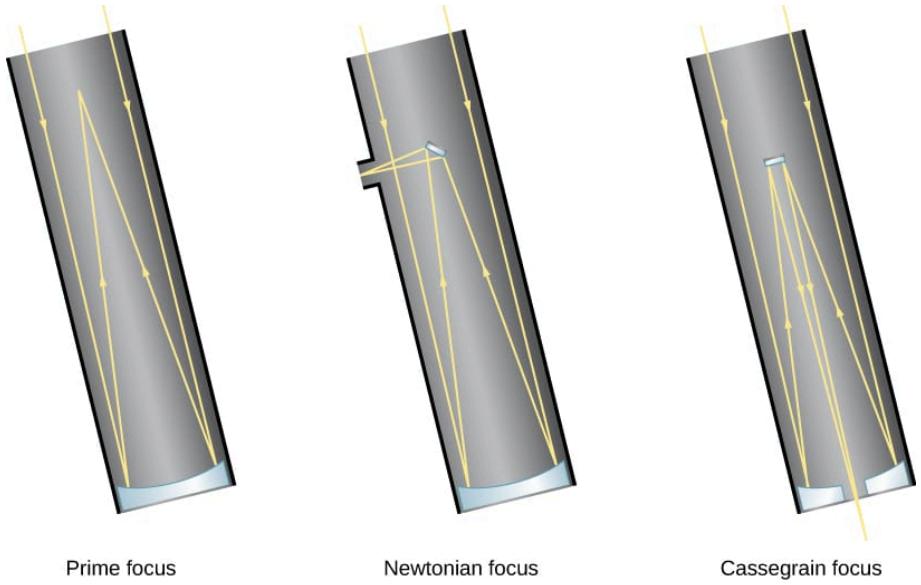


Figure 6.6 : Focus Arrangements for Reflecting Telescopes. Reflecting telescopes have different options for where the light is brought to a focus. With prime focus, light is detected where it comes to a focus after reflecting from the primary mirror. With Newtonian focus, light is reflected by a small secondary mirror off to one side, where it can be detected (see also Figure 6.5). Most large professional telescopes have a Cassegrain focus in which light is reflected by the secondary mirror down through a hole in the primary mirror to an observing station below the telescope.

Making Connections: Choosing Your Own Telescope

If the astronomy course you are taking whets your appetite for exploring the sky further, you may be thinking about buying your own telescope. Many excellent amateur telescopes are available, and some research is required to find the best model for your needs. Some good sources of information about personal telescopes are the two popular US magazines aimed at amateur astronomers: *Sky & Telescope* and *Astronomy*. Both carry regular articles with advice, reviews, and advertisements from reputable telescope dealers.

Some of the factors that determine which telescope is right for you depend upon your preferences:

- Will you be setting up the telescope in one place and leaving it there, or do you want an instrument that is portable and can come with you on outdoor excursions? How portable should it be, in terms of size and weight?
- Do you want to observe the sky with your eyes only, or do you want to take photographs? (Long-exposure photography, for example, requires a good clock drive to turn your telescope to compensate for Earth's rotation.)
- What types of objects will you be observing? Are you interested primarily in comets, planets, star clusters, or galaxies, or do you want to observe all kinds of celestial sights?

You may not know the answers to some of these questions yet. For this reason, you may want to "test-drive" some telescopes first. Most communities have amateur astronomy clubs that sponsor star parties open to the public. The members of those clubs often know a lot about telescopes and can share their ideas with you. Your instructor may know where the nearest amateur astronomy club meets; or, to find a club near you, use the websites suggested in Appendix B.

Furthermore, you may already have an instrument like a telescope at home (or have access to one through a relative or friend). Many amateur astronomers recommend starting your survey of the sky with a good pair of binoculars. These are easily carried around and can show you many objects not visible (or clear) to the unaided eye.

When you are ready to purchase a telescope, you might find the following ideas useful:

- The key characteristic of a telescope is the aperture of the main mirror or lens; when someone says they have a 6-inch or 8-inch telescope, they mean the diameter of the collecting surface. The larger the aperture, the more light you can gather, and the fainter the objects you can see or photograph.

- Telescopes of a given aperture that use lenses (refractors) are typically more expensive than those using mirrors (reflectors) because both sides of a lens must be polished to great accuracy. And, because the light passes through it, the lens must be made of high-quality glass throughout. In contrast, only the front surface of a mirror must be accurately polished.
- Magnification is not one of the criteria on which to base your choice of a telescope. As we discussed, the magnification of the image is done by a smaller eyepiece, so the magnification can be adjusted by changing eyepieces. However, a telescope will magnify not only the astronomical object you are viewing but also the turbulence of Earth's atmosphere. If the magnification is too high, your image will shimmer and shake and be difficult to view. A good telescope will come with a variety of eyepieces that stay within the range of useful magnification.
- The mount of a telescope (the structure on which it rests) is one of its most critical elements. Because a telescope shows a tiny field of view, which is magnified significantly, even the smallest vibration or jarring of the telescope can move the object you are viewing around or out of your field of view. A sturdy and stable mount is essential for serious viewing or photography (although it clearly affects how portable your telescope can be).
- A telescope requires some practice to set up and use effectively. Don't expect everything to go perfectly on your first try. Take some time to read the instructions. If a local amateur astronomy club is nearby, use it as a resource.

Answers

Exercise $5.7.1$

$\frac{12.6\,\mathrm{m}^2}{0.79\,\mathrm{m}^2} = 16$. Therefore, with 16 times the area, a 4-m telescope collects 16 times the light of a 1-m telescope.

This page titled 5.7: Telescopes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 6.2: Telescopes by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.8: Telescopes Today

Learning Objectives

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

- Recognize the largest visible-light and infrared telescopes in operation today
- Discuss the factors relevant to choosing an appropriate telescope site
- Define the technique of adaptive optics and describe the effects of the atmosphere on astronomical observations

Since Newton's time, when the sizes of the mirrors in telescopes were measured in inches, reflecting telescopes have grown ever larger. In 1948, US astronomers built a telescope with a 5-meter (200-inch) diameter mirror on Palomar Mountain in Southern California. It remained the largest visible-light telescope in the world for several decades. The giants of today, however, have primary mirrors (the largest mirrors in the telescope) that are 8- to 10-meters in diameter, and larger ones are being built (Figure 6.7).

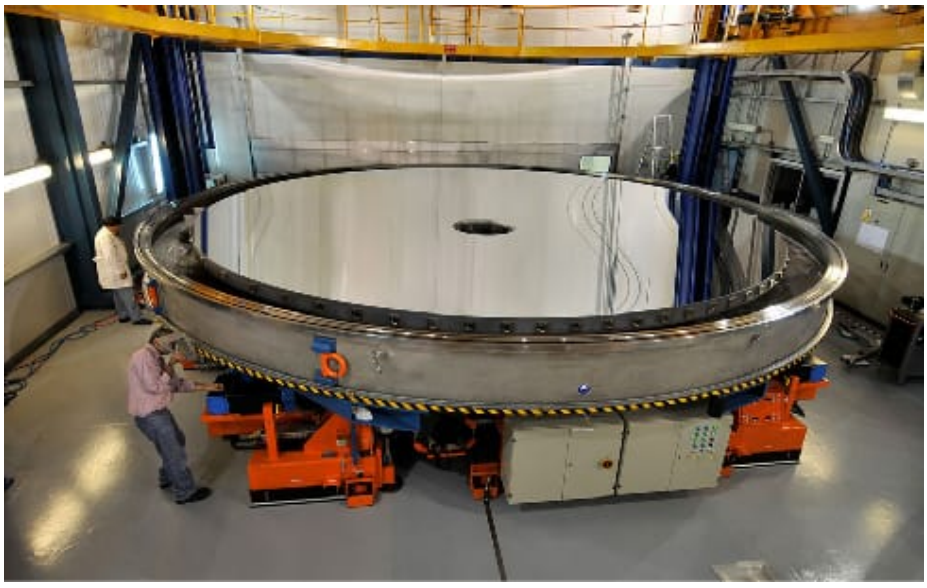


Figure 6.7 : Large Telescope Mirror. This image shows one of the primary mirrors of the European Southern Observatory's Very Large Telescope, named Yepun, just after it was recoated with aluminum. The mirror is a little over 8 meters in diameter. (credit: ESO/G. Huedepohl)

Modern Visible-Light and Infrared Telescopes

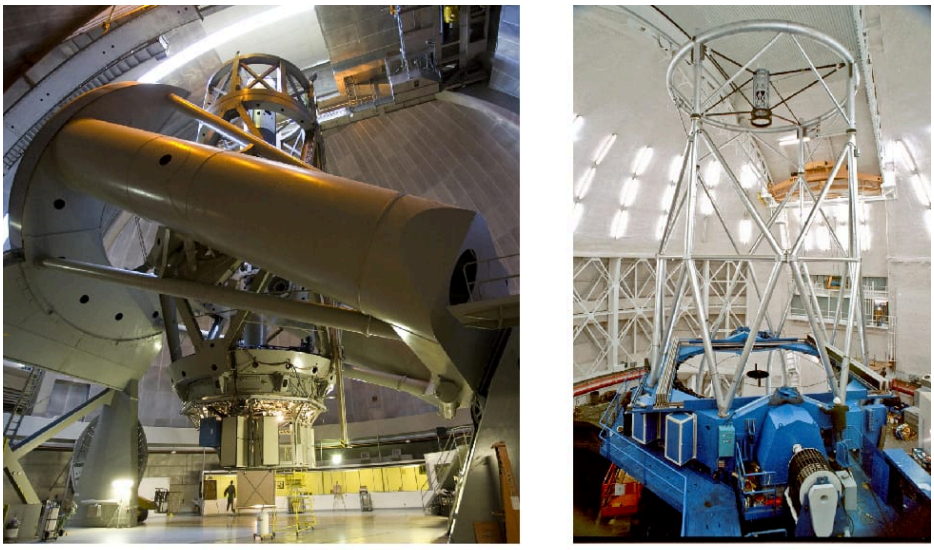
The decades starting in 1990 saw telescope building around the globe grow at an unprecedented rate. (See Table 6.1, which also includes websites for each telescope in case you want to visit or learn more about them.) Technological advancements had finally made it possible to build telescopes significantly larger than the 5-meter telescope at Palomar at a reasonable cost. New technologies have also been designed to work well in the infrared, and not just visible, wavelengths.

Table 6.1: Large Single-Dish Visible-Light and Infrared Telescopes

Aperture (m)	Telescope Name	Location	Status	Website
39	European Extremely Large Telescope (E-ELT)	Cerro Armazonas, Chile	First light 2025 (estimated)	www.eso.org/sci/facilities/eelt
30	Thirty-Meter Telescope (TMT)	Maunakea, HI	First light 2025 (estimated)	www.tmt.org

Aperture (m)	Telescope Name	Location	Status	Website
24.5	Giant Magellan Telescope (GMT)	Las Campanas Observatory, Chile	First light 2025 (estimated)	www.gmto.org
11.1×9.9	Southern African Large Telescope (SALT)	Sutherland, South Africa	2005	www.salt.ac.za
10.4	Gran Telescopio Canarias (GTC)	La Palma, Canary Islands	First light 2007	http://www.gtc.iac.es
10.0	Keck I and II (two telescopes)	Maunakea, HI	Completed 1993–96	www.keckobservatory.org
9.1	Hobby–Eberly Telescope (HET)	Mount Locke, TX	Completed 1997	www.as.utexas.edu/mcdonald/het
8.4	Large Binocular Telescope (LBT) (two telescopes)	Mount Graham, AZ	First light 2004	www.lbto.org
8.4	Large Synoptic Survey Telescope (Rubin Obs.)	The Cerro Pachón, Chile	First light 2021	www.lsst.org
8.3	Subaru Telescope	Maunakea, HI	First light 1998	www.naoj.org
8.2	Very Large Telescope (VLT)	Cerro Paranal, Chile	All four telescopes completed 2000	www.eso.org/public/teles-instr/paranal
8.1	Gemini North and Gemini South	Maunakea, HI (North) and Cerro Pachón, Chile (South)	First light 1999 (North), First light 2000 (South)	www.gemini.edu
6.5	Magellan Telescopes (two telescopes: Baade and Landon Clay)	Las Campanas, Chile	First light 2000 and 2002	obs.carnegiescience.edu/Magellan
6.5	Multi-Mirror Telescope (MMT)	Mount Hopkins, AZ	Completed 1979	www.mmto.org
6.0	Big Telescope Altazimuth (BTA-6)	Mount Pastukhov, Russia	Completed 1976	w0.sao.ru/Doc-en/Telescopes/bta/descrip.html
5.1	Hale Telescope	Mount Palomar, CA	Completed 1948	www.astro.caltech.edu/palomar/about/telescopes/hale.html

The differences between the Palomar telescope and the modern Gemini North telescope (to take an example) are easily seen in Figure 6.8. The Palomar telescope is a massive steel structure designed to hold the 14.5-ton primary mirror with a 5-meter diameter. Glass tends to sag under its own weight; hence, a huge steel structure is needed to hold the mirror. A mirror 8 meters in diameter, the size of the Gemini North telescope, if it were built using the same technology as the Palomar telescope, would have to weigh at least eight times as much and would require an enormous steel structure to support it.



(a) (b)

Figure 6.8 : Modern Reflecting Telescopes. (a) The Palomar 5-meter reflector: The Hale telescope on Palomar Mountain has a complex mounting structure that enables the telescope (in the open “tube” pointing upward in this photo) to swing easily into any position. (b) The Gemini North 8-meter telescope: The Gemini North mirror has a larger area than the Palomar mirror, but note how much less massive the whole instrument seems. (credit a: modification of work by Caltech/Palomar Observatory; credit b: modification of work by Gemini Observatory/AURA)

The 8-meter Gemini North telescope looks like a featherweight by contrast, and indeed it is. The mirror is only about 8 inches thick and weighs 24.5 tons, less than twice as much as the Palomar mirror. The Gemini North telescope was completed about 50 years after the Palomar telescope. Engineers took advantage of new technologies to build a telescope that is much lighter in weight relative to the size of the primary mirror. The Gemini mirror does sag, but with modern computers, it is possible to measure that sag many times each second and apply forces at 120 different locations to the back of the mirror to correct the sag, a process called *active control*. Seventeen telescopes with mirrors 6.5 meters in diameter and larger have been constructed since 1990.

The twin 10-meter Keck telescopes on Maunakea, which were the first of these new-technology instruments, use precision control in an entirely novel way. Instead of a single primary mirror 10 meters in diameter, each Keck telescope achieves its larger aperture by combining the light from 36 separate hexagonal mirrors, each 1.8 meters wide (Figure 6.9). Computer-controlled actuators (motors) constantly adjust these 36 mirrors so that the overall reflecting surface acts like a single mirror with just the right shape to collect and focus the light into a sharp image.

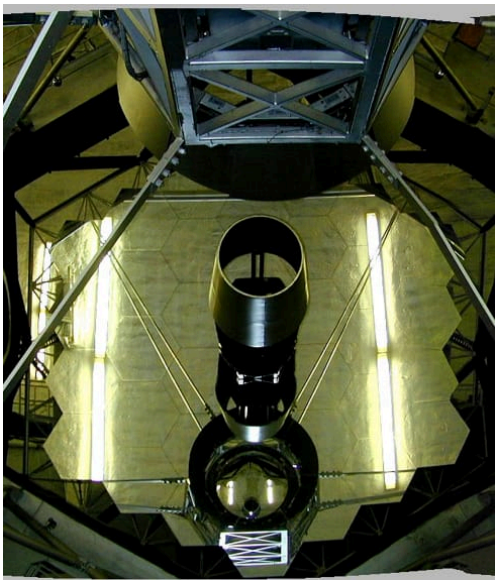


Figure 6.9 : Thirty-Six Eyes Are Better Than One. The mirror of the 10-meter Keck telescope is composed of 36 hexagonal sections. (credit: NASA)

Link to Learning

Learn more about the Keck Observatory on Maunakea through this History Channel clip on the telescopes and the work that they do.

In addition to holding the mirror, the steel structure of a telescope is designed so that the entire telescope can be pointed quickly toward any object in the sky. Since Earth is rotating, the telescope must have a motorized drive system that moves it very smoothly from east to west at exactly the same rate that Earth is rotating from west to east, so it can continue to point at the object being observed. All this machinery must be housed in a dome to protect the telescope from the elements. The dome has an opening in it that can be positioned in front of the telescope and moved along with it, so that the light from the objects being observed is not blocked.

Voyagers in Astronomy: George Ellery Hale: Master Telescope Builder

George Ellery Hale (Figure 6.10) was a giant among early telescope builders. Not once, but four times, he initiated projects that led to the construction of what was the world's largest telescope at the time. And he was a master at winning over wealthy benefactors to underwrite the construction of these new instruments.



Figure 6.10 : George Ellery Hale (1868–1938). Hale's work led to the construction of several major telescopes, including the 40-inch refracting telescope at Yerkes Observatory, and three reflecting telescopes: the 60-inch Hale and 100-inch Hooker telescopes at Mount Wilson Observatory, and the 200-inch Hale Telescope at Palomar Observatory.

Hale's training and early research were in solar physics. In 1892, at age 24, he was named associate professor of astral physics and director of the astronomical observatory at the University of Chicago. At the time, the largest telescope in the world was the 36-inch refractor at the Lick Observatory near San Jose, California. Taking advantage of an existing glass blank for a 40-inch telescope, Hale set out to raise money for a larger telescope than the one at Lick. One prospective donor was Charles T. Yerkes, who, among other things, ran the trolley system in Chicago.

Hale wrote to Yerkes, encouraging him to support the construction of the giant telescope by saying that "the donor could have no more enduring monument. It is certain that Mr. Lick's name would not have been nearly so widely known today were it not for the famous observatory established as a result of his munificence." Yerkes agreed, and the new telescope was completed in May 1897; it remains the largest refractor in the world (Figure 6.11).



Figure 6.11 : World's Largest Refractor. The Yerkes 40-inch (1-meter) telescope.

Even before the completion of the Yerkes refractor, Hale was not only dreaming of building a still larger telescope but was also taking concrete steps to achieve that goal. In the 1890s, there was a major controversy about the relative quality of refracting and reflecting telescopes. Hale realized that 40 inches was close to the maximum feasible aperture for refracting telescopes. If telescopes with significantly larger apertures were to be built, they would have to be reflecting telescopes.

Using funds borrowed from his own family, Hale set out to construct a 60-inch reflector. For a site, he left the Midwest for the much better conditions on Mount Wilson—at the time, a wilderness peak above the small city of Los Angeles. In 1904, at the age of 36, Hale received funds from the Carnegie Foundation to establish the Mount Wilson Observatory. The 60-inch mirror was placed in its mount in December 1908.

Two years earlier, in 1906, Hale had already approached John D. Hooker, who had made his fortune in hardware and steel pipe, with a proposal to build a 100-inch telescope. The technological risks were substantial. The 60-inch telescope was not yet complete, and the usefulness of large reflectors for astronomy had yet to be demonstrated. George Ellery Hale's brother called him "the greatest gambler in the world." Once again, Hale successfully obtained funds, and the 100-inch telescope was completed in November 1917. (It was with this telescope that Edwin Hubble was able to establish that the spiral nebulae were separate islands of stars—or galaxies—quite removed from our own Milky Way.)

Hale was not through dreaming. In 1926, he wrote an article in *Harper's Magazine* about the scientific value of a still larger telescope. This article came to the attention of the Rockefeller Foundation, which granted $6 million for the construction of a 200-inch telescope. Hale died in 1938, but the 200-inch (5-meter) telescope on Palomar Mountain was dedicated 10 years later and is now named in Hale's honor.

Picking the Best Observing Sites

A telescope like the Gemini or Keck telescope costs about $100 million to build. That kind of investment demands that the telescope be placed in the best possible site. Since the end of the nineteenth century, astronomers have realized that the best observatory sites are on mountains, far from the lights and pollution of cities. Although a number of urban observatories remain, especially in the large cities of Europe, they have become administrative centers or museums. The real action takes place far away, often on desert mountains or isolated peaks in the Atlantic and Pacific Oceans, where we find the staff's living quarters, computers, electronic and machine shops, and of course the telescopes themselves. A large observatory today requires a supporting staff of 20 to 100 people in addition to the astronomers.

The performance of a telescope is determined not only by the size of its mirror but also by its location. Earth's atmosphere, so vital to life, presents challenges for the observational astronomer. In at least four ways, our air imposes limitations on the usefulness of telescopes:

1. The most obvious limitation is weather conditions such as clouds, wind, and rain. At the best sites, the weather is clear as much as 75% of the time.
2. Even on a clear night, the atmosphere filters out a certain amount of starlight, especially in the infrared, where the absorption is due primarily to water vapor. Astronomers therefore prefer dry sites, generally found at high altitudes.
3. The sky above the telescope should be dark. Near cities, the air scatters the glare from lights, producing an illumination that hides the faintest stars and limits the distances that can be probed by telescopes. (Astronomers call this effect light pollution.) Observatories are best located at least 100 miles from the nearest large city.
4. Finally, the air is often unsteady; light passing through this turbulent air is disturbed, resulting in blurred star images. Astronomers call these effects "bad seeing." When seeing is bad, images of celestial objects are distorted by the constant twisting and bending of light rays by turbulent air.

The best observatory sites are therefore high, dark, and dry. The world's largest telescopes are found in such remote mountain locations as the Andes Mountains of Chile (Figure 6.12), the desert peaks of Arizona, the Canary Islands in the Atlantic Ocean, and Maunakea in Hawaii, a dormant volcano with an altitude of 13,700 feet (4200 meters).

Link to Learning

Light pollution is a problem not just for professional astronomers but for everyone who wants to enjoy the beauty of the night sky. In addition research is now showing that it can disrupt the life cycle of animals with whom we share the urban and suburban landscape. And the light wasted shining into the sky leads to unnecessary municipal expenses and use of fossil fuels. Concerned people have formed an organization, the International Dark-Sky Association, whose website is full of good information. A citizen science project called *Globe at Night* allows you to measure the light levels in your community by counting stars and to compare it to others around the world. Recently, a new threat to dark skies has emerged. A number of commercial companies are planning and launching large swarms of satellites into low-Earth orbit to provide internet access around the world. Unfortunately, these satellites reflect sunlight and have already begun to ruin astronomical photographs

the world over. See the Further Exploration section at the end of this chapter for more resources on this topic. If you get interested in this topic and want to do a paper for your astronomy course or another course while you are in college, the *Dark Night Skies* guide can point you to a variety of resources on the topic.



Figure 6.12 : High and Dry Site. Cerro Paranal, a mountain summit 2.7 kilometers above sea level in Chile's Atacama Desert, is the site of the European Southern Observatory's Very Large Telescope. This photograph shows the four 8-meter telescope buildings on the site and vividly illustrates that astronomers prefer high, dry sites for their instruments. The 4.1-meter Visible and Infrared Survey Telescope for Astronomy (VISTA) can be seen in the distance on the next mountain peak. (credit: ESO)

The Resolution of a Telescope

In addition to gathering as much light as they can, astronomers also want to have the sharpest images possible. Resolution refers to the precision of detail present in an image: that is, the smallest features that can be distinguished. Astronomers are always eager to make out more detail in the images they study, whether they are following the weather on Jupiter or trying to peer into the violent heart of a "cannibal galaxy" that recently ate its neighbor for lunch.

One factor that determines how good the resolution will be is the size of the telescope. Larger apertures produce sharper images. Until very recently, however, visible-light and infrared telescopes on Earth's surface could not produce images as sharp as the theory of light said they should.

The problem—as we saw earlier in this chapter—is our planet's atmosphere, which is turbulent. It contains many small-scale blobs or cells of gas that range in size from inches to several feet. Each cell has a slightly different temperature from its neighbor, and each cell acts like a lens, bending (refracting) the path of the light by a small amount. This bending slightly changes the position where each light ray finally reaches the detector in a telescope. The cells of air are in motion, constantly being blown through the light path of the telescope by winds, often in different directions at different altitudes. As a result, the path followed by the light is constantly changing.

For an analogy, think about watching a parade from a window high up in a skyscraper. You decide to throw some confetti down toward the marchers. Even if you drop a handful all at the same time and in the same direction, air currents will toss the pieces around, and they will reach the ground at different places. As we described earlier, we can think of the light from the stars as a series of parallel beams, each making its way through the atmosphere. Each path will be slightly different, and each will reach the detector of the telescope at a slightly different place. The result is a blurred image, and because the cells are being blown by the wind, the nature of the blur will change many times each second. You have probably noticed this effect as the "twinkling" of stars seen from Earth. The light beams are bent enough that part of the time they reach your eye, and part of the time some of them miss, thereby making the star seem to vary in brightness. In space, however, the light of the stars is steady.

Astronomers search the world for locations where the amount of atmospheric blurring, or turbulence, is as small as possible. It turns out that the best sites are in coastal mountain ranges and on isolated volcanic peaks in the middle of an ocean. Air that has

flowed long distances over water before it encounters land is especially stable.

The resolution of an image is measured in units of angle on the sky, typically in units of arcseconds. One arcsecond is 1/3600 degree, and there are 360 degrees in a full circle. So we are talking about tiny angles on the sky. To give you a sense of just how tiny, we might note that 1 arcsecond is how big a quarter would look when seen from a distance of 5 kilometers. The best images obtained from the ground with traditional techniques reveal details as small as several tenths of an arcsecond across. This image size is remarkably good. One of the main reasons for launching the Hubble Space Telescope was to escape Earth's atmosphere and obtain even sharper images.

But since we can't put every telescope into space, astronomers have devised a technique called adaptive optics that can beat Earth's atmosphere at its own game of blurring. This technique (which is most effective in the infrared region of the spectrum with our current technology) makes use of a small flexible mirror placed in the beam of a telescope. A sensor measures how much the atmosphere has distorted the image, and as often as 500 times per second, it sends instructions to the flexible mirror on how to change shape in order to compensate for distortions produced by the atmosphere. The light is thus brought back to an almost perfectly sharp focus at the detector. Figure 6.13 shows just how effective this technique is. With adaptive optics, ground-based telescopes can achieve resolutions of 0.1 arcsecond or a little better in the infrared region of the spectrum. This impressive figure is the equivalent of the resolution that the Hubble Space Telescope achieves in the visible-light region of the spectrum.

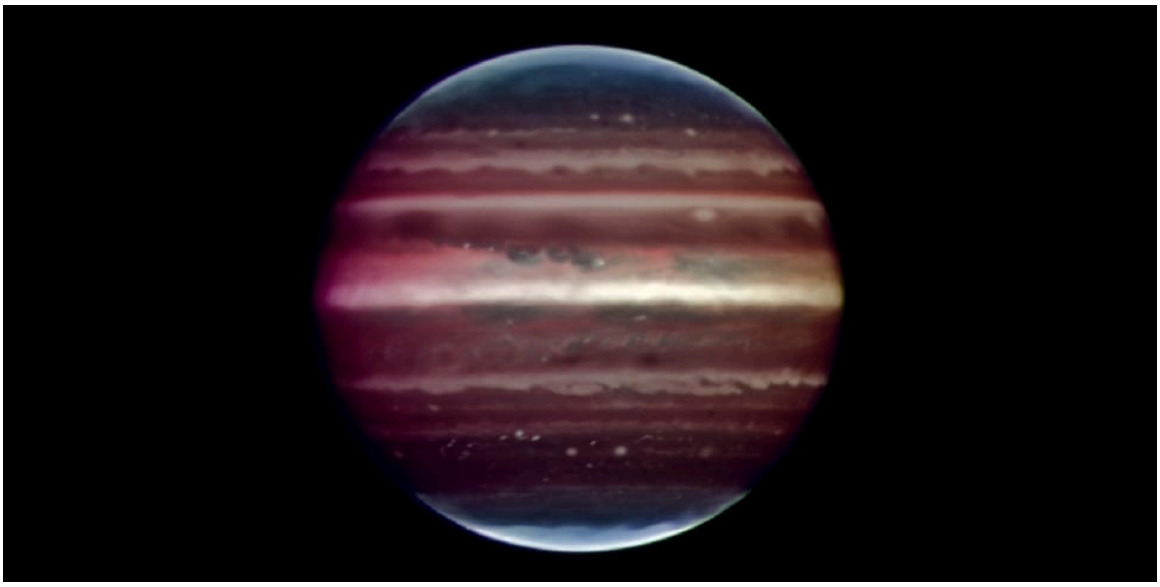


Figure 6.13 : Power of Adaptive Optics. One of the clearest pictures of Jupiter ever taken from the ground, this image was produced with adaptive optics using an 8-meter-diameter telescope at the Very Large Telescope in Chile. Adaptive optics uses infrared wavelengths to remove atmospheric blurring, resulting in a much clearer image. (credit: modification of work by ESO, F.Marchis, M.Wong (UC Berkeley); E.Marchetti, P.Amico, S.Tordo (ESO))

Astronomy Basics: How Astronomers Really Use Telescopes

In the popular view (and some bad movies), an astronomer spends most nights in a cold observatory peering through a telescope, but this is not very accurate today. Most astronomers do not live at observatories, but near the universities or laboratories where they work. An astronomer might spend only a week or so each year observing at the telescope and the rest of the time measuring or analyzing the data acquired from large project collaborations and dedicated surveys. Many astronomers use radio telescopes for space experiments, which work just as well during the daylight hours. Still others work at purely theoretical problems using supercomputers and never observe at a telescope of any kind.

Even when astronomers are observing with large telescopes, they seldom peer through them. Electronic detectors permanently record the data for detailed analysis later. At some observatories, observations may be made remotely, with the astronomer sitting at a computer thousands of miles away from the telescope.

Time on major telescopes is at a premium, and an observatory director will typically receive many more requests for telescope time than can be accommodated during the year. Astronomers must therefore write a convincing proposal explaining how they would like to use the telescope and why their observations will be important to the progress of astronomy. A committee

of astronomers is then asked to judge and rank the proposals, and time is assigned only to those with the greatest merit. Even if your proposal is among the high-rated ones, you may have to wait many months for your turn. If the skies are cloudy on the nights you have been assigned, it may be more than a year before you get another chance.

Some older astronomers still remember long, cold nights spent alone in an observatory dome, with only music from a tape recorder or an all-night radio station for company. The sight of the stars shining brilliantly hour after hour through the open slit in the observatory dome was unforgettable. So, too, was the relief as the first pale light of dawn announced the end of a 12-hour observation session. Astronomy is much easier today, with teams of observers working together, often at their computers, in a warm room. Those who are more nostalgic, however, might argue that some of the romance has gone from the field, too.

This page titled 5.8: Telescopes Today is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 6.3: Telescopes Today by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.9: Visible-Light Detectors and Instruments

Learning Objectives

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

- Describe the difference between photographic plates and charge-coupled devices
- Describe the unique difficulties associated with infrared observations and their solutions
- Describe how a spectrometer works

After a telescope collects radiation from an astronomical source, the radiation must be *detected* and measured. The first detector used for astronomical observations was the human eye, but it suffers from being connected to an imperfect recording and retrieving device—the human brain. Photography and modern electronic detectors have eliminated the quirks of human memory by making a permanent record of the information from the cosmos.

The eye also suffers from having a very short *integration time*; it takes only a fraction of a second to add light energy together before sending the image to the brain. One important advantage of modern detectors is that the light from astronomical objects can be collected by the detector over longer periods of time; this technique is called "taking a long exposure." Exposures of several hours are required to detect very faint objects in the cosmos.

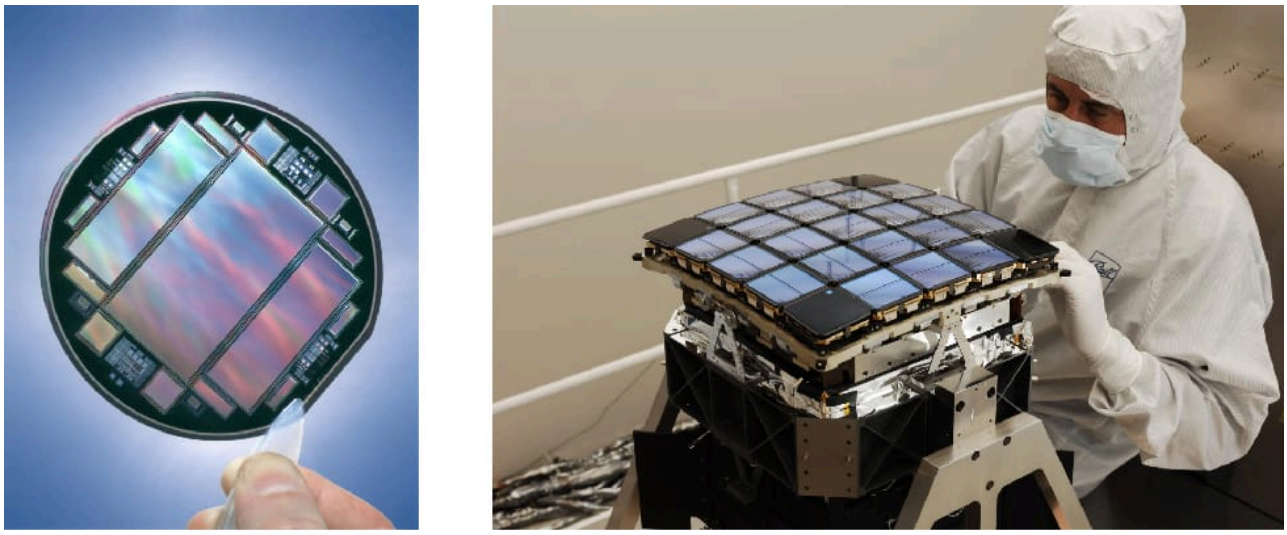
Before the light reaches the detector, astronomers today normally use some type of instrument to sort the light according to wavelength. The instrument may be as simple as colored filters, which transmit light within a specified range of wavelengths. A red transparent plastic is an everyday example of a filter that transmits only the red light and blocks the other colors. After the light passes through a filter, it forms an image that astronomers can then use to measure the apparent brightness and color of objects. We will show you many examples of such images in the later chapters of this book, and we will describe what we can learn from them.

Alternatively, the instrument between telescope and detector may be one of several devices that spread the light out into its full rainbow of colors so that astronomers can measure individual lines in the spectrum. Such an instrument (which you learned about in the chapter on Radiation and Spectra) is called a *spectrometer* because it allows astronomers to measure (to meter) the spectrum of a source of radiation. Whether a filter or a spectrometer, both types of wavelength-sorting instruments still have to use detectors to record and measure the properties of light.

Photographic and Electronic Detectors

Throughout most of the twentieth century, photographic film or *glass plates* served as the prime astronomical detectors, whether for photographing spectra or direct images of celestial objects. In a photographic plate, a light-sensitive chemical coating is applied to a piece of glass that, when developed, provides a lasting record of the image. At observatories around the world, vast collections of photographs preserve what the sky has looked like during the past 100 years. Photography represents a huge improvement over the human eye, but it still has limitations. Photographic films are inefficient: only about 1% of the light that actually falls on the film contributes to the chemical change that makes the image; the rest is wasted.

Astronomers today have much more efficient electronic detectors to record astronomical images. Most often, these are charge-coupled devices (CCDs), which are similar to the detectors used in video camcorders or in digital cameras (like the one more and more students have on their cell phones) (see Figure 6.14). In a CCD, photons of radiation hitting any part of the detector generate a stream of charged particles (electrons) that are stored and counted at the end of the exposure. Each place where the radiation is counted is called a pixel (picture element), and modern detectors can count the photons in millions of pixels (megapixels, or MPs).



(a) (b)

Figure 6.14 : Charge-Coupled Devices (CCDs). (a) This CCD is a mere 300-micrometers thick (thinner than a human hair) yet holds more than 21 million pixels. (b) This matrix of 42 CCDs served the Kepler telescope. (credit a: modification of work by US Department of Energy; credit b: modification of work by NASA and Ball Aerospace)

Because CCDs typically record as much as 60–70% of all the photons that strike them, and the best silicon and infrared CCDs exceed 90% sensitivity, we can detect much fainter objects. Among these are many small moons around the outer planets, icy dwarf planets beyond Pluto, and dwarf galaxies of stars. CCDs also provide more accurate measurements of the brightness of astronomical objects than photography, and their output is digital—in the form of numbers that can go directly into a computer for analysis.

Infrared Observations

Observing the universe in the infrared band of the spectrum presents some additional challenges. The infrared region extends from wavelengths near 1 micrometer (μm), which is about the long wavelength sensitivity limit of both CCDs and photography, to 100 micrometers or longer. Recall from the discussion on radiation and spectra that infrared is “heat radiation” (given off at temperatures that we humans are comfortable with). The main challenge to astronomers using infrared is to distinguish between the tiny amount of heat radiation that reaches Earth from stars and galaxies, and the much greater heat radiated by the telescope itself and our planet’s atmosphere.

Typical temperatures on Earth’s surface are near 300 K, and the atmosphere through which observations are made is only a little cooler. According to Wien’s law (from the chapter on Radiation and Spectra), the telescope, the observatory, and even the sky are radiating infrared energy with a peak wavelength of about 10 micrometers. To infrared eyes, everything on Earth is brightly aglow—including the telescope and camera (Figure 6.15). The challenge is to detect faint cosmic sources against this sea of infrared light. Another way to look at this is that an astronomer using infrared must always contend with the situation that a visible-light observer would face if working in broad daylight with a telescope and optics lined with bright fluorescent lights.



Figure 6.15 : Infrared Eyes. Infrared waves can penetrate places in the universe from which light is blocked, as shown in this infrared image where the plastic bag blocks visible light but not infrared. (credit: NASA/JPL-Caltech/R. Hurt (SSC))

To solve this problem, astronomers must protect the infrared detector from nearby radiation, just as you would shield photographic film from bright daylight. Since anything warm radiates infrared energy, the detector must be isolated in very cold surroundings; often, it is held near absolute zero (1 to 3 K) by immersing it in liquid helium. The second step is to reduce the radiation emitted by the telescope structure and optics, and to block this heat from reaching the infrared detector.

Link to Learning

Check out The Infrared Zoo to get a sense of what familiar objects look like with infrared radiation. Slide the slider to change the wavelength of radiation for the picture, and click the arrow to see other animals.

Spectroscopy

Spectroscopy is one of the astronomer's most powerful tools, providing information about the composition, temperature, motion, and other characteristics of celestial objects. More than half of the time spent on most large telescopes is used for spectroscopy.

The many different wavelengths present in light can be separated by passing them through a spectrometer to form a spectrum. The design of a simple spectrometer is illustrated in Figure 6.16. Light from the source (actually, the image of a source produced by the telescope) enters the instrument through a small hole or narrow slit, and is collimated (made into a beam of parallel rays) by a lens. The light then passes through a prism, producing a spectrum: different wavelengths leave the prism in different directions because each wavelength is bent by a different amount when it enters and leaves the prism. A second lens placed behind the prism focuses the many different images of the slit or entrance hole onto a CCD or other detecting device. This collection of images (spread out by color) is the spectrum that astronomers can then analyze at a later point. As spectroscopy spreads the light out into more and more collecting bins, fewer photons go into each bin, so either a larger telescope is needed or the integration time must be greatly increased—usually both.

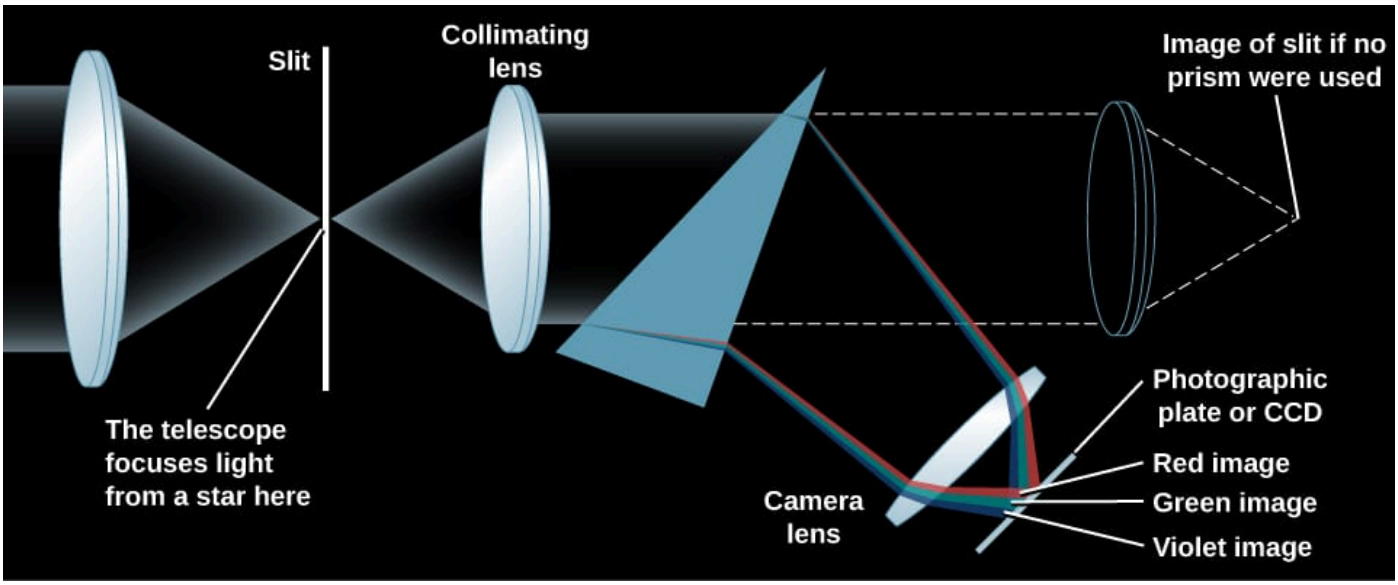


Figure 6.16 : Prism Spectrometer. The light from the telescope is focused on a slit. A prism (or grating) disperses the light into a spectrum, which is then photographed or recorded electronically.

In practice, astronomers today are more likely to use a different device, called a *grating*, to disperse the spectrum. A grating is a piece of material with thousands of grooves on its surface. While it functions completely differently, a grating, like a prism, also spreads light out into a spectrum.

This page titled 5.9: Visible-Light Detectors and Instruments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 6.4: Visible-Light Detectors and Instruments by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.10: Radio Telescopes

Learning Objectives

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

- Describe how radio waves from space are detected
- Identify the world's largest radio telescopes
- Define the technique of interferometry and discuss the benefits of interferometers over single-dish telescopes

In addition to visible and infrared radiation, radio waves from astronomical objects can also be detected from the surface of Earth. In the early 1930s, Karl G. Jansky, an engineer at Bell Telephone Laboratories, was experimenting with antennas for long-range radio communication when he encountered some mysterious static—radio radiation coming from an unknown source (Figure 6.17). He discovered that this radiation came in strongest about four minutes earlier on each successive day and correctly concluded that since Earth's sidereal rotation period (how long it takes us to rotate relative to the stars) is four minutes shorter than a solar day, the radiation must be originating from some region fixed on the celestial sphere. Subsequent investigation showed that the source of this radiation was part of the Milky Way Galaxy; Jansky had discovered the first source of cosmic radio waves.

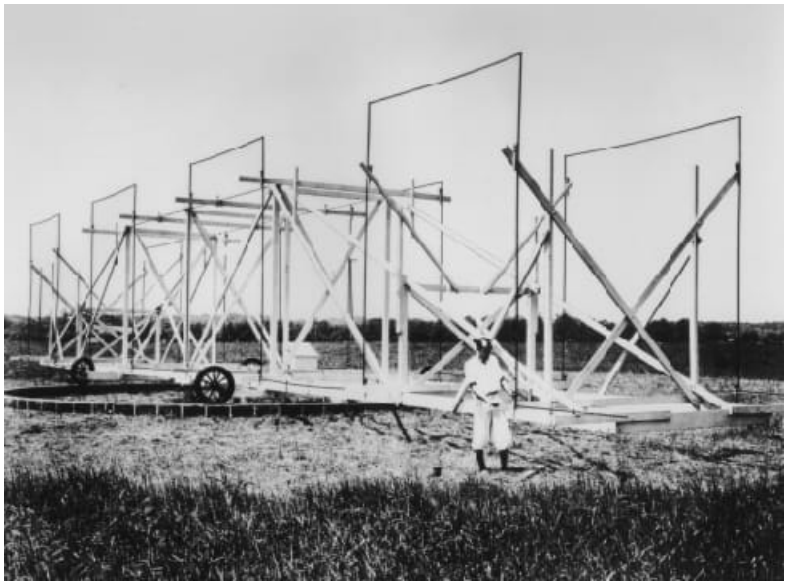


Figure 6.17 : First Radio Telescope. This rotating radio antenna was used by Jansky in his serendipitous discovery of radio radiation from the Milky Way.

In 1936, Grote Reber, who was an amateur astronomer interested in radio communications, used galvanized iron and wood to build the first antenna specifically designed to receive cosmic radio waves. Over the years, Reber built several such antennas and used them to carry out pioneering surveys of the sky for celestial radio sources; he remained active in radio astronomy for more than 30 years. During the first decade, he worked practically alone because professional astronomers had not yet recognized the vast potential of radio astronomy.

Detection of Radio Energy from Space

It is important to understand that radio waves cannot be "heard": they are not the sound waves you hear coming out of the radio receiver in your home or car. Like light, radio waves are a form of electromagnetic radiation, but unlike light, we cannot detect them with our senses—we must rely on electronic equipment to pick them up. In commercial radio broadcasting, we encode sound information (music or a newscaster's voice) into radio waves. These must be decoded at the other end and then turned back into sound by speakers or headphones.

The radio waves we receive from space do not, of course, have music or other program information encoded in them. If cosmic radio signals were translated into sound, they would sound like the static you hear when scanning between stations. Nevertheless, there is information in the radio waves we receive—information that can tell us about the chemistry and physical conditions of the sources of the waves.

Just as vibrating charged particles can produce electromagnetic waves (see the Radiation and Spectra chapter), electromagnetic waves can make charged particles move back and forth. Radio waves can produce a current in conductors of electricity such as metals. An antenna is such a conductor: it intercepts radio waves, which create a feeble current in it. The current is then amplified in a radio receiver until it is strong enough to measure or record. Like your television or radio, receivers can be tuned to select a single frequency (channel). In astronomy, however, it is more common to use sophisticated data-processing techniques that allow thousands of separate frequency bands to be detected simultaneously. Thus, the astronomical radio receiver operates much like a spectrometer on a visible-light or infrared telescope, providing information about how much radiation we receive at each wavelength or frequency. After computer processing, the radio signals are recorded on magnetic disks for further analysis.

Radio waves are reflected by conducting surfaces, just as light is reflected from a shiny metallic surface, and according to the same laws of optics. A radio-reflecting telescope consists of a concave metal reflector (called a *dish*), analogous to a telescope mirror. The radio waves collected by the dish are reflected to a focus, where they can then be directed to a receiver and analyzed. Because humans are such visual creatures, radio astronomers often construct a pictorial representation of the radio sources they observe. Figure 6.18 shows such a radio image of a distant galaxy, where radio telescopes reveal vast jets and complicated regions of radio emissions that are completely invisible in photographs taken with light.

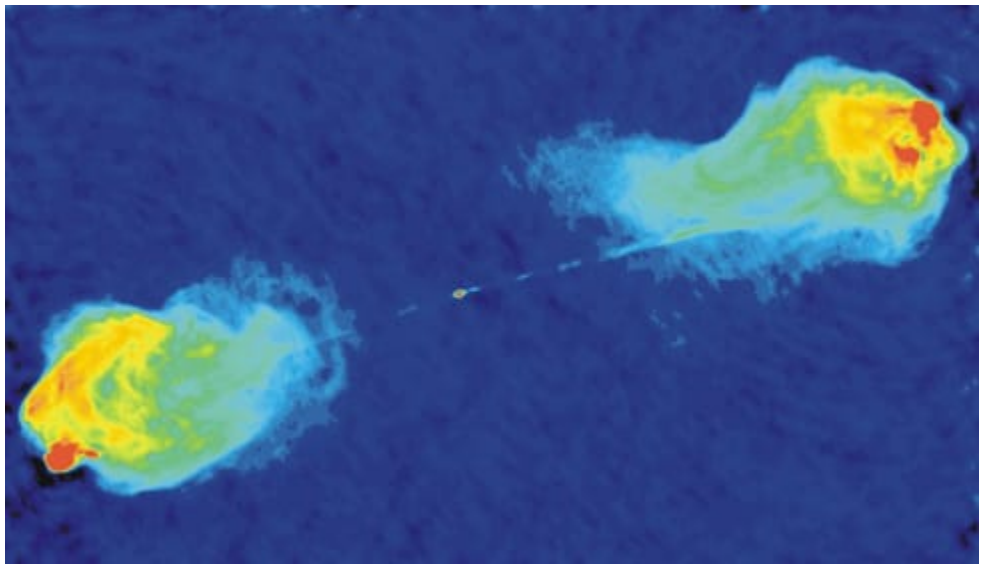


Figure 6.18 : Radio Image. This image has been constructed of radio observations at the Very Large Array of a galaxy called Cygnus A. Colors have been added to help the eye sort out regions of different radio intensities. Red regions are the most intense, blue the least. The visible galaxy would be a small dot in the center of the image. The radio image reveals jets of expelled material (more than 160,000 light-years long) on either side of the galaxy. (credit: NRAO/AUI)

Radio astronomy is a young field compared with visible-light astronomy, but it has experienced tremendous growth in recent decades. The world's largest radio reflectors that can be pointed to any direction in the sky have apertures of 100 meters. One of these has been built at the US National Radio Astronomy Observatory in West Virginia (Figure 6.19). Table 6.2 lists some of the major radio telescopes of the world.



Figure 6.19 : Robert C. Byrd Green Bank Telescope. This fully steerable radio telescope in West Virginia went into operation in August 2000. Its dish is about 100 meters across. (credit: modification of work by “b3nscott”/Flickr)

Table 6.2: Major Radio Observatories of the World

Observatory	Location	Description	Website
Individual Radio Dishes			
Five-hundred-meter Aperture Spherical radio Telescope (FAST)	Guizhou, China	500-m fixed dish	fast.bao.ac.cn/en/
Arecibo Observatory	Arecibo, Puerto Rico	305-m fixed dish	www.naic.edu
Green Bank Telescope (GBT)	Green Bank, WV	110 × 100-m steerable dish	www.science.nrao.edu/facilities/gbt
Effelsberg 100-m Telescope	Bonn, Germany	100-m steerable dish	www.mpifr-bonn.mpg.de/en/effelsberg
Lovell Telescope	Manchester, England	76-m steerable dish	www.jb.man.ac.uk/aboutus/lovell
Canberra Deep Space Communication Complex (CDSCC)	Tidbinbilla, Australia	70-m steerable dish	www.cdscc.nasa.gov
Goldstone Deep Space Communications Complex (GDSCC)	Barstow, CA	70-m steerable dish	www.gdscc.nasa.gov
Parkes Observatory	Parkes, Australia	64-m steerable dish	www.parkes.atnf.csiro.au
Arrays of Radio Dishes			
Square Kilometre Array (SKA)	South Africa and Western Australia	Thousands of dishes, km^2 collecting area, partial array in 2020	www.skatelescope.org
Atacama Large Millimeter/ submillimeter Array (ALMA)	Atacama desert, Northern Chile	66 7-m and 12-m dishes	www.almaobservatory.org
Jansky Very Large Array (VLA)	Socorro, New Mexico	27-element array of 25-m dishes (36-km baseline)	www.science.nrao.edu/facilities/vla
Westerbork Synthesis Radio Telescope (WSRT)	Westerbork, the Netherlands	12-element array of 25-m dishes (1.6-km baseline)	www.astron.nl/radio-observatory/public/public-0

Observatory	Location	Description	Website
Very Long Baseline Array (VLBA)	Ten US sites, HI to the Virgin Islands	10-element array of 25-m dishes (9000 km baseline)	www.science.nrao.edu/facilities/vlba
Australia Telescope Compact Array (ATCA)	Several sites in Australia	8-element array (seven 22-m dishes plus Parkes 64 m)	www.narrabri.atnf.csiro.au
Multi-Element Radio Linked Interferometer Network (MERLIN)	Cambridge, England, and other British sites	Network of seven dishes (the largest is 32 m)	www.e-merlin.ac.uk
Millimeter-wave Telescopes			
IRAM	Granada, Spain	30-m steerable mm-wave dish	www.iram-institute.org
James Clerk Maxwell Telescope (JCMT)	Maunakea, HI	15-m steerable mm-wave dish	www.eaobservatory.org/jcmt
Nobeyama Radio Observatory (NRO)	Minamimaki, Japan	6-element array of 10-m wave dishes	www.nro.nao.ac.jp/en
Hat Creek Radio Observatory (HCRO)	Cassel, CA	6-element array of 5-m wave dishes	www.sri.com/research-development/specialized-facilities/hat-creek-radio-observatory

Radio Interferometry

As we discussed earlier, a telescope's ability to show us fine detail (its resolution) depends upon its aperture, but it also depends upon the wavelength of the radiation that the telescope is gathering. The longer the waves, the harder it is to resolve fine detail in the images or maps we make. Because radio waves have such long wavelengths, they present tremendous challenges for astronomers who need good resolution. In fact, even the largest radio dishes on Earth, operating alone, cannot make out as much detail as the typical small visible-light telescope used in a college astronomy lab. To overcome this difficulty, radio astronomers have learned to sharpen their images by linking two or more radio telescopes together electronically. Two or more telescopes linked together in this way are called an interferometer.

"Interferometer" may seem like a strange term because the telescopes in an interferometer work cooperatively; they don't "interfere" with each other. Interference, however, is a technical term for the way that multiple waves interact with each other when they arrive in our instruments, and this interaction allows us to coax more detail out of our observations. The resolution of an interferometer depends upon the separation of the telescopes, not upon their individual apertures. Two telescopes separated by 1 kilometer provide the same resolution as would a single dish 1 kilometer across (although they are not, of course, able to collect as much radiation as a radio-wave bucket that is 1 kilometer across).

To get even better resolution, astronomers combine a large number of radio dishes into an interferometer array. In effect, such an array works like a large number of two-dish interferometers, all observing the same part of the sky together. Computer processing of the results permits the reconstruction of a high-resolution radio image. The most extensive such instrument in the United States is the National Radio Astronomy Observatory's Jansky Very Large Array (VLA) near Socorro, New Mexico. It consists of 27 movable radio telescopes (on railroad tracks), each having an aperture of 25 meters, spread over a total span of about 36 kilometers. By electronically combining the signals from all of its individual telescopes, this array permits the radio astronomer to make pictures of the sky at radio wavelengths comparable to those obtained with a visible-light telescope, with a resolution of about 1 arcsecond.

The Atacama Large Millimeter/submillimeter array (ALMA) in the Atacama Desert of Northern Chile (Figure 6.20), at an altitude of 16,400 feet, consists of 12 7-meter and 54 12-meter telescopes, and can achieve baselines up to 16 kilometers. Since it became operational in 2013, it has made observations at resolutions down to 6 milliarcseconds (0.006 arcseconds), a remarkable achievement for radio astronomy.



Figure 6.20 : Atacama Large Millimeter/Submillimeter Array (ALMA). Located in the Atacama Desert of Northern Chile, ALMA currently provides the highest resolution for radio observations. (credit: ESO/S. Guisard)

Link to Learning

Watch this documentary that explains the work that went into designing and building ALMA, discusses some of its first images, and explores its future.

Initially, the size of interferometer arrays was limited by the requirement that all of the dishes be physically wired together. The maximum dimensions of the array were thus only a few tens of kilometers. However, larger interferometer separations can be achieved if the telescopes do not require a physical connection. Astronomers, with the use of current technology and computing power, have learned to time the arrival of electromagnetic waves coming from space very precisely at each telescope and combine the data later. If the telescopes are as far apart as California and Australia, or as West Virginia and Crimea in Ukraine, the resulting resolution far surpasses that of visible-light telescopes.

The United States operates the Very Long Baseline Array (VLBA), made up of 10 individual telescopes stretching from the Virgin Islands to Hawaii (Figure 6.21). The VLBA, completed in 1993, can form astronomical images with a resolution of 0.0001 arcseconds, permitting features as small as 10 astronomical units (AU) to be distinguished at the center of our Galaxy.



Figure 6.21 : Very Long Baseline Array. This map shows the distribution of 10 antennas that constitute an array of radio telescopes stretching across the United States and its territories.

Recent advances in technology have also made it possible to do interferometry at visible-light and infrared wavelengths. At the beginning of the twenty-first century, three observatories with multiple telescopes each began using their dishes as interferometers, combining their light to obtain a much greater resolution. In addition, a dedicated interferometric array was built

on Mt. Wilson in California. Just as in radio arrays, these observations allow astronomers to make out more detail than a single telescope could provide.

Table 6.3: Visible-Light Interferometers

Longest Baseline (m)	Telescope Name	Location	Mirrors	Status
400	CHARA Array (Center for High Angular Resolution Astronomy)	Mount Wilson, CA	Six 1-m telescopes	Operational since 2004
200	Very Large Telescope	Cerro Paranal, Chile	Four 8.2-m telescopes	Completed 2000
85	Keck I and II telescopes	Maunakea, HI	Two 10-m telescopes	Operated from 2001 to 2012
22.8	Large Binocular Telescope	Mount Graham, AZ	Two 8.4-m telescopes	First light 2004

Radar Astronomy

Radar is the technique of transmitting radio waves to an object in our solar system and then detecting the radio radiation that the object reflects back. The time required for the round trip can be measured electronically with great precision. Because we know the speed at which radio waves travel (the speed of light), we can determine the distance to the object or a particular feature on its surface (such as a mountain).

Radar observations have been used to determine the distances to planets and how fast things are moving in the solar system (using the Doppler effect, discussed in the Radiation and Spectra chapter). Radar waves have played important roles in navigating spacecraft throughout the solar system. In addition, as will be discussed in later chapters, radar observations have determined the rotation periods of Venus and Mercury, probed tiny Earth-approaching asteroids, and allowed us to investigate the mountains and valleys on the surfaces of Mercury, Venus, Mars, and the large moons of Jupiter.

Any radio dish can be used as a radar telescope if it is equipped with a powerful transmitter as well as a receiver. For many years, the most spectacular facility in the world for radar astronomy was the 1000-foot (305-meter) telescope at Arecibo in Puerto Rico (Figure 6.22). The Arecibo telescope was too large to be pointed directly at different parts of the sky. Instead, it was constructed in a huge natural “bowl” (more than a mere dish) formed by several hills, and it was lined with reflecting metal panels. A limited ability to track astronomical sources was achieved by moving the receiver system, which was suspended on cables 100 meters above the surface of the bowl. Unfortunately, the telescope was seriously damaged in the powerful storms of 2020 and had to be decommissioned. An even larger (500-meter) radar telescope has recently gone into operation in China and is called the Five-hundred-meter Aperture Spherical Telescope (FAST).



Figure 6.22 : Largest Radio and Radar Dish. The Arecibo Observatory in Puerto Rico was the largest and most powerful astronomical radar faculty in the world and was often featured in films. In November 2020 it collapsed, damaging the 300-meter diameter “dish” and destroying the radar transmitter and receiver. (credit: National Astronomy and Ionosphere Center, Cornell U., NSF)

This page titled 5.10: Radio Telescopes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 6.5: Radio Telescopes by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.11: Observations outside Earth's Atmosphere

Learning Objectives

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

- List the advantages of making astronomical observations from space
- Explain the importance of the Hubble Space Telescope
- Describe some of the major space-based observatories astronomers use

Earth's atmosphere blocks most radiation at wavelengths shorter than visible light, so we can only make direct ultraviolet, X-ray, and gamma ray observations from space (though indirect gamma ray observations can be made from Earth). Getting above the distorting effects of the atmosphere is also an advantage at visible and infrared wavelengths. The stars don't "twinkle" in space, so the amount of detail you can observe is limited only by the size of your instrument. On the other hand, it is expensive to place telescopes into space, and repairs can present a major challenge. This is why astronomers continue to build telescopes for use on the ground as well as for launching into space.

Airborne and Space Infrared Telescopes

Water vapor, the main source of atmospheric interference for making infrared observations, is concentrated in the lower part of Earth's atmosphere. For this reason, a gain of even a few hundred meters in elevation can make an important difference in the quality of an infrared observatory site. Given the limitations of high mountains, most of which attract clouds and violent storms, and the fact that the ability of humans to perform complex tasks degrades at high altitudes, it was natural for astronomers to investigate the possibility of observing infrared waves from airplanes and ultimately from space.

Infrared observations from airplanes have been made since the 1960s, starting with a 15-centimeter telescope on board a Learjet. From 1974 through 1995, NASA operated a 0.9-meter airborne telescope flying regularly out of the Ames Research Center south of San Francisco. Observing from an altitude of 12 kilometers, the telescope was above 99% of the atmospheric water vapor. More recently, NASA (in partnership with the German Aerospace Center) has constructed a much larger 2.5-meter telescope, called the Stratospheric Observatory for Infrared Astronomy (SOFIA), which flies in a modified Boeing 747SP (Figure 6.23).



Figure 6.23 : Stratospheric Observatory for Infrared Astronomy (SOFIA). SOFIA allows observations to be made above most of Earth's atmospheric water vapor. (credit: NASA)

Link to Learning

To find out more about SOFIA, watch this video provided by NASA's Armstrong Flight Research Center.

Getting even higher and making observations from space itself have important advantages for infrared astronomy. First is the elimination of all interference from the atmosphere. Equally important is the opportunity to cool the entire optical system of the instrument in order to nearly eliminate infrared radiation from the telescope itself. If we tried to cool a telescope within the atmosphere, it would quickly become coated with condensing water vapor and other gases, making it useless. Only in the vacuum of space can optical elements be cooled to hundreds of degrees below freezing and still remain operational.

The first orbiting infrared observatory, launched in 1983, was the Infrared Astronomical Satellite (IRAS), built as a joint project by the United States, the Netherlands, and Britain. IRAS was equipped with a 0.6-meter telescope cooled to a temperature of less than 10 K. For the first time, the infrared sky could be seen as if it were night, rather than through a bright foreground of atmospheric and telescope emissions. IRAS carried out a rapid but comprehensive survey of the entire infrared sky over a 10-month period, cataloging about 350,000 sources of infrared radiation. Since then, several other infrared telescopes have operated in space with much better sensitivity and resolution due to improvements in infrared detectors. The most powerful of these infrared telescopes is the 0.85-meter Spitzer Space Telescope, which launched in 2003. A few of its observations are shown in Figure 6.24. With infrared observations, astronomers can detect cooler parts of cosmic objects, such as the dust clouds around star nurseries and the remnants of dying stars, that visible-light images don't reveal.

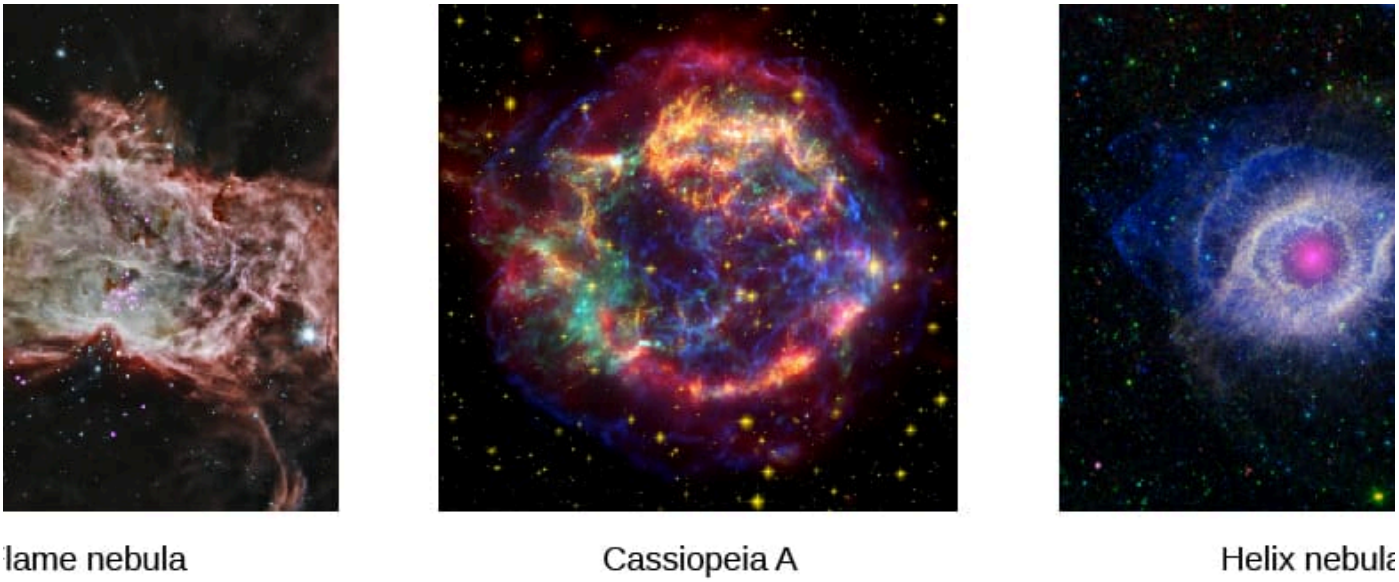


Figure 6.24 : Observations from the Spitzer Space Telescope (SST). These infrared images—a region of star formation, the remnant of an exploded star, and a region where an old star is losing its outer shell—show just a few of the observations made and transmitted back to Earth from the SST. Since our eyes are not sensitive to infrared rays, we don't perceive colors from them. The colors in these images have been selected by astronomers to highlight details like the composition or temperature in these regions. (credit "Flame nebula": modification of work by NASA (X-ray: NASA/CXC/PSU/K.Getman, E.Feigelson, M.Kuhn & the MYStIX team; Infrared:NASA/JPL-Caltech); credit "Cassiopeia A": modification of work by NASA/JPL-Caltech; credit "Helix nebula": modification of work by NASA/JPL-Caltech)

Hubble Space Telescope

In April 1990, a great leap forward in astronomy was made with the launch of the Hubble Space Telescope (HST). With an aperture of 2.4 meters, this is the largest telescope put into space so far. (Its aperture was limited by the size of the payload bay in the Space Shuttle that served as its launch vehicle.) It was named for Edwin Hubble, the astronomer who discovered the expansion of the universe in the 1920s (whose work we will discuss in the chapters on Galaxies).

HST is operated jointly by NASA's Goddard Space Flight Center and the Space Telescope Science Institute in Baltimore. It was the first orbiting observatory designed to be serviced by Shuttle astronauts and, over the years since it was launched, they made several visits to improve or replace its initial instruments and to repair some of the systems that operate the spacecraft (Figure 6.1)—though this repair program has now been discontinued, and no more visits or improvements will be made.

With the Hubble, astronomers have obtained some of the most detailed images of astronomical objects from the solar system outward to the most distant galaxies. Among its many great achievements is the Hubble Ultra-Deep Field, an image of a small region of the sky observed for almost 100 hours. It contains views of about 10,000 galaxies, some of which formed when the universe was just a few percent of its current age (Figure 6.25).

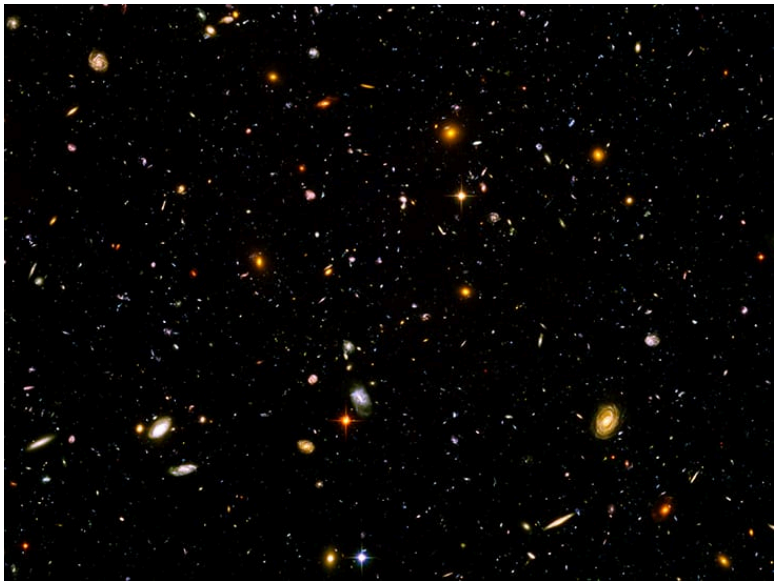


Figure 6.25 : Hubble Ultra-Deep Field (HUDF). The Hubble Space Telescope has provided an image of a specific region of space built from data collected between September 24, 2003, and January 16, 2004. These data allow us to search for galaxies that existed approximately 13 billion years ago. (credit: modification of work by NASA)

The HST's mirror was ground and polished to a remarkable degree of accuracy. If we were to scale up its 2.4-meter mirror to the size of the entire continental United States, there would be no hill or valley larger than about 6 centimeters in its smooth surface. Unfortunately, after it was launched, scientists discovered that the primary mirror had a slight error in its *shape,* equal to roughly 1/50 the width of a human hair. Small as that sounds, it was enough to ensure that much of the light entering the telescope did not come to a clear focus and that all the images were blurry. (In a misplaced effort to save money, a complete test of the optical system had not been carried out before launch, so the error was not discovered until HST was in orbit.)

The solution was to do something very similar to what we do for astronomy students with blurry vision: put corrective optics in front of their eyes. In December 1993, in one of the most exciting and difficult space missions ever flown, astronauts captured the orbiting telescope and brought it back into the shuttle payload bay. There they installed a package containing compensating optics as well as a new, improved camera before releasing HST back into orbit. The telescope now works as it was intended to, and further missions to it were able to install even more advanced instruments to take advantage of its capabilities.

High-Energy Observatories

Ultraviolet, X-ray, and direct gamma-ray (high-energy electromagnetic wave) observations can be made only from space. Such observations first became possible in 1946, with V2 rockets captured from Germany after World War II. The US Naval Research Laboratory put instruments on these rockets for a series of pioneering flights, used initially to detect ultraviolet radiation from the Sun. Since then, many other rockets have been launched to make X-ray and ultraviolet observations of the Sun, and later of other celestial objects.

Beginning in the 1960s, a steady stream of high-energy observatories has been launched into orbit to reveal and explore the universe at short wavelengths. Among recent X-ray telescopes is the Chandra X-ray Observatory, which was launched in 1999 (Figure 6.26). It is producing X-ray images with unprecedented resolution and sensitivity. Designing instruments that can collect and focus energetic radiation like X-rays and gamma rays is an enormous technological challenge. The 2002 Nobel Prize in physics was awarded to Riccardo Giacconi, a pioneer in the field of building and launching sophisticated X-ray instruments. In 2008, NASA launched the Fermi Gamma-ray Space Telescope, designed to measure cosmic gamma rays at energies greater than any previous telescope, and thus able to collect radiation from some of the most energetic events in the universe.



Figure 6.26 : Chandra X-Ray Satellite. Chandra, the world's most powerful X-ray telescope, was developed by NASA and launched in July 1999. (credit: modification of work by NASA)

One major challenge is to design "mirrors" to reflect such penetrating radiation as X-rays and gamma rays, which normally pass straight through matter. However, although the technical details of design are more complicated, the three basic components of an observing system, as we explained earlier in this chapter, are the same at all wavelengths: a telescope to gather up the radiation, filters or instruments to sort the radiation according to wavelength, and some method of detecting and making a permanent record of the observations. Table 6.4 lists some of the most important active space observatories that humanity has launched.

Gamma-ray detections can also be made from Earth's surface by using the atmosphere as the primary detector. When a gamma ray hits our atmosphere, it accelerates charged particles (mostly electrons) in the atmosphere. Those energetic particles hit other particles in the atmosphere and give off their own radiation. The effect is a cascade of light and energy that can be detected on the ground. The VERITAS array in Arizona and the H.E.S.S. array in Namibia are two such ground-based gamma-ray observatories.

Table 6.4: Recent Observatories in Space

Observatory	Date Operation Began	Bands of the Spectrum	Notes	Website
Hubble Space Telescope (HST)	1990	visible, UV, IR	2.4-m mirror; images and spectra	www.hubblesite.org
Chandra X-Ray Observatory	1999	X-rays	X-ray images and spectra	www.chandra.si.edu
XMM-Newton	1999	X-rays	X-ray spectroscopy	http://www.cosmos.esa.int/web/xmm-newton
International Gamma-Ray Astrophysics Laboratory (INTEGRAL)	2002	X- and gamma-rays	higher resolution gamma-ray images	http://sci.esa.int/integral/
Spitzer Space Telescope	2003	IR	0.85-m telescope	www.spitzer.caltech.edu
Fermi Gamma-ray Space Telescope	2008	gamma-rays	first high-energy gamma-ray observations	fermi.gsfc.nasa.gov
Kepler	2009	visible-light	planet finder	kepler.nasa.gov
Wide-field Infrared Survey Explorer (WISE)	2009	IR	whole-sky map, asteroid searches	www.nasa.gov/mission_pages/WISE/main
Gaia	2013	visible-light	Precise map of the Milky Way	http://sci.esa.int/gaia/

Observatory	Date Operation Began	Bands of the Spectrum	Notes	Website
Transiting Exoplanet Survey Satellite (TESS)	2018	visible-light	Planet finder	http://tess.mit.edu

This page titled 5.11: Observations outside Earth's Atmosphere is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 6.6: Observations outside Earth's Atmosphere by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

5.12: The Future of Large Telescopes

Learning Objectives

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

- Describe the next generation of ground- and space-based observatories
- Explain some of the challenges involved in building these observatories

If you've ever gone on a hike, you have probably been eager to see what lies just around the next bend in the path. Researchers are no different, and astronomers and engineers are working on the technologies that will allow us to explore even more distant parts of the universe and to see them more clearly.

The premier space facility planned for the next decade is the James Webb Space Telescope (Figure 6.27), which was launched on December 25, 2021. James Webb was one of the early administrators in NASA. As we write this (in December 2021), the telescope is on its way to a stable orbit point, some 1.5 million kilometers from Earth (where no astronauts can currently travel if the facility needs repair). During its 29-day journey, a number of activities must be successfully executed in order to prepare the JWST to make observations. A solar array will be deployed to provide energy, and a sunshield will be stretched out to protect the telescope from the intense heat of the Sun. The secondary and primary mirrors will then be unfolded. The primary mirror is 6 meters in diameter, made up, like the Keck telescopes, of 36 small hexagons. JWST has been designed to operate at infrared wavelengths and will have the sensitivity needed to detect the very first generation of stars, formed when the universe was only a few hundred million years old.



Figure 6.27 : James Webb Space Telescope (JWST). This image shows some of the mirrors of the JWST as they underwent cryogenic testing. The mirrors were exposed to extreme temperatures in order to gather accurate measurements on changes in their shape as they heated and cooled. (credit: NASA/MSFC/David Higginbotham/Emmett Given)

Link to Learning

Watch this video to learn more about the James Webb Space Telescope and how it will build upon the work that Hubble has allowed us to begin in exploring the universe. Visit NASA's James Webb Space Telescope site to keep track of how the JWST is doing.

On the ground, astronomers have started building the Vera Rubin Observatory[1], which has an 8.4-meter telescope with a

significantly larger field of view than any existing telescopes. It will rapidly scan the sky to find *transients*, phenomena that change quickly, such as exploding stars and chunks of rock that orbit near Earth. It is expected to see first light in 2022.

The international gamma-ray community is planning the Cherenkov Telescope Array (CTA), two arrays of telescopes, one in each hemisphere, which will indirectly measure gamma rays from the ground. The CTA will measure gamma-ray energies a thousand times as great as the Fermi telescope can detect.

Several groups of astronomers around the globe interested in studying visible light and the infrared are exploring the feasibility of building ground-based telescopes with mirrors larger than 30 meters across. Stop and think what this means: 30 meters is one-third the length of a football field. It is technically impossible to build and transport a single astronomical mirror that is 30 meters or larger in diameter. The primary mirror of these giant telescopes will consist of smaller mirrors, all aligned so that they act as a very large mirror in combination.

The most ambitious of these projects is the European Extremely Large Telescope (ELT) (Figure 6.28). (Astronomers try to outdo each other not only with the size of these telescopes, but also their names!) The design of the European ELT calls for a 39.3-meter primary mirror, which will follow the Keck design and be made up of 798 hexagonal mirrors, each 1.4 meters in diameter and all held precisely in position so that they form a continuous surface. Construction on the site in the Atacama Desert in Northern Chile started in 2014, and operations are expected to begin in about 2025.

International consortia with major contributions from U.S. astronomers have developed plans for the construction of two large new telescopes. One is a Thirty-Meter Telescope (TMT) for which the preferred site is Maunakea in Hawaii. The design of this telescope is similar to that of the European ELT and will make use of 492 hexagonal elements. Each segment is about 1.44 meters (56.6 inches) across corners. The segments are closely spaced, with gaps between the segments only 2.5 mm (0.1 inch) wide.

The Giant Magellan Telescope (GMT) is the second ELT project with major participation by U.S. astronomers. The GMT is also a segmented mirror telescope that employs seven stiff monolith 8.4-meter mirrors as segments. Construction has started at the selected site, which is near the Las Campanas Observatory on the southern edge of the Atacama Desert.

These giant telescopes will combine light-gathering power with high-resolution imaging. These powerful new instruments will enable astronomers to tackle many important astronomical problems. As just one example, they provide us images and spectra of planets around other stars and thus, perhaps, give us the first real evidence (from the chemistry of these planets' atmospheres) that life exists elsewhere.

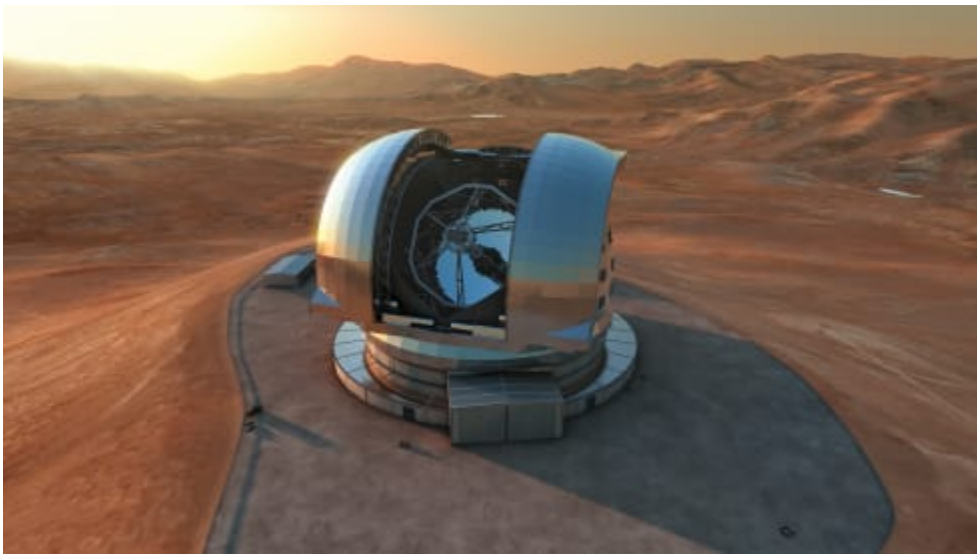


Figure 6.28 : Artist's Conception of the European Extremely Large Telescope. The primary mirror in this telescope is 39.3 meters across. The telescope is under construction in the Atacama Desert in Northern Chile. (credit: ESO/L. Calçada)

Link to Learning

Check out this fun diagram comparing the sizes of the largest planned and existing telescopes to a regulation basketball and tennis court.

Footnotes

- 1The observatory is named after the American astronomer whose work led us to the understanding that much of the universe is made of a mysterious substance that scientists call dark matter (which we explain in The Mass of the Galaxy).

This page titled 5.12: The Future of Large Telescopes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 6.7: The Future of Large Telescopes by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

6: INTRODUCTION TO THE SOLAR SYSTEM

6.1: Overview of Our Planetary System
6.2: Composition and Structure of Planets
6.3: Dating Planetary Surfaces
6.4: Origin of the Solar System
6.5: Formation of the Solar System

This page titled 6: Introduction to the Solar System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.1: Overview of Our Planetary System

Learning Objectives

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

- Describe how the objects in our solar system are identified, explored, and characterized
- Describe the types of small bodies in our solar system, their locations, and how they formed
- Model the solar system with distances from everyday life to better comprehend distances in space

The solar system[1] consists of the Sun and many smaller objects: the planets, their moons and rings, and such "debris" as asteroids, comets, and dust. Decades of observation and spacecraft exploration have revealed that most of these objects formed together with the Sun about 4.5 billion years ago. They represent clumps of material that condensed from an enormous cloud of gas and dust. The central part of this cloud became the Sun, and a small fraction of the material in the outer parts eventually formed the other objects.

During the past 50 years, we have learned more about the solar system than anyone imagined before the space age. In addition to gathering information with powerful new telescopes, we have sent spacecraft directly to many members of the planetary system. (Planetary astronomy is the only branch of our science in which we can, at least vicariously, travel to the objects we want to study.) With evocative names such as *Voyager*, *Pioneer*, *Curiosity*, and *Pathfinder*, our robot explorers have flown past, orbited, or landed on every planet, returning images and data that have dazzled both astronomers and the public. In the process, we have also investigated two dwarf planets, hundreds of fascinating moons, four ring systems, a dozen asteroids, and several comets (smaller members of our solar system that we will discuss later).

Our probes have penetrated the atmosphere of Jupiter and landed on the surfaces of Venus, Mars, our Moon, Saturn's moon Titan, the asteroids Eros, Itokawa, Ryugu, and Bennu, and the Comet Churyumov-Gerasimenko (usually referred to as 67P). Humans have set foot on the Moon and returned samples of its surface soil for laboratory analysis (Figure 7.2). We have flown a helicopter drone on Mars. We have even discovered other places in our solar system that might be able to support some kind of life.



Figure 7.2 : Astronauts on the Moon. The lunar lander and surface rover from the Apollo 15 mission are seen in this view of the one place beyond Earth that has been explored directly by humans. (credit: modification of work by David R. Scott, NASA)

Link to Learning

View this gallery of NASA images that trace the history of the Apollo mission.

An Inventory

The Sun, a star that is brighter than about 80% of the stars in the Galaxy, is by far the most massive member of the solar system, as shown in Table 7.1. It is an enormous ball about 1.4 million kilometers in diameter, with surface layers of incandescent gas and an interior temperature of millions of degrees. The Sun will be discussed in later chapters as our first, and best-studied, example of a star.

Table 7.1: Mass of Members of the Solar System

Object	Percentage of Total Mass of Solar System
Sun	99.80
Jupiter	0.10
Comets	0.0005–0.03 (estimate)
All other planets and dwarf planets	0.04
Moons and rings	0.00005
Asteroids	0.0000002 (estimate)
Cosmic dust	0.00000001 (estimate)

Table 7.1 also shows that most of the material of the planets is actually concentrated in the largest one, Jupiter, which is more massive than all the rest of the planets combined. Astronomers were able to determine the masses of the planets centuries ago using Kepler's laws of planetary motion and Newton's law of gravity to measure the planets' gravitational effects on one another or on moons that orbit them (see Orbits and Gravity). Today, we make even more precise measurements of their masses by tracking their gravitational effects on the motion of spacecraft that pass near them.

Beside Earth, five other planets were known to the ancients—Mercury, Venus, Mars, Jupiter, and Saturn—and two were discovered after the invention of the telescope: Uranus and Neptune. The eight planets all revolve in the same direction around the Sun. They orbit in approximately the same plane, like cars traveling on concentric tracks on a giant, flat racecourse. Each planet stays in its own "traffic lane," following a nearly circular orbit about the Sun and obeying the "traffic" laws discovered by Galileo, Kepler, and Newton. Besides these planets, we have also been discovering smaller worlds beyond Neptune that are called trans-Neptunian objects or TNOs (see Figure 7.3). The first to be found, in 1930, was Pluto, but others have been discovered during the twenty-first century. One of them, Eris, is about the same size as Pluto and has at least one moon (Pluto has five known moons.) The largest TNOs are also classed as *dwarf planets,* as is the largest asteroid, Ceres. (Dwarf planets will be discussed further in the chapter on Rings, Moons, and Pluto). To date, more than 2600 of these TNOs have been discovered, and one, called Arrokoth, was explored by the New Horizons spacecraft.

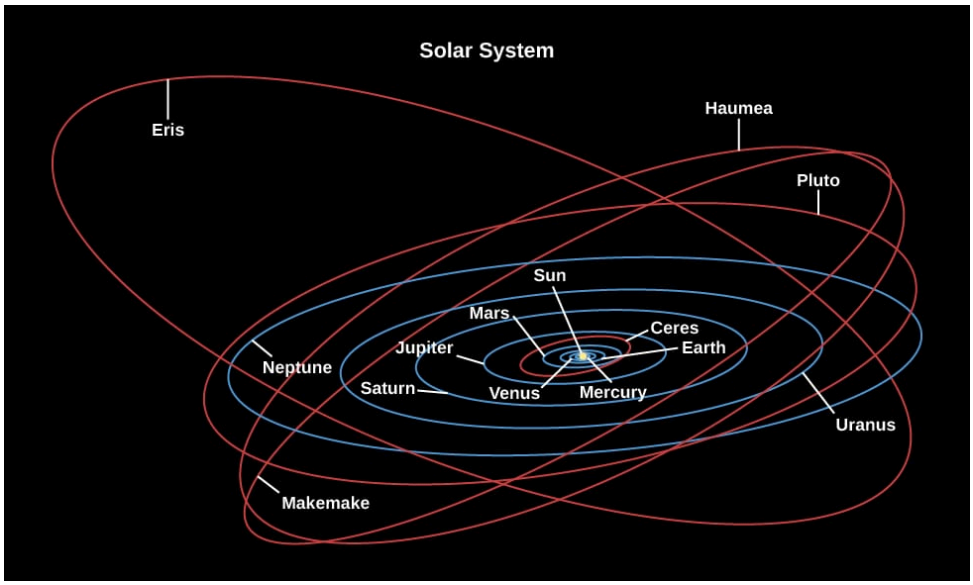


Figure 7.3 : Orbits of the Planets. All eight major planets orbit the Sun in roughly the same plane. The five currently known dwarf planets are also shown: Eris, Haumea, Pluto, Ceres, and Makemake. Note that Pluto's orbit is not in the plane of the planets.

Each of the planets and dwarf planets also rotates (spins) about an axis running through it, and in most cases the direction of rotation is the same as the direction of revolution about the Sun. The exceptions are Venus, which rotates backward very slowly (that is, in a retrograde direction), and Uranus and Pluto, which also have strange rotations, each spinning about an axis tipped nearly on its side. We do not yet know the spin orientations of Eris, Haumea, and Makemake.

The four planets closest to the Sun (Mercury through Mars) are called the inner or terrestrial planets. Often, the Moon is also discussed as a part of this group, bringing the total of terrestrial objects to five. (We generally call Earth's satellite "the Moon," with a capital M, and the other satellites "moons," with lowercase m's.) The terrestrial planets are relatively small worlds, composed primarily of rock and metal. All of them have solid surfaces that bear the records of their geological history in the forms of craters, mountains, and volcanoes (Figure 7.4).

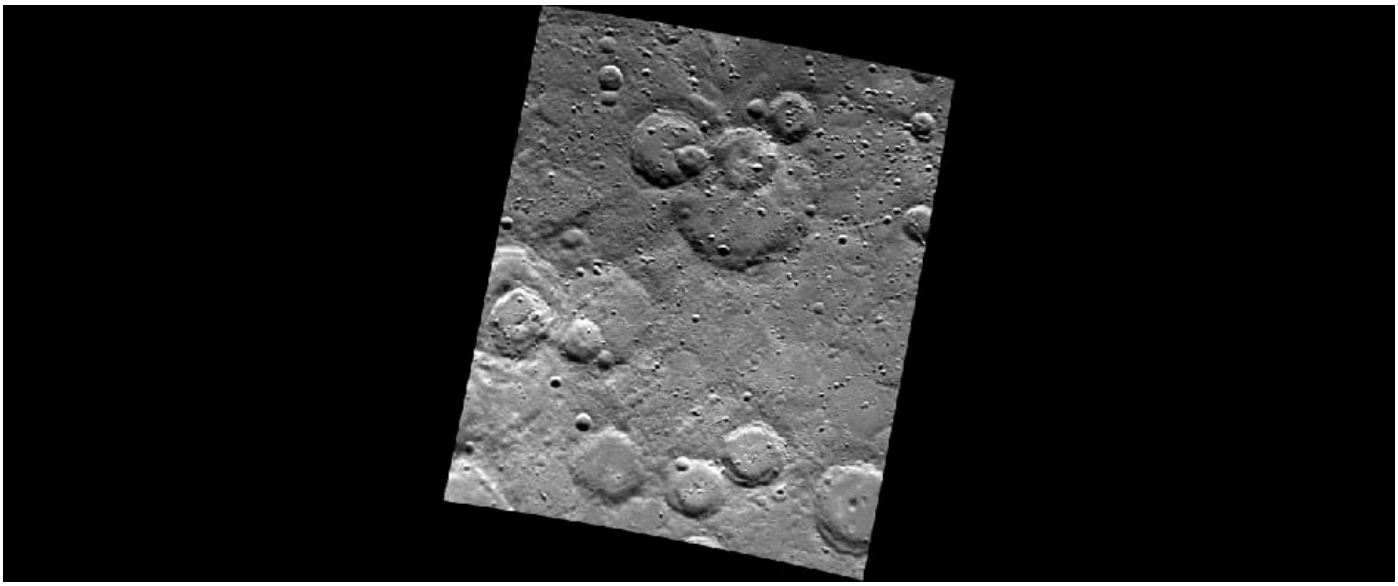


Figure 7.4 : Surface of Mercury. The pockmarked face of the terrestrial world of Mercury is more typical of the inner planets than the watery surface of Earth. This black-and-white image, taken with the Mariner 10 spacecraft, shows a region more than 400 kilometers wide. (credit: modification of work by NASA/John Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington)

The next four planets (Jupiter through Neptune) are much larger and are composed primarily of lighter ices, liquids, and gases. We call these four the jovian planets (after "Jove," another name for Jupiter in mythology) or giant planets—a name they richly deserve (Figure 7.5). About 1,300 Earths could fit inside Jupiter, for example. These planets do not have solid surfaces on which

future explorers might land. They are more like vast, spherical oceans with much smaller, dense cores.

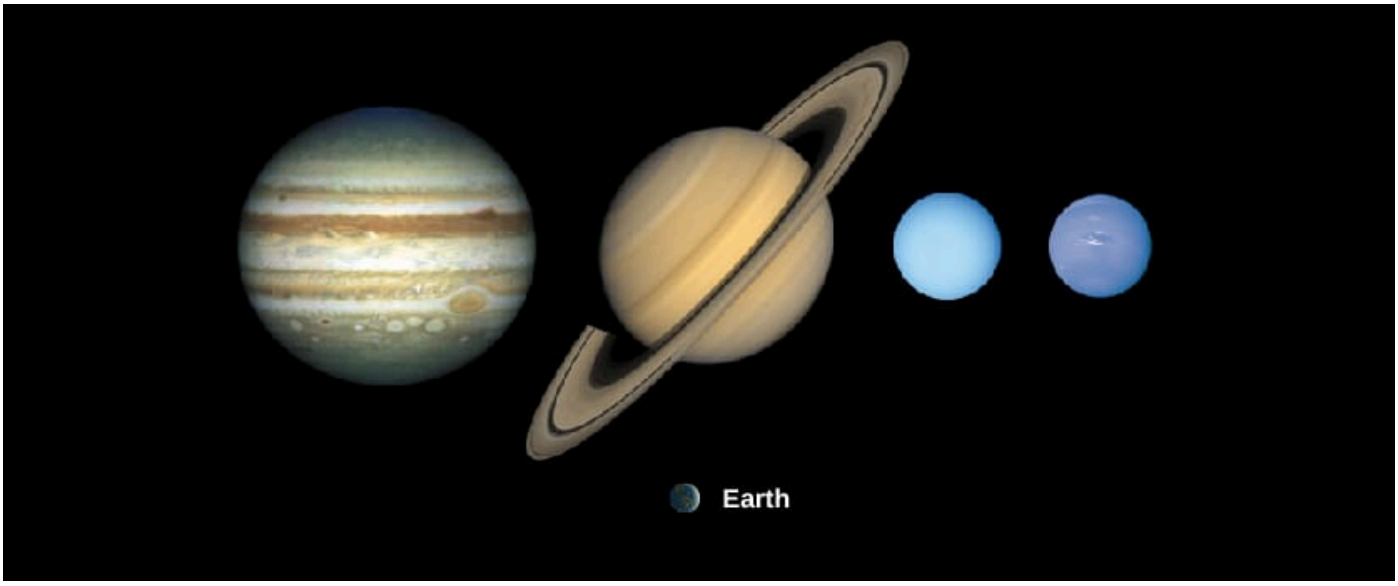


Figure 7.5 : The Four Giant Planets. This montage shows the four giant planets: Jupiter, Saturn, Uranus, and Neptune. Below them, Earth is shown to scale. (credit: modification of work by NASA, Solar System Exploration)

Near the outer edge of the system lies Pluto, which was the first of the distant icy worlds to be discovered beyond Neptune (Pluto was visited by a spacecraft, the NASA New Horizons mission, in 2015 [see Figure 7.6]). Table 7.2 summarizes some of the main facts about the planets.

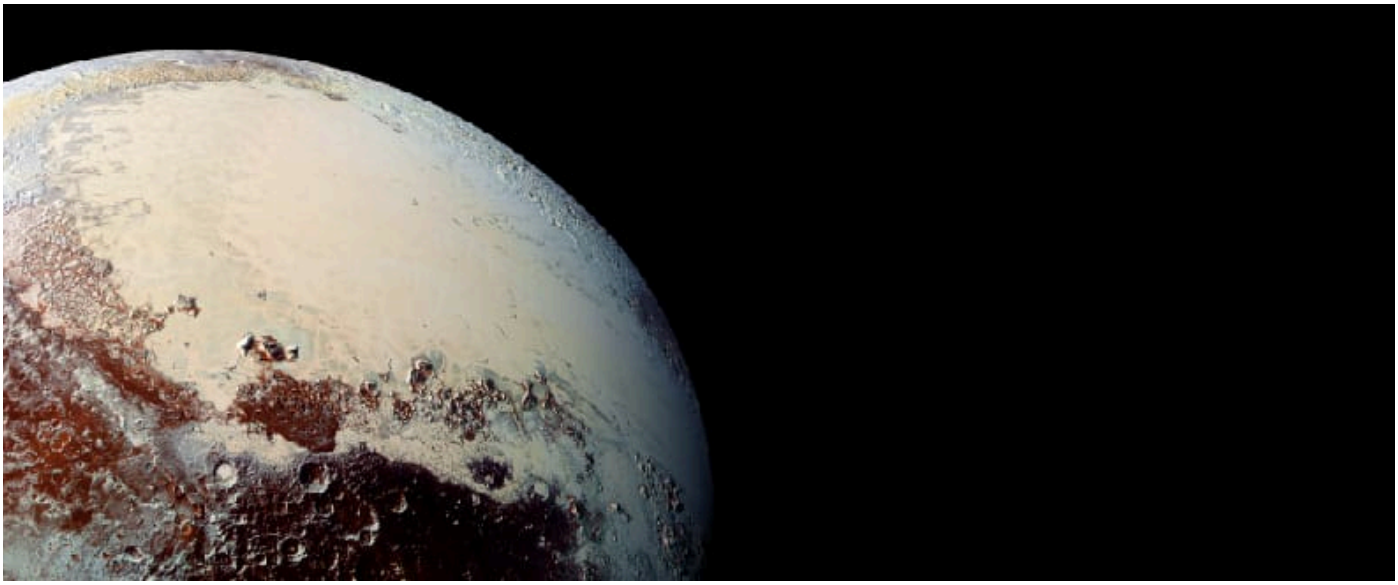


Figure 7.6 : Pluto Close-up. This intriguing image from the New Horizons spacecraft, taken when it flew by the dwarf planet in July 2015, shows some of its complex surface features. The rounded white area is called the Sputnik Plain, after humanity's first spacecraft. (credit: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

Table 7.2: The Planets

Name	Distance from Sun (AU)[2]	Revolution Period (y)	Diameter (km)	Mass (10^{23} kg)	Density (g/cm^3)[3]
Mercury	0.39	0.24	4,878	3.3	5.4
Venus	0.72	0.62	12,120	48.7	5.2

Name	Distance from Sun (AU)[2]	Revolution Period (y)	Diameter (km)	Mass (10^{23} kg)	Density (g/cm^3)[3]
Earth	1.00	1.00	12,756	59.8	5.5
Mars	1.52	1.88	6,787	6.4	3.9
Jupiter	5.20	11.86	142,984	18,991	1.3
Saturn	9.54	29.46	120,536	5686	0.7
Uranus	19.18	84.07	51,118	866	1.3
Neptune	30.06	164.82	49,660	1030	1.6

Example 7.1: Comparing Densities

Let's compare the densities of several members of the solar system. The density of an object equals its mass divided by its volume. The volume (V) of a sphere (like a planet) is calculated using the equation

$$V = \frac{4}{3}\pi R^3$$

where π (the Greek letter pi) has a value of approximately 3.14. Although planets are not perfect spheres, this equation works well enough. The masses and diameters of the planets are given in Table 7.2. For data on selected moons, see Appendix G. Let's use Saturn's moon Mimas as our example, with a mass of 4×10^{19} kg and a diameter of approximately 400 km (radius, $200\text{km} = 2 \times 10^5 m$).

Solution

The volume of Mimas is

$$\frac{4}{3} \times 3.14 \times \left(2 \times 10^5\ \text{m}\right)^3 = 3.3 \times 10^{16}\ \text{m}^3.$$

Density is mass divided by volume:

$$\frac{4 \times 10^{19}\ \text{kg}}{3.3 \times 10^{16}\ \text{m}^3} = 1.2 \times 10^3\ \text{kg/m}^3.$$

Note that the density of water in these units is 1000 kg/m^3, so Mimas must be made mainly of ice, not rock. (Note that the density of Mimas given in Appendix G is 1.2, but the units used there are different. In that table, we give density in units of g/cm^3, for which the density of water equals 1. Can you show, by converting units, that 1 g/cm^3 is the same as 1000 kg/m^3?)

Exercise 6.1.1

Calculate the average density of our own planet, Earth. Show your work. How does it compare to the density of an ice moon like Mimas? See Table 7.2 for data.

Answer available at the end of this section.

Link to Learning

Learn more about NASA's mission to Pluto and see high-resolution images of Pluto and its moon Charon.

Smaller Members of the Solar System

Most of the planets are accompanied by one or more moons; only Mercury and Venus move through space alone. There are more than 210 known moons orbiting planets and dwarf planets (see Appendix G for a listing of the larger ones), and undoubtedly many other small ones remain undiscovered. The largest of the moons are as big as small planets and just as interesting. In addition to our Moon, they include the four largest moons of Jupiter (called the Galilean moons, after their discoverer) and the largest moons of Saturn and Neptune (confusingly named Titan and Triton).

Each of the giant planets also has rings made up of countless small bodies ranging in size from mountains to mere grains of dust, all in orbit about the equator of the planet. The bright rings of Saturn are, by far, the easiest to see. They are among the most beautiful sights in the solar system (Figure 7.7). But, all four ring systems are interesting to scientists because of their complicated forms, influenced by the pull of the moons that also orbit these giant planets.

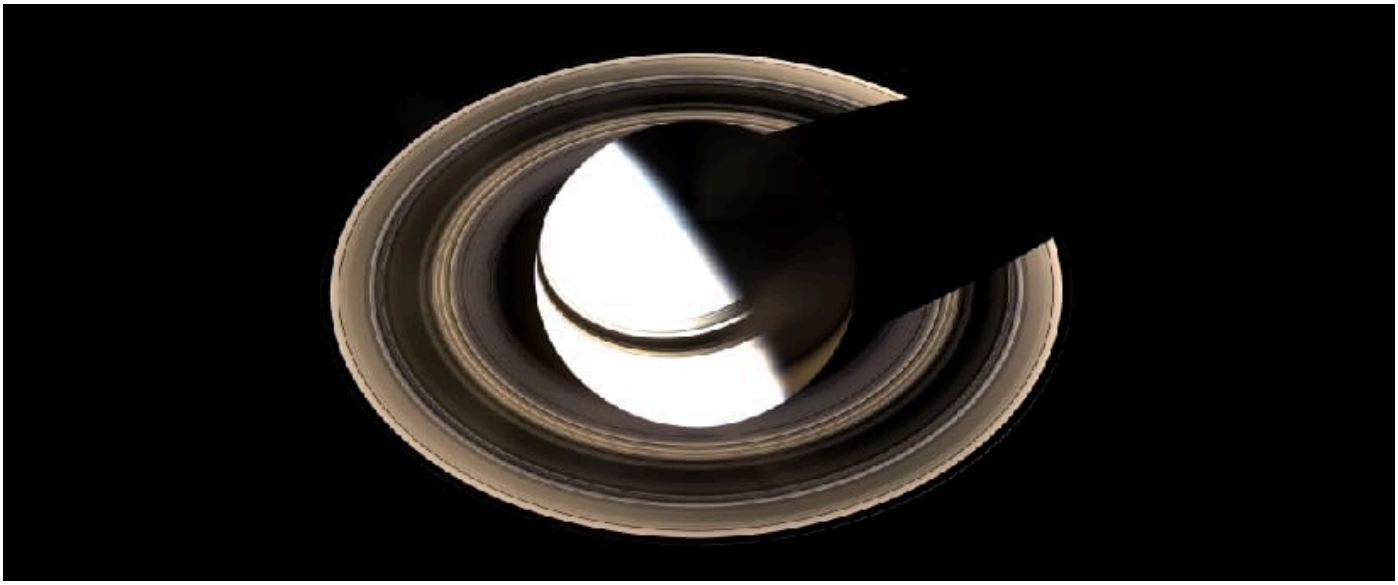


Figure 7.7 : Saturn and Its Rings. This 2007 Cassini image shows Saturn and its complex system of rings, taken from a distance of about 1.2 million kilometers. This natural-color image is a composite of 36 images taken over the course of 2.5 hours. (credit: modification of work by NASA/JPL/Space Science Institute)

The solar system has many other less-conspicuous members. Another group is the asteroids, rocky bodies that orbit the Sun like miniature planets, mostly in the space between Mars and Jupiter (although some do cross the orbits of planets like Earth—see Figure 7.8). Most asteroids are remnants of the initial population of the solar system that existed before the planets themselves formed. Some of the smallest moons of the planets, such as the moons of Mars, are very likely captured asteroids.

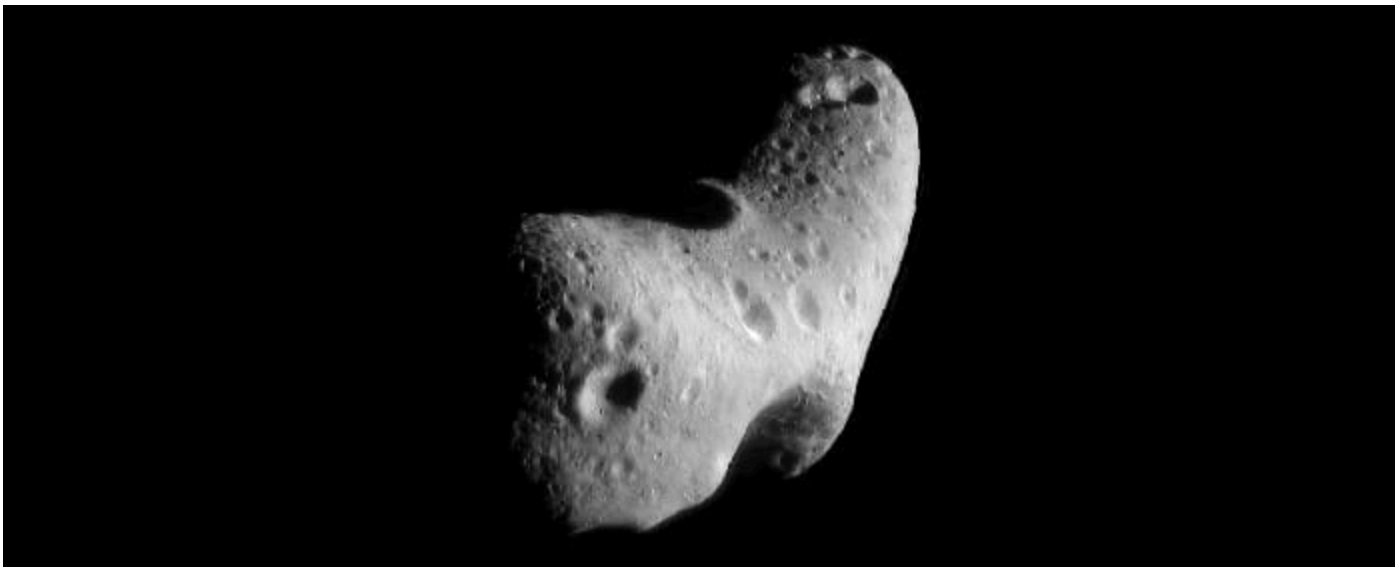


Figure 7.8 : Asteroid Eros. This small Earth-crossing asteroid image was taken by the NEAR-Shoemaker spacecraft from an altitude of about 100 kilometers. This view of the heavily cratered surface is about 10 kilometers wide. The spacecraft orbited Eros for a year before landing gently on its surface. (credit: modification of work by NASA/JHUAPL)

Another class of small bodies is composed mostly of ice, made of frozen gases such as water, carbon dioxide, and carbon monoxide; these objects are called comets (see Figure 7.9). Comets also are remnants from the formation of the solar system, but they were formed and continue (with rare exceptions) to orbit the Sun in distant, cooler regions—stored in a sort of cosmic deep freeze. This is also the realm of the larger icy worlds, called dwarf planets.

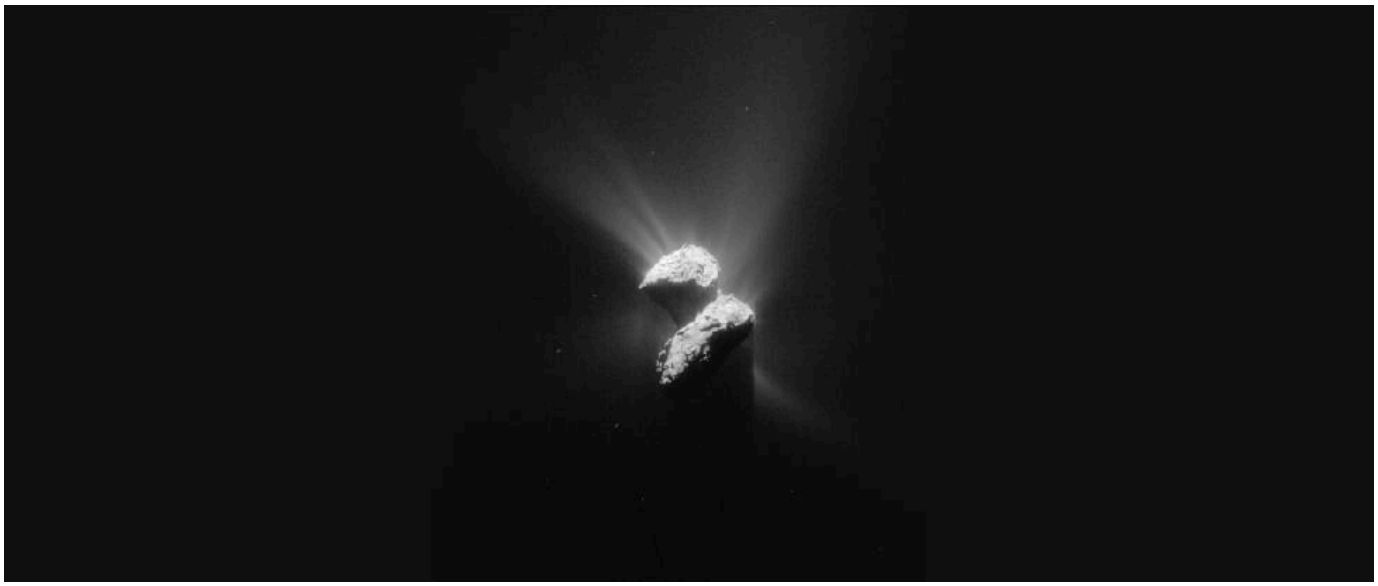


Figure 7.9 : Comet Churyumov-Gerasimenko (67P). This image shows Comet Churyumov-Gerasimenko, also known as 67P, near its closest approach to the Sun in 2015, as seen from the *Rosetta* spacecraft. Note the jets of gas escaping from the solid surface. (credit: modification of work by ESA/Rosetta/NAVACAM, CC BY-SA IGO 3.0)

Finally, there are countless grains of broken rock, which we call cosmic dust, scattered throughout the solar system. When these particles enter Earth's atmosphere (as millions do each day) they burn up, producing a brief flash of light in the night sky known as a meteor (meteors are often referred to as shooting stars). Occasionally, some larger chunk of rocky or metallic material survives its passage through the atmosphere and lands on Earth. Any piece that strikes the ground is known as a meteorite. (You can see meteorites on display in many natural history museums and can sometimes even purchase pieces of them from gem and mineral dealers.)

Voyagers in Astronomy: Carl Sagan: Solar System Advocate

The best-known astronomer in the world during the 1970s and 1980s, Carl Sagan devoted most of his professional career to studying the planets and considerable energy to raising public awareness of what we can learn from exploring the solar system (see Figure 7.10). Born in Brooklyn, New York, in 1934, Sagan became interested in astronomy as a youngster; he also credits science fiction stories for sustaining his fascination with what's out in the universe.

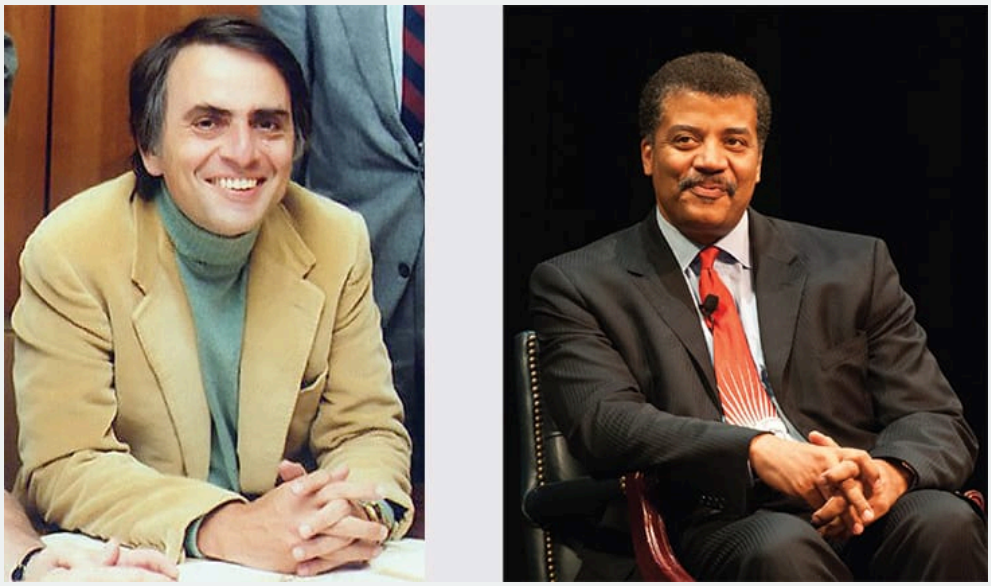


Figure 7.10 : Carl Sagan (1934–1996) and Neil deGrasse Tyson. Sagan was Tyson's inspiration to become a scientist. (credit "Sagan": modification of work by NASA, JPL; credit "Tyson": modification of work by Bruce F. Press)

In the early 1960s, when many scientists still thought Venus might turn out to be a hospitable place, Sagan calculated that the thick atmosphere of Venus could act like a giant greenhouse, keeping the heat in and raising the temperature enormously. He showed that the seasonal changes astronomers had seen on Mars were caused, not by vegetation, but by wind-blown dust. He was a member of the scientific teams for many of the robotic missions that explored the solar system and was instrumental in getting NASA to put a message-bearing plaque aboard the Pioneer spacecraft, as well as audio-video records on the Voyager spacecraft—all of them destined to leave our solar system entirely and send these little bits of Earth technology out among the stars.

To encourage public interest and public support of planetary exploration, Sagan helped found The Planetary Society, now the largest space-interest organization in the world. He was a tireless and eloquent advocate of the need to study the solar system close-up and the value of learning about other worlds in order to take better care of our own.

Sagan simulated conditions on early Earth to demonstrate how some of life's fundamental building blocks might have formed from the "primordial soup" of natural compounds on our planet. In addition, he and his colleagues developed computer models showing the consequences of nuclear war for Earth would be even more devastating than anyone had thought (this is now called the nuclear winter hypothesis) and demonstrating some of the serious consequences of continued pollution of our atmosphere.

Sagan was perhaps best known, however, as a brilliant popularizer of astronomy and the author of many books on science, including the best-selling *Cosmos*, and several evocative tributes to solar system exploration such as *The Cosmic Connection* and *Pale Blue Dot*. His book *The Demon Haunted World*, completed just before his death in 1996, is perhaps the best antidote to fuzzy thinking about pseudo-science and irrationality in print today. An intriguing science fiction novel he wrote, titled *Contact*, which became a successful film as well, is still recommended by many science instructors as a scenario for making contact with life elsewhere that is much more reasonable than most science fiction.

Sagan was a master, too, of the television medium. His 13-part public television series, *Cosmos*, was seen by an estimated 500 million people in 60 countries and has become one of the most-watched series in the history of public broadcasting. A few astronomers scoffed at a scientist who spent so much time in the public eye, but it is probably fair to say that Sagan's enthusiasm and skill as an explainer won more friends for the science of astronomy than anyone or anything else in the second half of the twentieth century.

In the two decades since Sagan's death, no other scientist has achieved the same level of public recognition. Perhaps closest is the director of the Hayden Planetarium, Neil deGrasse Tyson, who followed in Sagan's footsteps by making an updated version of the *Cosmos* program in 2014. Tyson is quick to point out that Sagan was his inspiration to become a scientist, telling how Sagan invited him to visit for a day at Cornell when he was a high school student looking for a career. However, the media environment has fragmented a great deal since Sagan's time. It is interesting to speculate whether Sagan could have adapted his communication style to the world of cable television, Twitter, Facebook, and podcasts.

Link to Learning

Two imaginative videos provide a tour of the solar system objects we have been discussing. Shane Gellert's I Need Some Space uses NASA photography and models to show the various worlds with which we share our system. In the more science fiction-oriented Wanderers video, we see some of the planets and moons as tourist destinations for future explorers, with commentary taken from recordings by Carl Sagan.

A Scale Model of the Solar System

Astronomy often deals with dimensions and distances that far exceed our ordinary experience. What does 1.4 billion kilometers—the distance from the Sun to Saturn—really mean to anyone? It can be helpful to visualize such large systems in terms of a scale model.

In our imaginations, let us build a scale model of the solar system, adopting a scale factor of 1 billion (10^9)—that is, reducing the actual solar system by dividing every dimension by a factor of 10^9. Earth, then, has a diameter of 1.3 centimeters, about the size of a grape. The Moon is a pea orbiting this at a distance of 40 centimeters, or a little more than a foot away. The Earth-Moon system fits into a standard backpack.

In this model, the Sun is nearly 1.5 meters in diameter, about the average height of an adult, and our Earth is at a distance of 150 meters—about one city block—from the Sun. Jupiter is five blocks away from the Sun, and its diameter is 15 centimeters, about the size of a very large grapefruit. Saturn is 10 blocks from the Sun; Uranus, 20 blocks; and Neptune, 30 blocks. Pluto, with a distance that varies quite a bit during its 249-year orbit, is currently just beyond 30 blocks and getting farther with time. Most of the moons of the outer solar system are the sizes of various kinds of seeds orbiting the grapefruit, oranges, and lemons that represent the outer planets.

In our scale model, a human is reduced to the dimensions of a single atom, and cars and spacecraft to the size of molecules. Sending the Voyager spacecraft to Neptune involves navigating a single molecule from the Earth–grape toward a lemon 5 kilometers away with an accuracy equivalent to the width of a thread in a spider's web.

If that model represents the solar system, where would the nearest stars be? If we keep the same scale, the closest stars would be tens of thousands of kilometers away. If you built this scale model in the city where you live, you would have to place the representations of these stars on the other side of Earth or beyond.

By the way, model solar systems like the one we just presented have been built in cities throughout the world. In Sweden, for example, Stockholm's huge Globe Arena has become a model for the Sun, and Pluto is represented by a 12-centimeter sculpture in the small town of Delsbo, 300 kilometers away. Another model solar system is in Washington on the Mall between the White House and Congress (perhaps proving they are worlds apart?).

Link to Learning

This model of the solar system shows all orbits and sizes to scale, and it lets you fly between the planets at an enhanced speed.

Making Connections: Names in the Solar System

We humans just don't feel comfortable until something has a name. Types of butterflies, new elements, and the mountains of Venus all need names for us to feel we are acquainted with them. How do we give names to objects and features in the solar system?

Planets and moons are named after gods and heroes in Greek and Roman mythology (with a few exceptions among the moons of Uranus, which have names drawn from English literature). When William Herschel, a German immigrant to England, first discovered the planet we now call Uranus, he wanted to name it Georgium Sidus (George's star) after King George III of his adopted country. This caused such an outcry among astronomers in other nations, however, that the classic tradition was upheld—and has been maintained ever since. Luckily, there were a lot of minor gods in the ancient pantheon, so plenty of names are left for the many small moons we are discovering around the giant planets. (Appendix G lists the larger moons). More recently, the names of dwarf planets and their moons have been drawn from the mythology of other cultures besides Greek and Roman.

Comets are often named after their discoverers (offering an extra incentive to comet hunters). Asteroids are named by their discoverers after just about anyone or anything they want. Recently, asteroid names have been used to recognize people who have made significant contributions to astronomy, including the three senior authors of this book.

That was pretty much all the naming that was needed while our study of the solar system was confined to Earth. But now, our spacecraft have surveyed and photographed many worlds in great detail, and each world has a host of features that also need names. To make sure that naming things in space remains multinational, rational, and somewhat dignified, astronomers have given the responsibility of approving names to a special committee of the International Astronomical Union (IAU), the body that includes scientists from every country that does astronomy.

This IAU committee has developed a set of rules for naming features on other worlds. For example, craters on Venus are named for women who have made significant contributions to human knowledge and welfare. Volcanic features on Jupiter's moon Io, which is in a constant state of volcanic activity, are named after gods of fire and thunder from the mythologies of many cultures. Craters on Mercury commemorate famous novelists, playwrights, artists, and composers. On Saturn's moon Tethys, all the features are named after characters and places in Homer's great epic poem, *The Odyssey*. As we explore further, it may well turn out that more places in the solar system need names than Earth history can provide. Perhaps by then, explorers and settlers on these worlds will be ready to develop their own names for the places they may (if but for a while) call home.

You may be surprised to know that the meaning of the word *planet* has recently become controversial because we have discovered many other planetary systems that don't look very much like our own. Even within our solar system, the planets differ greatly in size and chemical properties. The biggest dispute concerns Pluto, which is much smaller than the other eight major planets. The category of dwarf planet was invented to include Pluto and similar icy objects beyond Neptune. But is a dwarf planet also a planet? Logically, it should be, but even this simple issue of grammar has been the subject of heated debate among both astronomers and the general public.

Footnotes

- 1The generic term for a group of planets and other bodies circling a star is *planetary system*. Ours is called the *solar system* because our Sun is sometimes called *Sol*. Strictly speaking, then, there is only one solar system; planets orbiting other stars are in planetary systems.
- 2An AU (or astronomical unit) is the distance from Earth to the Sun.
- 3We give densities in units where the density of water is 1 g/cm^3. To get densities in units of kg/m^3, multiply the given value by 1000.

Answers

Exercise 6.1.1

For a sphere,
$\text{density} = \frac{\text{mass}}{(\frac{4}{3}\pi R^3)}\ \text{kg/m}^3$.
For Earth, then,
$\text{density} = \frac{6 \times 10^{24}\,\text{kg}}{4.2 \times 2.6 \times 10^{20}\,\text{m}^3} = 5.5 \times 10^3\,\text{kg/m}^3$.
This density is four to five times greater than Mimas'. In fact, Earth is the densest of the planets.

This page titled 6.1: Overview of Our Planetary System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 7.2: Overview of Our Planetary System by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

6.2: Composition and Structure of Planets

Learning Objectives

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

- Describe the characteristics of the giant planets, terrestrial planets, and small bodies in the solar system
- Explain what influences the temperature of a planet's surface
- Explain why there is geological activity on some planets and not on others

The fact that there are two distinct kinds of planets—the rocky terrestrial planets and the gas-rich jovian planets—leads us to believe that they formed under different conditions. Certainly their compositions are dominated by different elements. Let us look at each type in more detail.

The Giant Planets

The two largest planets, Jupiter and Saturn, have nearly the same chemical makeup as the Sun; they are composed primarily of the two elements hydrogen and helium, with 75% of their mass being hydrogen and 25% helium. On Earth, both hydrogen and helium are gases, so Jupiter and Saturn are sometimes called gas planets. But, this name is misleading. Jupiter and Saturn are so large that the gas is compressed in their interior until the hydrogen becomes a liquid. Because the bulk of both planets consists of compressed, liquefied hydrogen, we should really call them liquid planets.

Under the force of gravity, the heavier elements sink toward the inner parts of a liquid or gaseous planet. Both Jupiter and Saturn, therefore, have cores composed of heavier rock, metal, and ice, but we cannot see these regions directly. In fact, when we look down from above, all we see is the atmosphere with its swirling clouds (Figure 7.11). We must infer the existence of the denser core inside these planets from studies of each planet's gravity.



Figure 7.11 : Jupiter. This true-color image of Jupiter was taken from the Cassini spacecraft in 2000. The dark spot is the shadow of one of the giant planet's moons. (credit: modification of work by NASA/JPL/University of Arizona)

Uranus and Neptune are much smaller than Jupiter and Saturn, but each also has a core of rock, metal, and ice. Uranus and Neptune were less efficient at attracting hydrogen and helium gas, so they have much smaller atmospheres in proportion to their cores.

Chemically, each giant planet is dominated by hydrogen and its many compounds. Nearly all the oxygen present is combined

chemically with hydrogen to form water (H_2O). Chemists call such a hydrogen-dominated composition *reduced*. Throughout the outer solar system, we find abundant water (mostly in the form of ice) and reducing chemistry.

The Terrestrial Planets

The terrestrial planets are quite different from the giants. In addition to being much smaller, they are composed primarily of rocks and metals. These, in turn, are made of elements that are less common in the universe as a whole. The most abundant rocks, called silicates, are made of silicon and oxygen, and the most common metal is iron. We can tell from their densities (see Table 7.2) that Mercury has the greatest proportion of metals (which are denser) and the Moon has the lowest. Earth, Venus, and Mars all have roughly similar bulk compositions: about one third of their mass consists of iron-nickel or iron-sulfur combinations; two thirds is made of silicates. Because these planets are largely composed of oxygen compounds (such as the silicate minerals of their crusts), their chemistry is said to be *oxidized*.

When we look at the internal structure of each of the terrestrial planets, we find that the densest metals are in a central core, with the lighter silicates near the surface. If these planets were liquid, like the giant planets, we could understand this effect as the result the sinking of heavier elements due to the pull of gravity. This leads us to conclude that, although the terrestrial planets are solid today, at one time they must have been hot enough to melt.

Differentiation is the process by which gravity helps separate a planet's interior into layers of different compositions and densities. The heavier metals sink to form a core, while the lightest minerals float to the surface to form a crust. Later, when the planet cools, this layered structure is preserved. In order for a rocky planet to differentiate, it must be heated to the melting point of rocks, which is typically more than 1300 K.

Moons, Asteroids, and Comets

Chemically and structurally, Earth's Moon is like the terrestrial planets, but most moons are in the outer solar system, and they have compositions similar to the cores of the giant planets around which they orbit. The three largest moons—Ganymede and Callisto in the jovian system, and Titan in the saturnian system—are composed half of frozen water, and half of rocks and metals. Most of these moons differentiated during formation, and today they have cores of rock and metal, with upper layers and crusts of very cold and—thus very hard—ice (Figure 7.12).



Figure 7.12 : Ganymede. This view of Jupiter's moon Ganymede was taken in June 1996 by the Galileo spacecraft. The brownish gray color of the surface indicates a dusty mixture of rocky material and ice. The bright spots are places where recent impacts have uncovered fresh ice from underneath. (credit: modification of work by NASA/JPL)

Most of the asteroids and comets, as well as the smallest moons, were probably never heated to the melting point. However, some of the largest asteroids, such as Vesta, appear to be differentiated; others are fragments from differentiated bodies. Many of the smaller objects seem to be fragments or rubble piles that are the result of collisions. Because most asteroids and comets retain

their original composition, they represent relatively unmodified material dating back to the time of the formation of the solar system. In a sense, they act as chemical fossils, helping us to learn about a time long ago whose traces have been erased on larger worlds.

Temperatures: Going to Extremes

Generally speaking, the farther a planet or moon is from the Sun, the cooler its surface. The planets are heated by the radiant energy of the Sun, which gets weaker with the square of the distance. You know how rapidly the heating effect of a fireplace or an outdoor radiant heater diminishes as you walk away from it; the same effect applies to the Sun. Mercury, the closest planet to the Sun, has a blistering surface temperature that ranges from 280–430 °C on its sunlit side, whereas the surface temperature on Pluto is only about –220 °C, colder than liquid air.

Mathematically, the temperatures decrease approximately in proportion to the square root of the distance from the Sun. Pluto is about 30 AU at its closest to the Sun (or 100 times the distance of Mercury) and about 49 AU at its farthest from the Sun. Thus, Pluto's temperature is less than that of Mercury by the square root of 100, or a factor of 10: from 500 K to 50 K.

In addition to its distance from the Sun, the surface temperature of a planet can be influenced strongly by its atmosphere. Without our atmospheric insulation (the greenhouse effect, which keeps the heat in), the oceans of Earth would be permanently frozen. Conversely, if Mars once had a larger atmosphere in the past, it could have supported a more temperate climate than it has today. Venus is an even more extreme example, where its thick atmosphere of carbon dioxide acts as insulation, reducing the escape of heat built up at the surface, resulting in temperatures greater than those on Mercury. Today, Earth is the only planet where surface temperatures generally lie between the freezing and boiling points of water. As far as we know, Earth is the only planet to support life.

Astronomy Basics: There's No Place Like Home

In the classic film *The Wizard of Oz*, Dorothy, the heroine, concludes after her many adventures in "alien" environments that "there's no place like home." The same can be said of the other worlds in our solar system. There are many fascinating places, large and small, that we might like to visit, but humans could not survive on any without a great deal of artificial assistance.

A thick carbon dioxide atmosphere keeps the surface temperature on our neighbor Venus at a sizzling 700 K (near 900 °F). Mars, on the other hand, has temperatures generally below freezing, with air (also mostly carbon dioxide) so thin that it resembles that found at an altitude of 30 kilometers (100,000 feet) in Earth's atmosphere. And the red planet is so dry that it has not had any rain for billions of years.

The outer layers of the jovian planets are neither warm enough nor solid enough for human habitation. Any bases we build in the systems of the giant planets may well have to be in space or one of their moons—none of which is particularly hospitable to a luxury hotel with a swimming pool and palm trees. Perhaps we will find warmer havens deep inside the clouds of Jupiter or in the ocean under the frozen ice of its moon Europa.

All of this suggests that we had better take good care of Earth because it is the only site where life as we know it could survive. Recent human activity may be reducing the habitability of our planet by adding pollutants to the atmosphere, especially the potent greenhouse gas carbon dioxide. Human civilization is changing our planet dramatically, and these changes are not necessarily for the better. In a solar system that seems unready to receive us, making Earth less hospitable to life may be a grave mistake.

Geological Activity

The crusts of all of the terrestrial planets, as well as of the larger moons, and Pluto, have been modified over their histories by both internal and external forces. Externally, each has been battered by a slow rain of projectiles from space, leaving their surfaces pockmarked by impact craters of all sizes (see Figure 7.4). We have good evidence that this bombardment was far greater in the early history of the solar system, but it certainly continues to this day, even if at a lower rate. The collision of more than 20 large pieces of Comet Shoemaker–Levy 9 with Jupiter in the summer of 1994 (see Figure 7.13) is one dramatic example of this process.



Figure 7.13 : Comet Shoemaker–Levy 9. In this image of Comet Shoemaker–Levy 9 taken on May 17, 1994, by NASA's Hubble Space Telescope, you can see about 20 icy fragments into which the comet broke. The comet was approximately 660 million kilometers from Earth, heading on a collision course with Jupiter. (credit: modification of work by NASA, ESA, H. Weaver (STScI), E. Smith (STScI))

Figure 7.14 shows the aftermath of these collisions, when debris clouds larger than Earth could be seen in Jupiter's atmosphere.

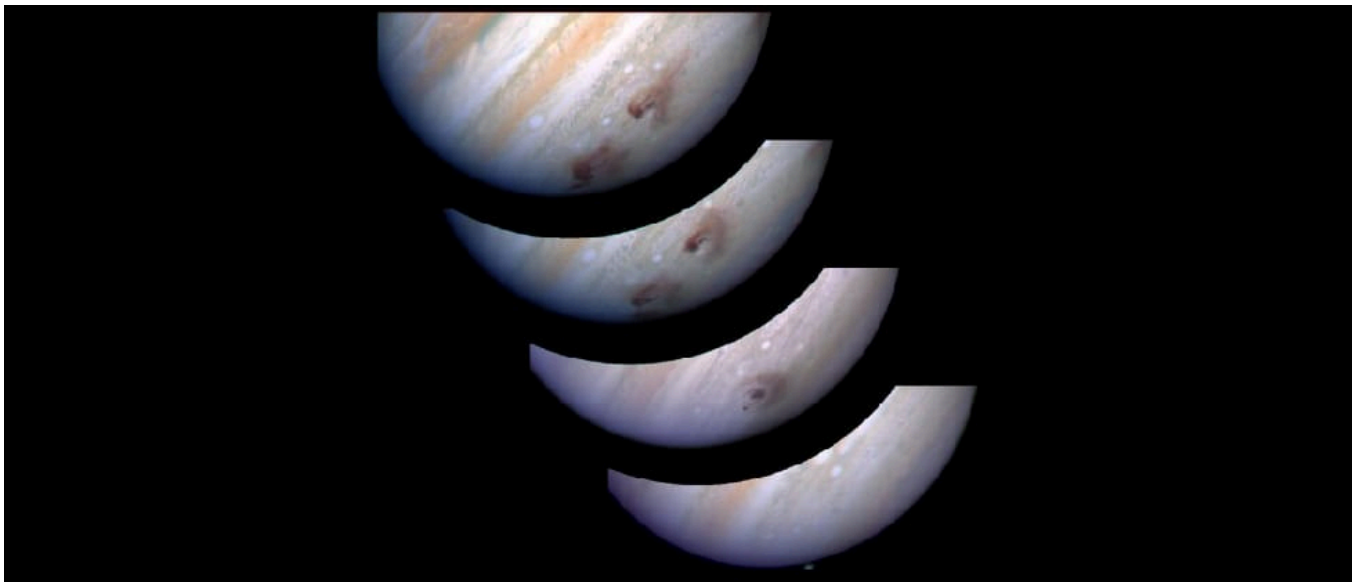


Figure 7.14 : Jupiter with Huge Dust Clouds. The Hubble Space Telescope took this sequence of images of Jupiter in summer 1994, when fragments of Comet Shoemaker–Levy 9 collided with the giant planet. Here we see the site hit by fragment G, from five minutes to five days after impact. Several of the dust clouds generated by the collisions became larger than Earth. (credit: modification of work by H. Hammel, NASA)

During the time all the planets have been subject to such impacts, internal forces on the terrestrial planets have buckled and twisted their crusts, built up mountain ranges, erupted as volcanoes, and generally reshaped the surfaces in what we call geological activity. (The prefix *geo* means "Earth," so this is a bit of an "Earth-chauvinist" term, but it is so widely used that we bow to tradition.) Among the terrestrial planets, Earth and Venus have experienced the most geological activity over their histories, although some of the moons in the outer solar system are also surprisingly active. In contrast, our own Moon is a dead world where geological activity ceased billions of years ago.

Geological activity on a planet is the result of a hot interior. The forces of volcanism and mountain building are driven by heat escaping from the interiors of planets. As we will see, each of the planets was heated at the time of its birth, and this primordial heat initially powered extensive volcanic activity, even on our Moon. But, small objects such as the Moon soon cooled off. The larger the planet or moon, the longer it retains its internal heat, and therefore the more we expect to see surface evidence of continuing geological activity. The effect is similar to our own experience with a hot baked potato: the larger the potato, the more slowly it cools. If we want a potato to cool quickly, we cut it into small pieces.

For the most part, the history of volcanic activity on the terrestrial planets conforms to the predictions of this simple theory. The

Moon, the smallest of these objects, is a geologically dead world. Although we know less about Mercury, it seems likely that this planet, too, ceased most volcanic activity about the same time the Moon did. Mars represents an intermediate case. It has been much more active than the Moon, but less so than Earth. Earth and Venus, the largest terrestrial planets, still have molten interiors even today, some 4.5 billion years after their birth.

This page titled 6.2: Composition and Structure of Planets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 7.3: Compostion and Structure of Planets by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

6.3: Dating Planetary Surfaces

Learning Objectives

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

- Explain how astronomers can tell whether a planetary surface is geologically young or old
- Describe different methods for dating planets

How do we know the age of the surfaces we see on planets and moons? If a world has a surface (as opposed to being mostly gas and liquid), astronomers have developed some techniques for estimating how long ago that surface solidified. Note that the age of these surfaces is not necessarily the age of the planet as a whole. On geologically active objects (including Earth), vast outpourings of molten rock or the erosive effects of water and ice, which we call planet weathering, have erased evidence of earlier epochs and present us with only a relatively young surface for investigation.

Counting the Craters

One way to estimate the age of a surface is by counting the number of impact craters. This technique works because the rate at which impacts have occurred in the solar system has been roughly constant for several billion years. Thus, in the absence of forces to eliminate craters, the number of craters is simply proportional to the length of time the surface has been exposed. This technique has been applied successfully to many solid planets and moons (Figure 7.15).

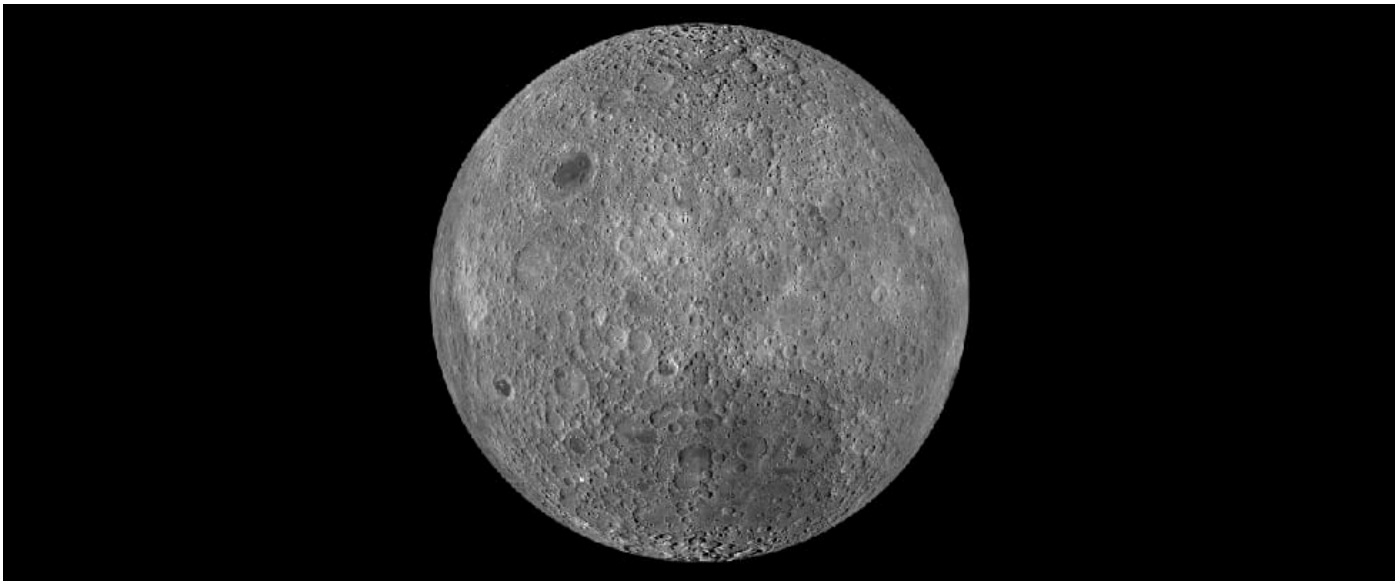


Figure 7.15 : Our Cratered Moon. This composite image of the Moon's surface was made from many smaller images taken between November 2009 and February 2011 by the Lunar Reconnaissance Orbiter (LRO) and shows craters of many different sizes. (credit: modification of work by NASA/GSFC/Arizona State University)

Bear in mind that crater counts can tell us only the time since the surface experienced a major change that could modify or erase preexisting craters. Estimating ages from crater counts is a little like walking along a sidewalk in a snowstorm after the snow has been falling steadily for a day or more. You may notice that in front of one house the snow is deep, while next door the sidewalk may be almost clear. Do you conclude that less snow has fallen in front of Ms. Jones' house than Mr. Smith's? More likely, you conclude that Jones has recently swept the walk clean and Smith has not. Similarly, the numbers of craters indicate how long it has been since a planetary surface was last "swept clean" by ongoing lava flows or by molten materials ejected when a large impact happened nearby.

Still, astronomers can use the numbers of craters on different parts of the same world to provide important clues about how

regions on that world evolved. On a given planet or moon, the more heavily cratered terrain will generally be older (that is, more time will have elapsed there since something swept the region clean).

Radioactive Rocks

Another way to trace the history of a solid world is to measure the age of individual rocks. After samples were brought back from the Moon by Apollo astronauts, the techniques that had been developed to date rocks on Earth were applied to rock samples from the Moon to establish a geological chronology for the Moon. Furthermore, a few samples of material from the Moon, Mars, and the large asteroid Vesta have fallen to Earth as meteorites and can be examined directly (see the chapter on Cosmic Samples and the Origin of the Solar System).

Scientists measure the age of rocks using the properties of natural radioactivity. Around the beginning of the twentieth century, physicists began to understand that some atomic nuclei are not stable but can split apart (decay) spontaneously into smaller nuclei. The process of radioactive decay involves the emission of particles such as electrons, or of radiation in the form of gamma rays (see the chapter on Radiation and Spectra).

For any one radioactive nucleus, it is not possible to predict when the decay process will happen. Such decay is random in nature, like the throw of dice: as gamblers have found all too often, it is impossible to say just when the dice will come up 7 or 11. But, for a very large number of dice tosses, we can calculate the odds that 7 or 11 will come up. Similarly, if we have a very large number of radioactive atoms of one type (say, uranium), there is a specific time period, called its half-life, during which the chances are fifty-fifty that decay will occur for any of the nuclei.

A particular nucleus may last a shorter or longer time than its half-life, but in a large sample, almost exactly half of the nuclei will have decayed after a time equal to one half-life. Half of the remaining nuclei will have decayed after two half-lives pass, leaving only one half of a half—or one quarter—of the original sample (Figure 7.16).

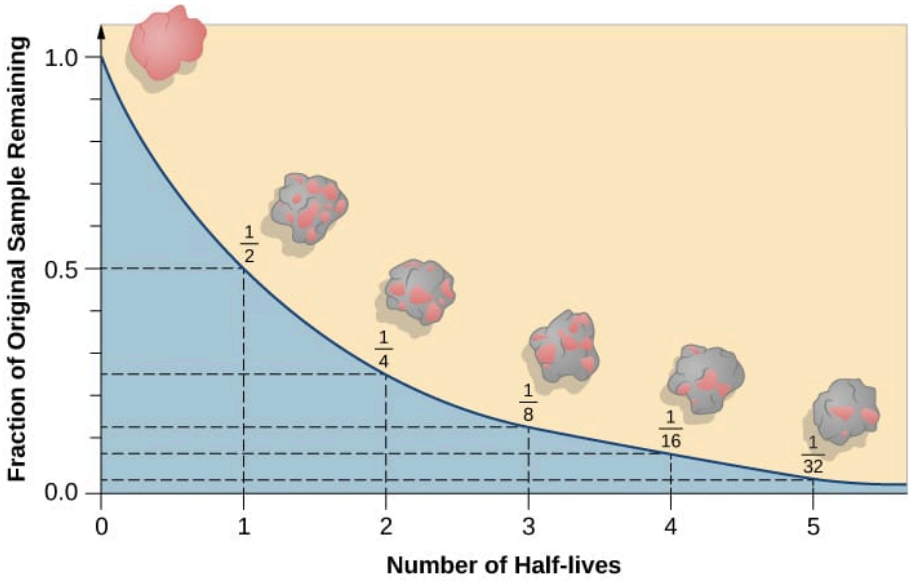


Figure 7.16 : Radioactive Decay. This graph shows (in pink) the amount of a radioactive sample that remains after several half-lives have passed. After one half-life, half the sample is left; after two half-lives, one half of the remainder (or one quarter) is left; and after three half-lives, one half of that (or one eighth) is left. Note that, in reality, the decay of radioactive elements in a rock sample would not cause any visible change in the appearance of the rock; the splashes of color are shown here for conceptual purposes only.

If you had 1 gram of pure radioactive nuclei with a half-life of 100 years, then after 100 years you would have 1/2 gram; after 200 years, 1/4 gram; after 300 years, only 1/8 gram; and so forth. However, the material does not disappear. Instead, the radioactive atoms are replaced with their decay products. Sometimes the radioactive atoms are called *parents* and the decay products are called *daughter* elements.

In this way, radioactive elements with half-lives we have determined can provide accurate nuclear clocks. By comparing how much of a radioactive parent element is left in a rock to how much of its daughter products have accumulated, we can learn how long the decay process has been going on and hence how long ago the rock formed. Table 7.3 summarizes the decay reactions used

most often to date lunar and terrestrial rocks.

Table 7.3: Radioactive Decay Reaction Used to Date Rocks[4]

Parent	Daughter	Half-Life (billions of years)
Samarium-147	Neodymium-143	106
Rubidium-87	Strontium-87	48.8
Thorium-232	Lead-208	14.0
Uranium-238	Lead-206	4.47
Potassium-40	Argon-40	1.31

Link to Learning

PBS provides an evolution series excerpt that explains how we use radioactive elements to date Earth.

This Science Channel video features Bill Nye the Science Guy showing how scientists have used radioactive dating to determine the age of Earth.

This radioactive decay simulator lets you examine how the number of atoms of a radioactive element decreases over time and see how the ratio of parent to child atoms could be used to figure out the age of the whole set of atoms.

When astronauts first flew to the Moon, one of their most important tasks was to bring back lunar rocks for radioactive age-dating. Until then, astronomers and geologists had no reliable way to measure the age of the lunar surface. Counting craters had let us calculate relative ages (for example, the heavily cratered lunar highlands were older than the dark lava plains), but scientists could not measure the actual age in years. Some thought that the ages were as young as those of Earth's surface, which has been resurfaced by many geological events. For the Moon's surface to be so young would imply active geology on our satellite. Only in 1969, when the first Apollo samples were dated, did we learn that the Moon is an ancient, geologically dead world. Using such dating techniques, we have been able to determine the ages of both Earth and the Moon: each was formed about 4.5 billion years ago (although, as we shall see, Earth probably formed earlier than the Moon).

We should also note that the decay of radioactive nuclei generally releases energy in the form of heat. Although the energy from a single nucleus is not very large (in human terms), the enormous numbers of radioactive nuclei in a planet or moon (especially early in its existence) can be a significant source of internal energy for that world. Geologists estimate that about half of Earth's current internal heat budget comes from the decay of radioactive isotopes in its interior.

Footnotes

- 4The number after each element is its atomic weight, equal to the number of protons plus neutrons in its nucleus. This specifies the *isotope* of the element; different isotopes of the same element differ in the number of neutrons.

This page titled 6.3: Dating Planetary Surfaces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 7.4: Dating Planetary Surfaces by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

6.4: Origin of the Solar System

Learning Objectives

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

- Describe the characteristics of planets that are used to create formation models of the solar system
- Describe how the characteristics of extrasolar systems help us to model our own solar system
- Explain the importance of collisions in the formation of the solar system

Much of astronomy is motivated by a desire to understand the origin of things: to find at least partial answers to age-old questions of where the universe, the Sun, Earth, and we ourselves came from. Each planet and moon is a fascinating place that may stimulate our imagination as we try to picture what it would be like to visit. Taken together, the members of the solar system preserve patterns that can tell us about the formation of the entire system. As we begin our exploration of the planets, we want to introduce our modern picture of how the solar system formed.

The recent discovery of thousands of planets in orbit around other stars has shown astronomers that many exoplanetary systems can be quite different from our own solar system. For example, it is common for these systems to include planets intermediate in size between our terrestrial and giant planets. These are often called *superearths*. Some exoplanet systems even have giant planets close to the star, reversing the order we see in our system. In The Birth of Stars and the Discovery of Planets outside the Solar System, we will look at these exoplanet systems. But for now, let us focus on theories of how our own particular system has formed and evolved.

Looking for Patterns

One way to approach our question of origin is to look for regularities among the planets. We found, for example, that all the planets lie in nearly the same plane and revolve in the same direction around the Sun. The Sun also spins in the same direction about its own axis. Astronomers interpret this pattern as evidence that the Sun and planets formed together from a spinning cloud of gas and dust that we call the solar nebula (Figure 7.17).



Figure 7.17 : Solar Nebula. This artist's conception of the solar nebula shows the flattened cloud of gas and dust from which our planetary system formed. Icy and rocky planetesimals (precursors of the planets) can be seen in the foreground. The bright center is where the Sun is forming. (credit: William K. Hartmann, Planetary Science Institute)

The composition of the planets gives another clue about origins. Spectroscopic analysis allows us to determine which elements

are present in the Sun and the planets. The Sun has the same hydrogen-dominated composition as Jupiter and Saturn, and therefore appears to have been formed from the same reservoir of material. In comparison, the terrestrial planets and our Moon are relatively deficient in the light gases and the various ices that form from the common elements oxygen, carbon, and nitrogen. Instead, on Earth and its neighbors, we see mostly the rarer heavy elements such as iron and silicon. This pattern suggests that the processes that led to planet formation in the inner solar system must somehow have excluded much of the lighter materials that are common elsewhere. These lighter materials must have escaped, leaving a residue of heavy stuff.

The reason for this is not hard to guess, bearing in mind the heat of the Sun. The inner planets and most of the asteroids are made of rock and metal, which can survive heat, but they contain very little ice or gas, which evaporate when temperatures are high. (To see what we mean, just compare how long a rock and an ice cube survive when they are placed in the sunlight.) In the outer solar system, where it has always been cooler, the planets and their moons, as well as icy dwarf planets and comets, are composed mostly of ice and gas.

The Evidence from Far Away

A second approach to understanding the origins of the solar system is to look outward for evidence that other systems of planets are forming elsewhere. We cannot look back in time to the formation of our own system, but many stars in space are much younger than the Sun. In these systems, the processes of planet formation might still be accessible to direct observation. We observe that there are many other “solar nebulas” or *circumstellar disks*—flattened, spinning clouds of gas and dust surrounding young stars. These disks resemble our own solar system’s initial stages of formation billions of years ago (Figure 7.18).



Figure 7.18 : Atlas of Planetary Nurseries. These Hubble Space Telescope photos show sections of the Orion Nebula, a relatively close-by region where stars are currently forming. Each image shows an embedded circumstellar disk orbiting a very young star. Seen from different angles, some are energized to glow by the light of a nearby star while others are dark and seen in silhouette against the bright glowing gas of the Orion Nebula. Each is a contemporary analog of our own solar nebula—a location where planets are probably being formed today. (credit: modification of work by NASA/ESA, L. Ricci (ESO))

Building Planets

Circumstellar disks are a common occurrence around very young stars, suggesting that disks and stars form together. Astronomers can use theoretical calculations to see how solid bodies might form from the gas and dust in these disks as they cool. These models show that material begins to coalesce first by forming smaller objects, precursors of the planets, which we call planetesimals.

Today’s fast computers can simulate the way millions of planetesimals, probably no larger than 100 kilometers in diameter, might gather together under their mutual gravity to form the planets we see today. We are beginning to understand that this process was a violent one, with planetesimals crashing into each other and sometimes even disrupting the growing planets themselves. As a consequence of those violent impacts (and the heat from radioactive elements in them), all the planets were heated until

they were liquid and gas, and therefore differentiated, which helps explain their present internal structures.

The process of impacts and collisions in the early solar system was complex and, apparently, often random. The solar nebula model can explain many of the regularities we find in the solar system, but the random collisions of massive collections of planetesimals could be the reason for some exceptions to the “rules” of solar system behavior. For example, why do the planets Uranus and Pluto spin on their sides? Why does Venus spin slowly and in the opposite direction from the other planets? Why does the composition of the Moon resemble Earth in many ways and yet exhibit substantial differences? The answers to such questions probably lie in enormous collisions that took place in the solar system long before life on Earth began.

Today, some 4.5 billion years after its origin, the solar system is—thank goodness—a much less violent place. As we will see, however, some planetesimals have continued to interact and collide, and their fragments move about the solar system as roving “transients” that can make trouble for the established members of the Sun’s family, such as our own Earth. (We discuss this “troublemaking” in Comets and Asteroids: Debris of the Solar System.)

Link to Learning

A great variety of infographics at space.com let you explore what it would be like to live on various worlds in the solar system.

This page titled 6.4: Origin of the Solar System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 7.5: Origin of the Solar System by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

6.5: Formation of the Solar System

Learning Objectives

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

- Describe the motion, chemical, and age constraints that must be met by any theory of solar system formation
- Summarize the physical and chemical changes during the solar nebula stage of solar system formation
- Explain the formation process of the terrestrial and giant planets
- Describe the main events of the further evolution of the solar system

As we have seen, the comets, asteroids, and meteorites are surviving remnants from the processes that formed the solar system. The planets, moons, and the Sun, of course, also are the products of the formation process, although the material in them has undergone a wide range of changes. We are now ready to put together the information from all these objects to discuss what is known about the origin of the solar system.

Observational Constraints

There are certain basic properties of the planetary system that any theory of its formation must explain. These may be summarized under three categories: motion constraints, chemical constraints, and age constraints. We call them *constraints* because they place restrictions on our theories; unless a theory can explain the observed facts, it will not survive in the competitive marketplace of ideas that characterizes the endeavor of science. Let's take a look at these constraints one by one.

There are many regularities to the motions in the solar system. We saw that the planets all revolve around the Sun in the same direction and approximately in the plane of the Sun's own rotation. In addition, most of the planets rotate in the same direction as they revolve, and most of the moons also move in counterclockwise orbits (when seen from the north). With the exception of the comets and other trans-neptunian objects, the motions of the system members define a disk or Frisbee shape. Nevertheless, a full theory must also be prepared to deal with the exceptions to these trends, such as the *retrograde rotation* (not revolution) of Venus.

In the realm of chemistry, we saw that Jupiter and Saturn have approximately the same composition—dominated by hydrogen and helium. These are the two largest planets, with sufficient gravity to hold on to any gas present when and where they formed; thus, we might expect them to be representative of the original material out of which the solar system formed. Each of the other members of the planetary system is, to some degree, lacking in the light elements. A careful examination of the composition of solid solar-system objects shows a striking progression from the metal-rich inner planets, through those made predominantly of rocky materials, out to objects with ice-dominated compositions in the outer solar system. The comets in the Oort cloud and the trans-neptunian objects in the Kuiper belt are also icy objects, whereas the asteroids represent a transitional rocky composition with abundant dark, carbon-rich material.

As we saw in Other Worlds: An Introduction to the Solar System, this general chemical pattern can be interpreted as a temperature sequence: hot near the Sun and cooler as we move outward. The inner parts of the system are generally missing those materials that could not condense (form a solid) at the high temperatures found near the Sun. However, there are (again) important exceptions to the general pattern. For example, it is difficult to explain the presence of water on Earth and Mars if these planets formed in a region where the temperature was too hot for ice to condense, unless the ice or water was brought in later from cooler regions. The extreme example is the observation that there are polar deposits of ice on both Mercury and the Moon; these are almost certainly formed and maintained by occasional comet impacts.

As far as age is concerned, we discussed that radioactive dating demonstrates that some rocks on the surface of Earth have been present for at least 3.8 billion years, and that certain lunar samples are 4.4 billion years old. The primitive meteorites all have radioactive ages near 4.5 billion years. The age of these unaltered building blocks is considered the age of the planetary system. The similarity of the measured ages tells us that planets formed and their crusts cooled within a few tens of millions of years (at most) of the beginning of the solar system. Further, detailed examination of primitive meteorites indicates that they are made primarily from material that condensed or coagulated out of a hot gas; few identifiable fragments appear to have survived from before this hot-vapor stage 4.5 billion years ago.

The Solar Nebula

All the foregoing constraints are consistent with the general idea, introduced in Other Worlds: An Introduction to the Solar System, that the solar system formed 4.5 billion years ago out of a rotating cloud of vapor and dust—which we call the solar nebula—with an initial composition similar to that of the Sun today. As the solar nebula collapsed under its own gravity, material fell toward the center, where things became more and more concentrated and hot. Increasing temperatures in the shrinking nebula vaporized most of the solid material that was originally present.

At the same time, the collapsing nebula began to rotate faster through the conservation of angular momentum (see the Orbits and Gravity and Earth, Moon, and Sky chapters). Like a figure skater pulling her arms in to spin faster, the shrinking cloud spun more quickly as time went on. Now, think about how a round object spins. Close to the poles, the spin rate is slow, and it gets faster as you get closer to the equator. In the same way, near the poles of the nebula, where orbits were slow, the nebular material fell directly into the center. Faster moving material, on the other hand, collapsed into a flat disk revolving around the central object (Figure 14.11). The existence of this disk-shaped rotating nebula explains the primary motions in the solar system that we discussed in the previous section. And since they formed from a rotating disk, the planets all orbit the same way.

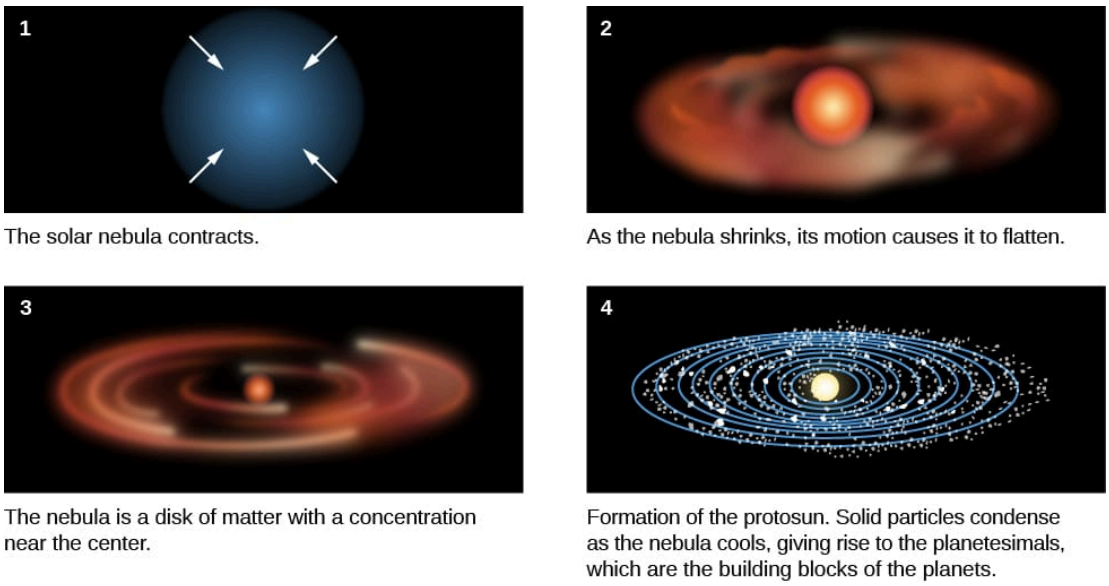


Figure 14.11 : Steps in Forming the Solar System. This illustration shows the steps in the formation of the solar system from the solar nebula. As the nebula shrinks, its rotation causes it to flatten into a disk. Much of the material is concentrated in the hot center, which will ultimately become a star. Away from the center, solid particles can condense as the nebula cools, giving rise to planetesimals, the building blocks of the planets and moons.

Picture the solar nebula at the end of the collapse phase, when it was at its hottest. With no more gravitational energy (from material falling in) to heat it, most of the nebula began to cool. The material in the center, however, where it was hottest and most crowded, formed a *star* that maintained high temperatures in its immediate neighborhood by producing its own energy. Turbulent motions and magnetic fields within the disk can drain away angular momentum, robbing the disk material of some of its spin. This allowed some material to continue to fall into the growing star, while the rest of the disk gradually stabilized.

The temperature within the disk decreased with increasing distance from the Sun, much as the planets' temperatures vary with position today. As the disk cooled, the gases interacted chemically to produce compounds; eventually these compounds condensed into liquid droplets or solid grains. This is similar to the process by which raindrops on Earth condense from moist air as it rises over a mountain.

Let's look in more detail at how material condensed at different places in the maturing disk (Figure 14.12). The first materials to form solid grains were the metals and various rock-forming silicates. As the temperature dropped, these were joined throughout much of the solar nebula by sulfur compounds and by carbon- and water-rich silicates, such as those now found abundantly among the asteroids. However, in the inner parts of the disk, the temperature never dropped low enough for such materials as ice or carbonaceous organic compounds to condense, so they were lacking on the innermost planets.

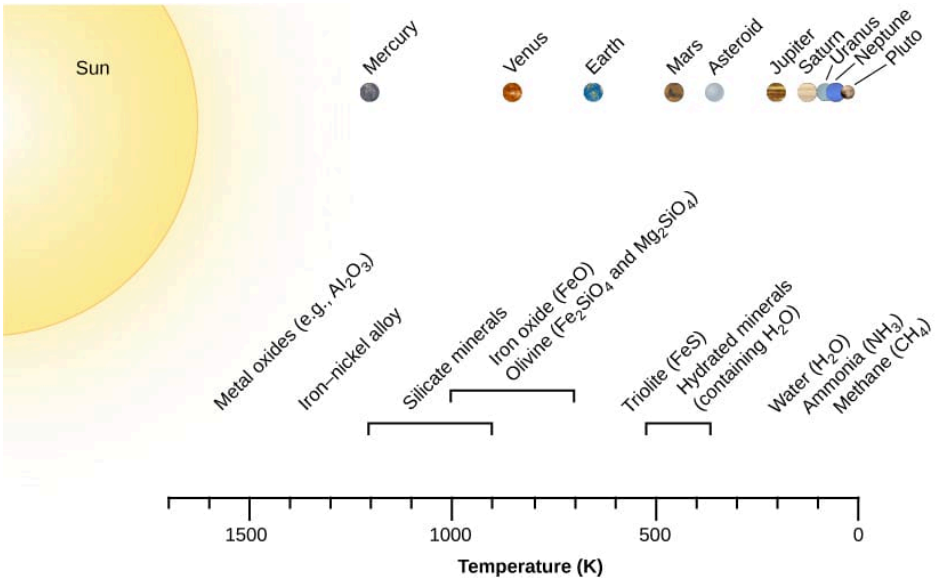


Figure 14.12 : Chemical Condensation Sequence in the Solar Nebula. The scale along the bottom shows temperature; above are the materials that would condense out at each temperature under the conditions expected to prevail in the nebula.

Far from the Sun, cooler temperatures allowed the oxygen to combine with hydrogen and condense in the form of water (H_2O) ice. Beyond the orbit of Saturn, carbon and nitrogen combined with hydrogen to make ices such as methane (CH_4) and ammonia (NH_3). This sequence of events explains the basic chemical composition differences among various regions of the solar system.

Example 14.1: Rotation of the Solar Nebula

We can use the concept of angular momentum to trace the evolution of the collapsing solar nebula. The angular momentum of an object is proportional to the square of its size (diameter) divided by its period of rotation Math input error

Solution

We are given that the final diameter of the solar nebula is about 80 AU. Noting the initial state before the collapse and the final state at Pluto's orbit, then

$$\frac{P_{\text{final}}}{P_{\text{initial}}} = \left(\frac{D_{\text{final}}}{D_{\text{initial}}}\right)^2 = \left(\frac{80}{10{,}000}\right)^2 = (0.008)^2 = 0.000064$$

With P_{initial} equal to 1,000,000 years, P_{final}, the new rotation period, is 64 years. This is a lot shorter than the actual time Pluto takes to go around the Sun, but it gives you a sense of the kind of speeding up the conservation of angular momentum can produce. As we noted earlier, other mechanisms helped the material in the disk lose angular momentum before the planets fully formed.

Exercise 6.5.1

What would the rotation period of the nebula in our example be when it had shrunk to the size of Jupiter's orbit?

Answer available at the end of this section.

Formation of the Terrestrial Planets

The grains that condensed in the solar nebula rather quickly joined into larger and larger chunks, until most of the solid material was in the form of *planetesimals,* chunks a few kilometers to a few tens of kilometers in diameter. Some planetesimals still survive today as comets and asteroids. Others have left their imprint on the cratered surfaces of many of the worlds we studied in earlier

chapters. A substantial step up in size is required, however, to go from planetesimal to planet.

Some planetesimals were large enough to attract their neighbors gravitationally and thus to grow by the process called accretion. While the intermediate steps are not well understood, ultimately several dozen centers of accretion seem to have grown in the inner solar system. Each of these attracted surrounding planetesimals until it had acquired a mass similar to that of Mercury or Mars. At this stage, we may think of these objects as *protoplanets*—"not quite ready for prime time" planets.

Each of these protoplanets continued to grow by the accretion of planetesimals. Every incoming planetesimal was accelerated by the gravity of the protoplanet, striking with enough energy to melt both the projectile and a part of the impact area. Soon the entire protoplanet was heated to above the melting temperature of rocks. The result was *planetary differentiation*, with heavier metals sinking toward the core and lighter silicates rising toward the surface. As they were heated, the inner protoplanets lost some of their more volatile constituents (the lighter gases), leaving more of the heavier elements and compounds behind.

Formation of the Giant Planets

In the outer solar system, where the available raw materials included ices as well as rocks, the protoplanets grew to be much larger, with masses ten times greater than Earth. These protoplanets of the outer solar system were so large that they were able to attract and hold the surrounding gas. As the hydrogen and helium rapidly collapsed onto their cores, the giant planets were heated by the energy of contraction. But although these giant planets got hotter than their terrestrial siblings, they were far too small to raise their central temperatures and pressures to the point where nuclear reactions could begin (and it is such reactions that give us our definition of a star). After glowing dull red for a few thousand years, the giant planets gradually cooled to their present state (Figure 14.13).

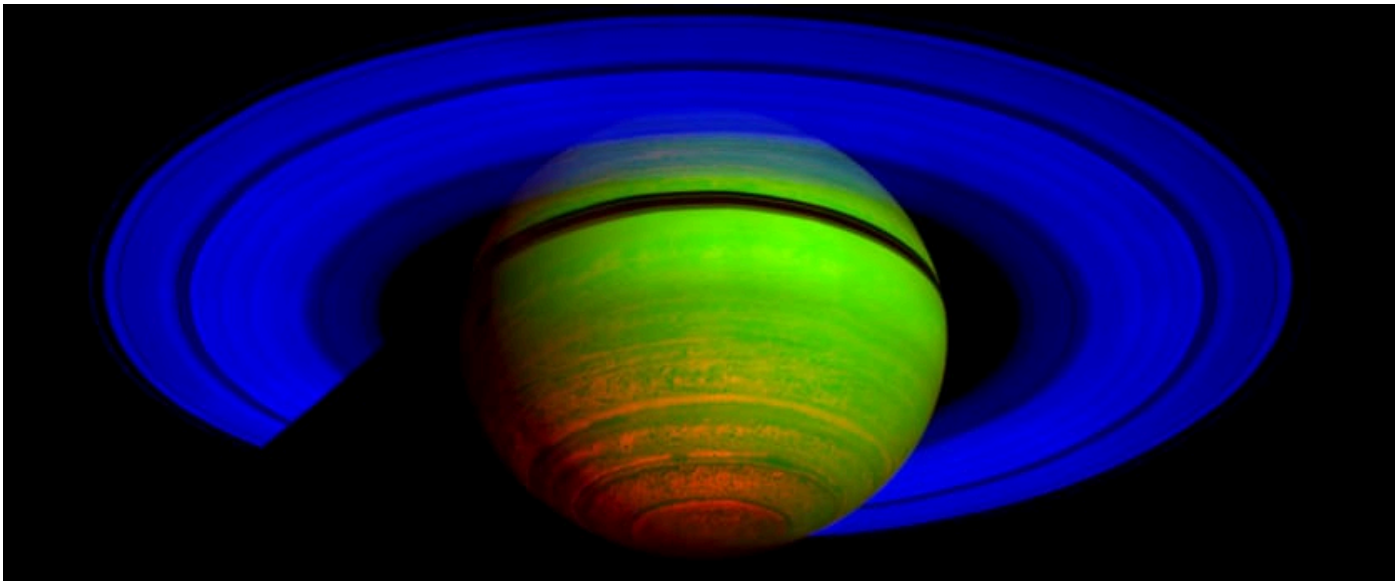


Figure 14.13 : Saturn Seen in Infrared. This image from the Cassini spacecraft is stitched together from 65 individual observations. Sunlight reflected at a wavelength of 2 micrometers is shown as blue, sunlight reflected at 3 micrometers is shown as green, and heat radiated from Saturn's interior at 5 micrometers is red. For example, Saturn's rings reflect sunlight at 2 micrometers, but not at 3 and 5 micrometers, so they appear blue. Saturn's south polar regions are seen glowing with internal heat. (credit: modification of work by NASA/JPL/University of Arizona)

The collapse of gas from the nebula onto the cores of the giant planets explains how these objects acquired nearly the same hydrogen-rich composition as the Sun. The process was most efficient for Jupiter and Saturn; hence, their compositions are most nearly "cosmic." Much less gas was captured by Uranus and Neptune, which is why these two planets have compositions dominated by the icy and rocky building blocks that made up their large cores rather than by hydrogen and helium. The initial formation period ended when much of the available raw material was used up and the solar wind (the flow of atomic particles) from the young Sun blew away the remaining supply of lighter gases.

Further Evolution of the System

All the processes we have just described, from the collapse of the solar nebula to the formation of protoplanets, took place within

a few million years. However, the story of the formation of the solar system was not complete at this stage; there were many planetesimals and other debris that did not initially accumulate to form the planets. What was their fate?

The comets visible to us today are merely the tip of the cosmic iceberg (if you'll pardon the pun). Most comets are believed to be in the Oort cloud, far from the region of the planets. Additional comets and icy dwarf planets are in the Kuiper belt, which stretches beyond the orbit of Neptune. These icy pieces probably formed near the present orbits of Uranus and Neptune but were ejected from their initial orbits by the gravitational influence of the giant planets.

In the inner parts of the system, remnant planetesimals and perhaps several dozen protoplanets continued to whiz about. Over the vast span of time we are discussing, collisions among these objects were inevitable. Giant impacts at this stage may have stripped Mercury of part of its mantle and crust, reversed the rotation of Venus, and broke off part of Earth to create the Moon (all events we discussed in other chapters).

Smaller-scale impacts also added mass to the inner protoplanets. Because the gravity of the giant planets could "stir up" the orbits of the planetesimals, the material impacting on the inner protoplanets could have come from almost anywhere within the solar system. In contrast to the previous stage of accretion, therefore, this new material did not represent just a narrow range of compositions.

As a result, much of the debris striking the inner planets was ice-rich material that had condensed in the outer part of the solar nebula. As this comet-like bombardment progressed, Earth accumulated the water and various organic compounds that would later be critical to the formation of life. Mars and Venus probably also acquired abundant water and organic materials from the same source, as Mercury and the Moon are still doing to form their icy polar caps.

Gradually, as the planets swept up or ejected the remaining debris, most of the planetesimals disappeared. In two regions, however, stable orbits are possible where leftover planetesimals could avoid impacting the planets or being ejected from the system. These regions are the asteroid belt between Mars and Jupiter and the Kuiper belt beyond Neptune. The planetesimals (and their fragments) that survive in these special locations are what we now call asteroids, comets, and trans-neptunian objects.

Astronomers used to think that the solar system that emerged from this early evolution was similar to what we see today. Detailed recent studies of the orbits of the planets and asteroids, however, suggest that there were more violent events soon afterward, perhaps involving substantial changes in the orbits of Jupiter and Saturn. These two giant planets control, through their gravity, the distribution of asteroids. Working backward from our present solar system, it appears that orbital changes took place during the first few hundred million years. One consequence may have been scattering of asteroids into the inner solar system, causing the period of "heavy bombardment" recorded in the oldest lunar craters.

Answers

Exercise $6.5.1$

The period of the rotating nebula is inversely proportional to D^2. As we have just seen, $\frac{P_{\text{final}}}{P_{\text{initial}}} = \left(\frac{D_{\text{final}}}{D_{\text{initial}}}\right)^2$. Initially, we have $P_{\text{initial}} = 10^6$ yr and $D_{\text{initial}} = 10^4$ AU. Then, if D_{final} is in AU, P_{final} (in years) is given by $P_{\text{final}} = 0.01 D_{\text{final}}^2$. If Jupiter's orbit has a radius of 5.2 AU, then the diameter is 10.4 AU. The period is then 1.08 years.

This page titled 6.5: Formation of the Solar System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 14.4: Formation of the Solar System by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

7: WORLDS AROUND OTHER STARS

7.1: Evidence That Planets Form around Other Stars
7.2: Planets beyond the Solar System- Search and Discovery
7.3: Exoplanets Everywhere - What we are Learning
7.4: New Perspectives on Planet Formation

This page titled 7: Worlds Around Other Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.1: Evidence That Planets Form around Other Stars

Learning Objectives

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

- Trace the evolution of dust surrounding a protostar, leading to the development of rocky planets and gas giants
- Estimate the timescale for growth of planets using observations of the disks surrounding young stars
- Evaluate evidence for planets around forming stars based on the structures seen in images of the circumstellar dust disks

Having developed on a planet and finding it essential to our existence, we have a special interest in how planets fit into the story of star formation. Yet planets outside the solar system are extremely difficult to detect. Recall that we see planets in our own system only because they reflect sunlight and are close by. When we look to the other stars, we find that the amount of light a planet reflects is a depressingly tiny fraction of the light its star gives off. Furthermore, from a distance, planets are lost in the glare of their much-brighter parent stars.

Disks around Protostars: Planetary Systems in Formation

It is a lot easier to detect the spread-out raw material from which planets might be assembled than to detect planets after they are fully formed. From our study of the solar system, we understand that planets form by the gathering together of gas and dust particles in orbit around a newly created star. Each dust particle is heated by the young protostar and radiates in the infrared region of the spectrum. Before any planets form, we can detect such radiation from all of the spread-out individual dust particles that are destined to become parts of planets. We can also detect the silhouette of the disk if it blocks bright light coming from a source behind it (Figure 21.13).

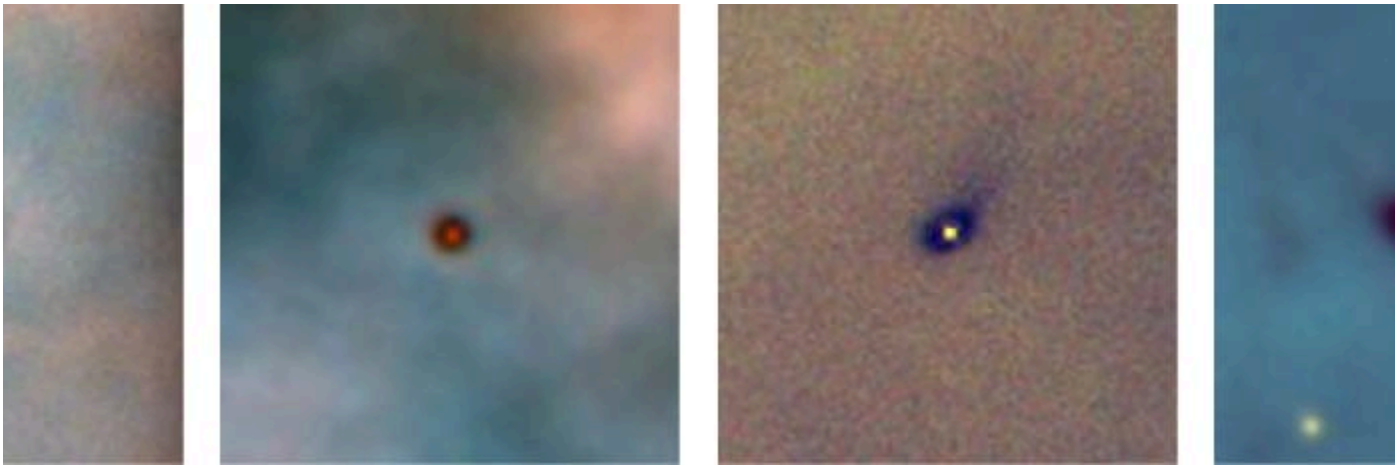


Figure 21.13 : Disks around Protostars. These Hubble Space Telescope images show four disks around young stars in the Orion Nebula. The dark, dusty disks are seen silhouetted against the bright backdrop of the glowing gas in the nebula. The size of each image is about 30 times the diameter of our planetary system; this means the disks we see here range in size from two to eight times the orbit of Pluto. The red glow at the center of each disk is a young star, no more than a million years old. These images correspond to the stage in the life of a protostar shown in part (d) of Figure 21.8. (credit: modification of work by Mark McCaughrean (Max-Planck-Institute for Astronomy), C. Robert O'Dell (Rice University), and NASA)

Once the dust particles gather together and form a few planets (and maybe some moons), the overwhelming majority of the dust is hidden in the interiors of the planets where we cannot see it. All we can now detect is the radiation from the outside surfaces, which cover a drastically smaller area than the huge, dusty disk from which they formed. The amount of infrared radiation is therefore greatest before the dust particles combine into planets. For this reason, our search for planets begins with a search for infrared radiation from the material required to make them.

A disk of gas and dust appears to be an essential part of star formation. Observations show that nearly all very young protostars have disks and that the disks range in size from 10 to 1000 AU. (For comparison, the average diameter of the orbit of Pluto, which can be considered the rough size of our own planetary system, is 80 AU, whereas the outer diameter of the Kuiper belt of smaller icy bodies is about 100 AU.) The mass contained in these disks is typically 1–10% of the mass of our own Sun, which is more than the mass of all the planets in our solar system put together. Such observations already demonstrate that a large fraction of stars begin their lives with enough material in the right place to form a planetary system.

The Timing of Planet Formation and Growth

We can use observations of how the disks change with time to estimate how long it takes for planets to form. If we measure the temperature and luminosity of a protostar, then, as we saw, we can place it in an H–R diagram like the one shown in Figure 21.12. By comparing the real star with our models of how protostars should evolve with time, we can estimate its age. We can then look at how the disks we observe change with the ages of the stars that they surround.

What such observations show is that if a protostar is less than about 1 to 3 million years old, its disk extends all the way from very close to the surface of the star out to tens or hundreds of AU away. In older protostars, we find disks with outer parts that still contain large amounts of dust, but the inner regions have lost most of their dust. In these objects, the disk looks like a donut, with the protostar centered in its hole. The inner, dense parts of most disks have disappeared by the time the stars are 10 million years old (Figure 21.14).

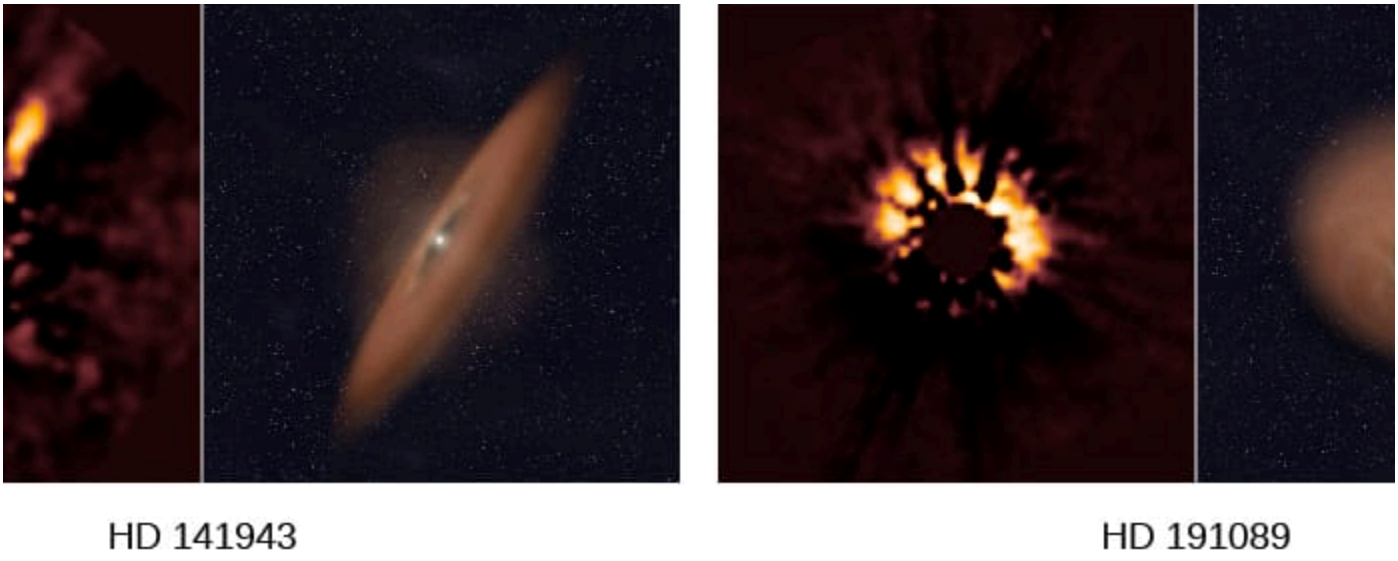


Figure 21.14 : Protoplanetary Disks around Two Stars. The left view of each star shows infrared observations by the Hubble Space Telescope of their protoplanetary disks. The central star is much brighter than the surrounding disk, so the instrument includes a coronagraph, which has a small shield that blocks the light of the central star but allows the surrounding disk to be imaged. The right image of each star shows models of the disks based on the observations. The star HD 141943 has an age of about 17 million years, while HD 191089 is about 12 million years old. (credit: modification of work by NASA, ESA, R. Soummer and M. Perrin (STScI), L. Pueyo (STScI/Johns Hopkins University), C. Chen and D. Golimowski (STScI), J.B. Hagan (STScI/Purdue University), T. Mittal (University of California, Berkeley/Johns Hopkins University), E. Choquet, M. Moerchen, and M. N'Diaye (STScI), A. Rajan (Arizona State University), S. Wolff (STScI/Purdue University), J. Debes and D. Hines (STScI), and G. Schneider (Steward Observatory/University of Arizona))

Calculations show that the formation of one or more planets could produce such a donut-like distribution of dust. Suppose a planet forms a few AU away from the protostar, presumably due to the gathering together of matter from the disk. As the planet grows in mass, the process clears out a dust-free region in its immediate neighborhood. Calculations also show that any small dust particles and gas that were initially located in the region between the protostar and the planet, and that are not swept up by the planet, will then fall onto the star very quickly in about 50,000 years.

Matter lying outside the planet's orbit, in contrast, is prevented from moving into the hole by the gravitational forces exerted by the planet. (We saw something similar in Saturn's rings, where the action of the shepherd moons keeps the material near the edge of the rings from spreading out.) If the formation of a planet is indeed what produces and sustains holes in the disks that surround very young stars, then planets must form in 3 to 30 million years. This is a short period compared with the lifetimes of most stars and shows that the formation of planets may be a quick byproduct of the birth of stars.

Calculations show that *accretion* can drive the rapid growth of planets—small, dust-grain-size particles orbiting in the disk collide and stick together, with the larger collections growing more rapidly as they attract and capture smaller ones. Once these clumps grow to about 10 centimeters in size or so, they enter a perilous stage in their development. At that size, unless they can grow to larger than about 100 meters in diameter, they are subject to drag forces produced by friction with the gas in the disk—and their orbits can rapidly decay, plunging them into the host star. Therefore, these bodies must rapidly grow to nearly 1 kilometer in size in diameter to avoid a fiery fate. At this stage, they are considered planetesimals (the small chunks of solid matter—ice and dust particles—that you learned about in Other Worlds: An Introduction to the Solar System). Once they survive to those sizes, the largest survivors will continue to grow by accreting smaller planetesimals; ultimately, this process results in a few large planets.

If the growing planets reach a mass bigger than about 10 times the mass of Earth, their gravity is strong enough to capture and hold on to hydrogen gas that remains in the disk. At that point, they will grow in mass and radius rapidly, reaching giant planet dimensions. However, to do so requires that the rapidly evolving central star hasn't yet driven away the gas in the disk with its increasingly vigorous wind (see the earlier section on Star Formation). From observations, we see that the disk can be blown away within 10 million years, so growth of a giant planet must also be a very fast process, astronomically speaking.

Debris Disks and Shepherd Planets

The dust around newly formed stars is gradually either incorporated into the growing planets in the newly forming planetary system or ejected through gravitational interactions with the planets into space. The dust will disappear after about 30 million years unless the disk is continually supplied with new material. Local comets and asteroids are the most likely sources of new dust. As the planet-size bodies grow, they stir up the orbits of smaller objects in the area. These small bodies collide at high speeds, shatter, and produce tiny particles of silicate dust and ices that can keep the disk supplied with the debris from these collisions.

Over several hundred million years, the comets and asteroids will gradually be reduced in number, the frequency of collisions will go down, and the supply of fresh dust will diminish. Remember that the heavy bombardment in the early solar system ended when the Sun was only about 500 million years old. Observations show that the dusty "debris disks" around stars also become largely undetectable by the time the stars reach an age of 400 to 500 million years. It is likely, however, that some small amount of cometary material will remain in orbit, much like our Kuiper belt, a flattened disk of comets outside the orbit of Neptune.

In a young planetary system, even if we cannot see the planets directly, the planets can concentrate the dust particles into clumps and arcs that are much larger than the planets themselves and more easily imaged. This is similar to how the tiny moons of Saturn shepherd the particles in the rings and produce large arcs and structures in Saturn's rings.

Debris disks—many with just such clumps and arcs—have now been found around many stars, such as HL Tau, located about 450 light-years from Earth in the constellation Taurus (Figure 21.15). In some stars, the brightness of the rings varies with position; around other stars, there are bright arcs and gaps in the rings. The brightness indicates the relative concentration of dust, since what we are seeing is infrared (heat radiation) from the dust particles in the rings. More dust means more radiation.

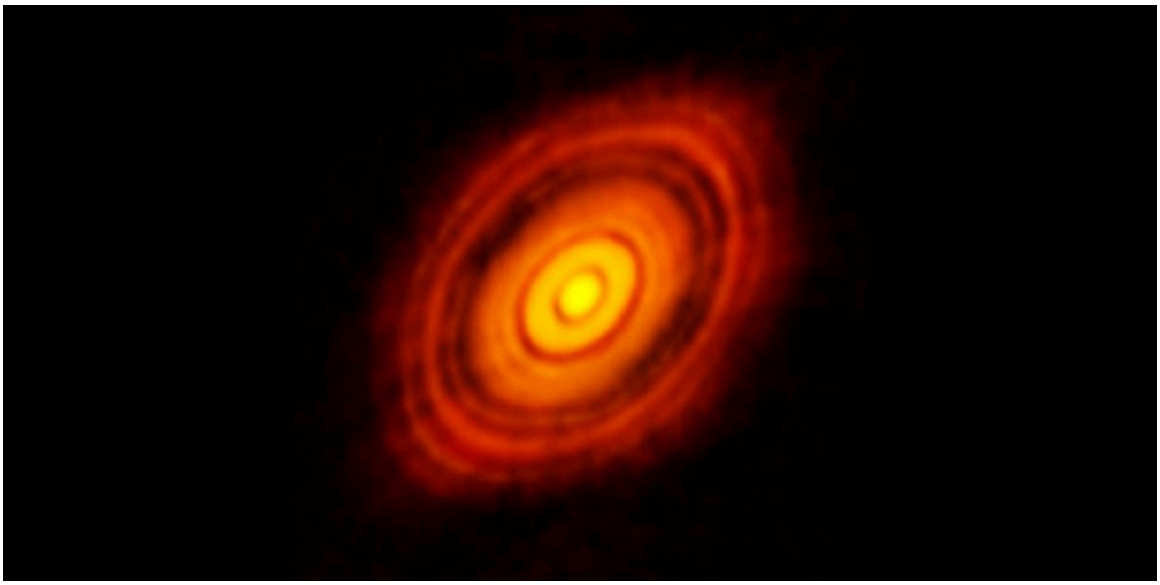


Figure 21.15 : Dust Ring around a Young Star. This image was made by ALMA (the Atacama Large Millimeter/Submillimeter Array) at a wavelength of 1.3 millimeters and shows the young star HL Tau and its protoplanetary disk. It reveals multiple rings and gaps that indicate the presence of emerging planets, which are sweeping their orbits clear of dust and gas. (credit: modification of work by ALMA (ESO/NAOJ/NRAO))

Link to Learning

Watch a short video clip of the director of NRAO (National Radio Astronomy Observatory) describing the high-resolution observations of the young star HL Tau. While you're there, watch an artist's animation of a protoplanetary disk to see newly formed planets traveling around a host (parent) star.

This page titled 7.1: Evidence That Planets Form around Other Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 21.4: Evidence That Planets Form around Other Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

7.2: Planets beyond the Solar System- Search and Discovery

Learning Objectives

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

- Describe the orbital motion of planets in our solar system using Kepler's laws
- Compare the indirect and direct observational techniques for exoplanet detection

For centuries, astronomers have dreamed of finding planets around other stars, including other planets like Earth. Direct observations of such distant planets are very difficult, however. You might compare a planet orbiting a star to a mosquito flying around one of those giant spotlights at a shopping center opening. From close up, you might spot the mosquito. But imagine viewing the scene from some distance away—say, from an airplane. You could see the spotlight just fine, but what are your chances of catching the mosquito in that light? Instead of making direct images, astronomers have relied on indirect observations and have now succeeded in detecting a multitude of planets around other stars.

In 1995, after decades of effort, we found the first such exoplanet (a planet outside our solar system) orbiting a main-sequence star, and today we know that most stars form with planets. This is an example of how persistence and new methods of observation advance the knowledge of humanity. By studying exoplanets, astronomers hope to better understand our solar system in context of the rest of the universe. For instance, how does the arrangement of our solar system compare to planetary systems in the rest of the universe? What do exoplanets tell us about the process of planet formation? And how does knowing the frequency of exoplanets influence our estimates of whether there is life elsewhere?

Searching for Orbital Motion

Most exoplanet detections are made using techniques where we observe the *effect* that the planet exerts on the host star. For example, the gravitational tug of an unseen planet will cause a small wobble in the host star. Or, if its orbit is properly aligned, a planet will periodically cross in front of the star, causing the brightness of the star to dim.

To understand how a planet can move its host star, consider a single Jupiter-like planet. Both the planet and the star actually revolve about their *common center of mass*. Remember that gravity is a mutual attraction. The star and the planet each exert a force on the other, and we can find a stable point, the center of mass, between them about which both objects move. The smaller the mass of a body in such a system, the larger its orbit. A massive star barely swings around the center of mass, while a low-mass planet makes a much larger "tour."

Suppose the planet is like Jupiter and has a mass about one-thousandth that of its star; in this case, the size of the star's orbit is one-thousandth the size of the planet's. To get a sense of how difficult observing such motion might be, let's see how hard Jupiter would be to detect in this way from the distance of a nearby star. Consider an alien astronomer trying to observe our own system from Alpha Centauri, the closest star system to our own (about 4.3 light-years away). There are two ways this astronomer could try to detect the orbital motion of the Sun. One way would be to look for changes in the Sun's position on the sky. The second would be to use the Doppler effect to look for changes in its velocity. Let's discuss each of these in turn.

The diameter of Jupiter's apparent orbit viewed from Alpha Centauri is 1$0$ seconds of arc, and that of the Sun's orbit is $0.0$1$0$ seconds of arc. (Remember, 1 second of arc is 1/3$6$0$0$ degree.) If they could measure the apparent position of the Sun (which is bright and easy to detect) to sufficient precision, they would describe an orbit of diameter $0.0$1$0$ seconds of arc with a period equal to that of Jupiter, which is 12 years.

In other words, if they watched the Sun for 12 years, they would see it wiggle back and forth in the sky by this minuscule fraction of a degree. From the observed motion and the period of the "wiggle," they could deduce the mass of Jupiter and its distance using Kepler's laws. (To refresh your memory about these laws, see the chapter on Orbits and Gravity.)

Measuring positions in the sky this accurately is extremely difficult, and so far, astronomers have not made any confirmed detections of planets using this technique. However, we have been successful in using spectrometers to measure the changing velocity of stars with planets around them.

As the star and planet orbit each other, part of their motion will be in our line of sight (toward us or away from us). Such motion can be measured using the *Doppler effect* and the star's spectrum. As the star moves back and forth in orbit around the system's center of mass in response to the gravitational tug of an orbiting planet, the lines in its spectrum will shift back and forth.

Let's again consider the example of the Sun. Its *radial velocity* (motion toward or away from us) changes by about 13 meters per second with a period of 12 years because of the gravitational pull of Jupiter. This corresponds to about 30 miles per hour, roughly the speed at which many of us drive around town. Detecting motion at this level in a star's spectrum presents an enormous technical challenge, but several groups of astronomers around the world, using specialized spectrographs designed for this purpose, have succeeded. Note that the change in speed does not depend on the distance of the star from the observer. Using the Doppler effect to detect planets will work at any distance, as long as the star is bright enough to provide a good spectrum and a large telescope is available to make the observations (Figure 21.16).

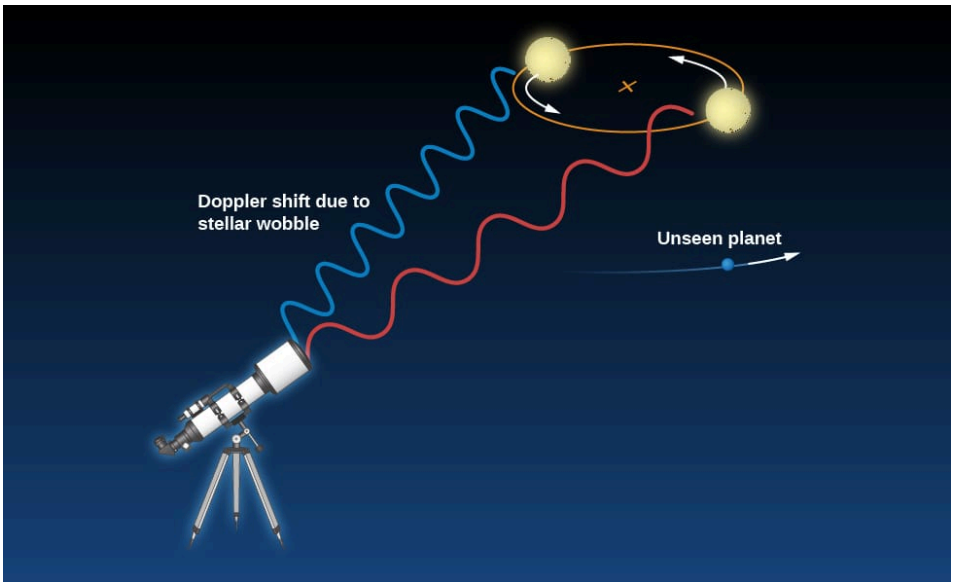


Figure 21.16 : Doppler Method of Detecting Planets. The motion of a star around a common center of mass with an orbiting planet can be detected by measuring the changing speed of the star. When the star is moving away from us, the lines in its spectrum show a tiny redshift; when it is moving toward us, they show a tiny blueshift. The change in color (wavelength) has been exaggerated here for illustrative purposes. In reality, the Doppler shifts we measure are extremely small and require sophisticated equipment to be detected.

The first successful use of the Doppler effect to find a planet around another star was in 1995. Michel Mayor and Didier Queloz of the Geneva Observatory (Figure 21.17) used this technique to find a planet orbiting a star resembling our Sun called 51 Pegasi, about 40 light-years away. (The star can be found in the sky near the great square of Pegasus, the flying horse of Greek mythology, one of the easiest-to-find star patterns.) To everyone's surprise, the planet takes a mere 4.2 days to orbit around the star. (Remember that Mercury, the innermost planet in our solar system, takes 88 days to go once around the Sun, so 4.2 days seems fantastically short.)

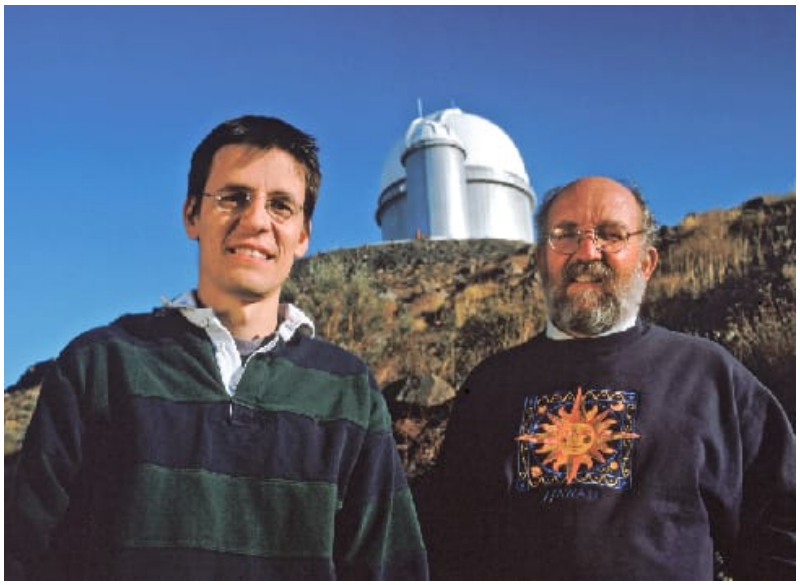


Figure 21.17 : Planet Discoverers. In 1995, Didier Queloz and Michel Mayor of the Geneva Observatory were the first to discover a planet around a regular star (51 Pegasi), for which they received the 2019 Nobel Prize in physics. They are seen here at an observatory in Chile where they are continuing their planet hunting. (credit: Weinstein/Ciel et Espace Photos)

Mayor and Queloz's findings mean the planet must be very close to 51 Pegasi, circling it about 7 million kilometers away (Figure 21.18). At that distance, the energy of the star should heat the planet's surface to a temperature of a few thousand degrees Celsius (a bit hot for future tourism). From its motion, astronomers calculate that it has at least half the mass of Jupiter[1], making it clearly a jovian and not a terrestrial-type planet.

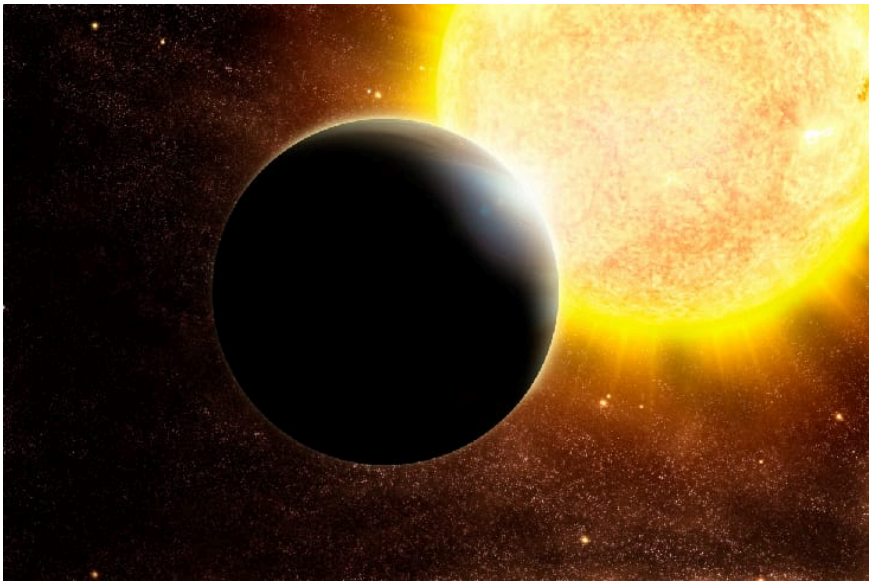


Figure 21.18 : Hot Jupiter. Artist Greg Bacon painted this impression of a hot, Jupiter-type planet orbiting close to a sunlike star. The artist shows bands on the planet like Jupiter, but we only estimate the mass of most hot, Jupiter-type planets from the Doppler method and don't know what conditions on the planet are like. (credit: ESO)

Since that initial planet discovery, the rate of progress has been breathtaking. Almost a thousand giant planets have been discovered using the Doppler technique. Many of these giant planets are orbiting close to their stars—astronomers have called these *hot Jupiters*.

The existence of giant planets so close to their stars was a surprise, and these discoveries have forced us to rethink our ideas about how planetary systems form. But for now, bear in mind that the Doppler-shift method—which relies on the pull of a planet making its star "wiggle" back and forth around the center of mass—is most effective at finding planets that are both close to their stars and massive. These planets cause the biggest "wiggles" in the motion of their stars and the biggest Doppler shifts in the

spectrum. Plus, they will be found sooner, since astronomers like to monitor the star for at least one full orbit (and perhaps more) and hot Jupiters take the shortest time to complete their orbit.

So if such planets exist, we would expect to be finding this type first. Scientists call this a *selection effect*—where our technique of discovery selects certain kinds of objects as "easy finds." As an example of a selection effect in everyday life, imagine you decide you are ready for a new romantic relationship in your life. To begin with, you only attend social events on campus, all of which require a student ID to get in. Your selection of possible partners will then be limited to students at your college. That may not give you as diverse a group to choose from as you want. In the same way, when we first used the Doppler technique, it selected massive planets close to their stars as the most likely discoveries. As we spend longer times watching target stars and as our ability to measure smaller Doppler shifts improves, this technique can reveal more distant and less massive planets too.

Link to Learning

View a series of animations demonstrating solar system motion and Kepler's laws, and select animation 1 (Kepler's laws) from the dropdown playlist. To view an animation demonstrating the radial velocity curve for an exoplanet, select animation 29 (radial velocity curve for an exoplanet) and animation 30 (radial velocity curve for an exoplanet—elliptical orbit) from the dropdown playlist.

Transiting Planets

The second method for indirect detection of exoplanets is based not on the motion of the star but on its brightness. When the orbital plane of the planet is tilted or inclined so that it is viewed edge-on, we will see the planet cross in front of the star once per orbit, causing the star to dim slightly; this event is known as transit. Figure 21.19 shows a sketch of the transit at three time steps: (1) out of transit, (2) the start of transit, and (3) full transit, along with a sketch of the light curve, which shows the drop in the brightness of the host star. The amount of light blocked—the depth of the transit—depends on the area of the planet (its size) compared to the star. If we can determine the size of the star, the transit method tells us the size of the planet.

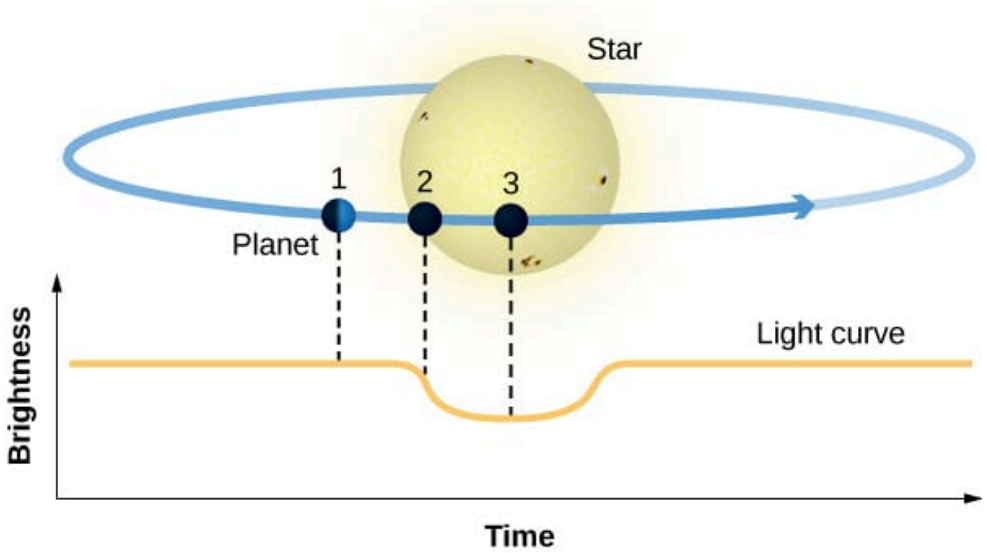


Figure 21.19 : Planet Transits. As the planet transits, it blocks out some of the light from the star, causing a temporary dimming in the brightness of the star. The top figure shows three moments during the transit event and the bottom panel shows the corresponding light curve: (1) out of transit, (2) transit ingress, and (3) the full drop in brightness.

The interval between successive transits is the length of the year for that planet, which can be used (again using Kepler's laws) to find its distance from the star. Larger planets like Jupiter block out more starlight than small earthlike planets, making transits by giant planets easier to detect, even from ground-based observatories. But by going into space, above the distorting effects of Earth's atmosphere, the transit technique has been extended to exoplanets as small as Mars.

Example 21.1: Transit Depth

In a transit, the planet's circular disk blocks the light of the star's circular disk. The area of a circle is πR^2. The amount of light the planet blocks, called the transit depth, is then given by

$$\frac{\pi R^2_{\text{planet}}}{\pi R^2_{\text{star}}} = \frac{R^2_{\text{planet}}}{R^2_{\text{star}}} = \left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2$$

Now calculate the transit depth for a star the size of the Sun with a gas giant planet the size of Jupiter.

Solution

The radius of Jupiter is 71,400 km, while the radius of the Sun is 695,700 km. Substituting into the equation, we get $\left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2 = \left(\frac{71{,}400 \text{ km}}{695{,}700 \text{ km}}\right)^2 = 0.01$ or 1%, which can easily be detected with the instruments on board the Kepler spacecraft.

Exercise 7.2.1

What is the transit depth for a star half the size of the Sun with a much smaller planet, like the size of Earth?

Answer available at the end of this section.

The Doppler method allows us to estimate the mass of a planet. If the same object can be studied by both the Doppler and transit techniques, we can measure both the mass and the size of the exoplanet. This is a powerful combination that can be used to derive the average density (mass/volume) of the planet. In 1999, using measurements from ground-based telescopes, the first transiting planet was detected orbiting the star HD 209458. The planet transits its parent star for about 3 hours every 3.5 days as we view it from Earth. Doppler measurements showed that the planet around HD 209458 has about 70% the mass of Jupiter, but its radius is about 35% larger than Jupiter's. This was the first case where we could determine what an exoplanet was made of—with that mass and radius, HD 209458 must be a gas and liquid world like Jupiter or Saturn.

It is even possible to learn something about the planet's atmosphere. When the planet passes in front of HD 209458, the atoms in the planet's atmosphere absorb starlight. Observations of this absorption were first made at the wavelengths of yellow sodium lines and showed that the atmosphere of the planet contains sodium; now, other elements can be measured as well.

Link to Learning

Try a transit simulator that demonstrates how a planet passing in front of its parent star can lead to the planet's detection. Follow the instructions to run the animation on your computer.

Transiting planets reveal such a wealth of information that the French Space Agency (CNES) and the European Space Agency (ESA) launched the CoRoT space telescope in 2007 to detect transiting exoplanets. CoRoT discovered 32 transiting exoplanets, including the first transiting planet with a size and density similar to Earth. In 2012, the spacecraft suffered an onboard computer failure, ending the mission. Meanwhile, NASA built a much more powerful transit observatory called Kepler.

In 2009, NASA launched the Kepler space telescope, dedicated to the discovery of transiting exoplanets. This spacecraft stared continuously at more than 150,000 stars in a small patch of sky near the constellation of Cygnus—just above the plane of our Milky Way Galaxy (Figure 21.20). Kepler's cameras and ability to measure small changes in brightness very precisely enabled the discovery of thousands of exoplanets, including many multi-planet systems. The spacecraft required three reaction wheels—a type of wheel used to help control slight rotation of the spacecraft—to stabilize the pointing of the telescope and monitor the brightness of the same group of stars over and over again. Kepler was launched with four reaction wheels (one a spare), but by May 2013, two wheels had failed and the telescope could no longer be accurately pointed toward the target area. Kepler had been designed to operate for 4 years, and ironically, the pointing failure occurred exactly 4 years and 1 day after it began observing.

However, this failure did not end the mission. The Kepler telescope continued to observe for two more years, looking for short-period transits in different parts of the sky. We discuss the Kepler results in the next section Exoplanets Everywhere: What We

Are Learning. A new NASA mission called TESS (Transiting Exoplanet Survey Satellite) is now carrying out a transit survey of the nearer (and therefore brighter) stars all over the sky. TESS completed its first planned survey of the sky in August 2020 and reported 66 exoplanet discoveries and almost 2100 candidates that need further observations to confirm.

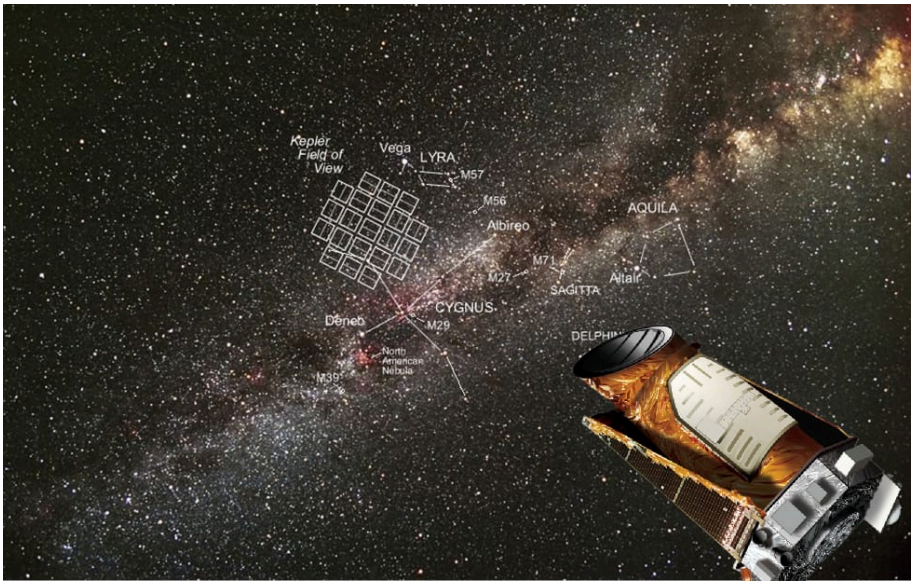


Figure 21.20 : Kepler's Field of View. The boxes show the region where the Kepler spacecraft cameras took images of over 150,000 stars regularly, to find transiting planets. (credit "field of view": modification of work by NASA/Kepler mission; credit "spacecraft": modification of work by NASA/Kepler mission/Wendy Stenzel)

What do we mean, exactly, by "discovery" of transiting exoplanets? A single transit shows up as a very slight drop in the brightness of the star, lasting several hours. However, astronomers must be on guard against other factors that might produce a false transit, especially when working at the limit of precision of the telescope. We must wait for a second transit of similar depth. But when another transit is observed, we don't initially know whether it might be due to another planet in a different orbit. The "discovery" occurs only when a third transit is found with similar depth and the same spacing in time as the first pair.

Computers normally conduct the analysis, which involves searching for tiny, periodic dips in the light from each star, extending over 4 years of observation. But the Kepler mission also has a program in which non-astronomers—citizen scientists—can examine the data. These dedicated volunteers have found several transits that were missed by the computer analyses, showing that the human eye and brain sometimes recognize unusual events that a computer was not programmed to look for.

Measuring three or four evenly spaced transits is normally enough to "discover" an exoplanet. But in a new field like exoplanet research, we would like to find further independent verification. The strongest confirmation happens when ground-based telescopes are also able to detect a Doppler shift with the same period as the transits. However, this is generally not possible for Earth-size planets. One of the most convincing ways to verify that a dip in brightness is due to a planet is to find more planets orbiting the same star—a *planetary system*. Multi-planet systems also provide alternative ways to estimate the masses of the planets, as we will discuss in the next section.

The selection effects (or biases) in the Kepler data are similar to those in Doppler observations. Large planets are easier to find than small ones, and short-period planets are easier than long-period planets. If we require three transits to establish the presence of a planet, we are of course limited to discovering planets with orbital periods less than one-third of the observing interval. Thus, it was only in its fourth and final year of operation that Kepler was able to find planets with orbits like Earth's that require 1 year to go around their star.

Direct Detection

The best possible evidence for an earthlike planet elsewhere would be an image. After all, "seeing is believing" is a very human prejudice. But imaging a distant planet is a formidable challenge indeed. Suppose, for example, you were a great distance away and wished to detect reflected light from Earth. Earth intercepts and reflects less than one billionth of the Sun's radiation, so its apparent brightness in visible light is less than one billionth that of the Sun. Compounding the challenge of detecting such a faint speck of light, the planet is swamped by the blaze of radiation from its parent star.

Even today, the best telescope mirrors' optics have slight imperfections that prevent the star's light from coming into focus in a completely sharp point.

Direct imaging works best for young gas giant planets that emit infrared light and reside at large separations from their host stars. Young giant planets emit more infrared light because they have more internal energy, stored from the process of planet formation. Even then, clever techniques must be employed to subtract out the light from the host star. In 2008, three such young planets were discovered orbiting HR 8799, a star in the constellation of Pegasus (Figure 21.21). Two years later, a fourth planet was detected closer to the star. Additional planets may reside even closer to HR 8799, but if they exist, they are currently lost in the glare of the star.

Since then, a number of planets around other stars have been found using direct imaging. However, one challenge is to tell whether the objects we are seeing are indeed planets or if they are brown dwarfs (failed stars) in orbit around a star.

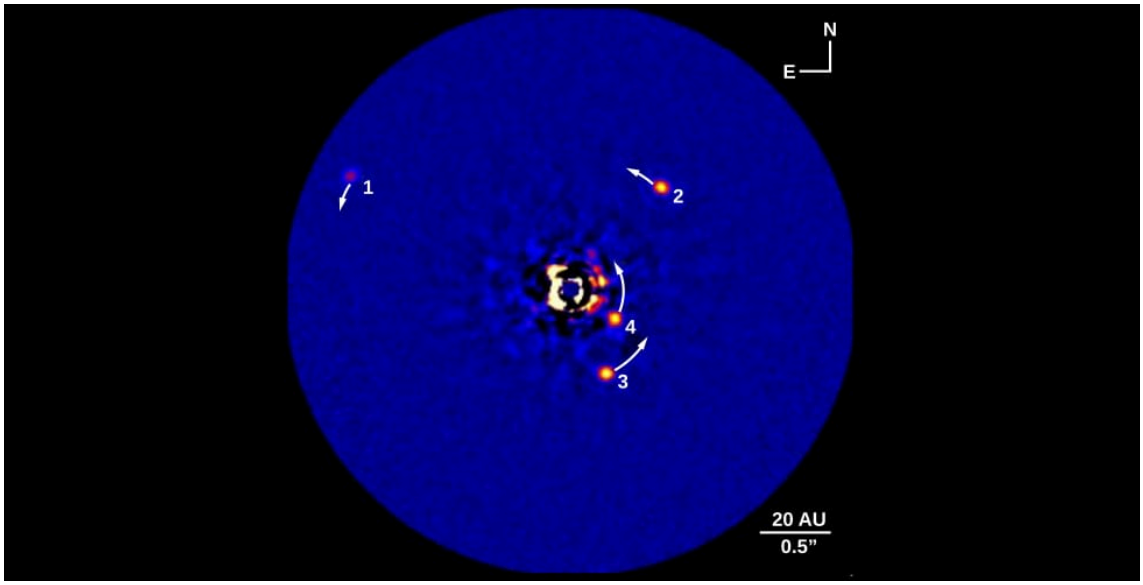


Figure 21.21 : Exoplanets around HR 8799. This image shows Keck telescope observations of four directly imaged planets orbiting HR 8799. A size scale for the system gives the distance in AU (remember that one astronomical unit is the distance between Earth and the Sun.) (credit: modification of work by Ben Zuckerman)

Direct imaging is an important technique for characterizing an exoplanet. The brightness of the planet can be measured at different wavelengths. These observations provide an estimate for the temperature of the planet's atmosphere; in the case of HR 8799 planet 1, the color suggests the presence of thick clouds. Spectra can also be obtained from the faint light to analyze the atmospheric constituents. A spectrum of HR 8799 planet 1 indicates a hydrogen-rich atmosphere, while the closer planet 4 shows evidence for methane in the atmosphere.

Another way to overcome the blurring effect of Earth's atmosphere is to observe from space. Infrared may be the optimal wavelength range in which to observe because planets get brighter in the infrared while stars like our Sun get fainter, thereby making it easier to detect a planet against the glare of its star. Special optical techniques can be used to suppress the light from the central star and make it easier to see the planet itself. However, even if we go into space, it will be difficult to obtain images of Earth-size planets.

Footnotes

- 1The Doppler method only allows us to find the minimum mass of a planet. To determine the exact mass using the Doppler shift and Kepler's laws, we must also have the angle at which the planet's orbit is oriented to our view—something we don't have any independent way of knowing in most cases. Still, if the minimum mass is half of Jupiter's, the actual mass can only be larger than that, and we are sure that we are dealing with a jovian planet.

Answers

Exercise 7.2.1

The radius of Earth is 6371 km. Therefore,
$\left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2 = \left(\frac{6371\ \text{km}}{695{,}700/2\ \text{km}}\right)^2 = \left(\frac{6371\ \text{km}}{347{,}850\ \text{km}}\right)^2 = 0.0003$, or significantly less than 1%.

This page titled 7.2: Planets beyond the Solar System- Search and Discovery is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 21.5: Planets beyond the Solar System- Search and Discovery by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

7.3: Exoplanets Everywhere - What we are Learning

Learning Objectives

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

- Explain what we have learned from our discovery of exoplanets
- Identify which kind of exoplanets appear to be the most common in the Galaxy
- Discuss the kinds of planetary systems we are finding around other stars

Before the discovery of exoplanets, most astronomers expected that other planetary systems would be much like our own—planets following roughly circular orbits, with the most massive planets several AU from their parent star. Such systems do exist in large numbers, but many exoplanets and planetary systems are very different from those in our solar system. Another surprise is the existence of whole classes of exoplanets that we simply don't have in our solar system: planets with masses between the mass of Earth and Neptune, and planets that are several times more massive than Jupiter.

Kepler Results

The Kepler telescope has been responsible for the discovery of most exoplanets, especially at smaller sizes, as illustrated in Figure 21.22, where the Kepler discoveries are plotted in yellow. You can see the wide range of sizes, including planets substantially larger than Jupiter and smaller than Earth. The absence of Kepler-discovered exoplanets with orbital periods longer than a few hundred days is a consequence of the 4-year lifetime of the mission. (Remember that three evenly spaced transits must be observed to register a discovery.) At the smaller sizes, the absence of planets much smaller than one earth radius is due to the difficulty of detecting transits by very small planets. In effect, the "discovery space" for Kepler was limited to planets with orbital periods less than 400 days and sizes larger than Mars.

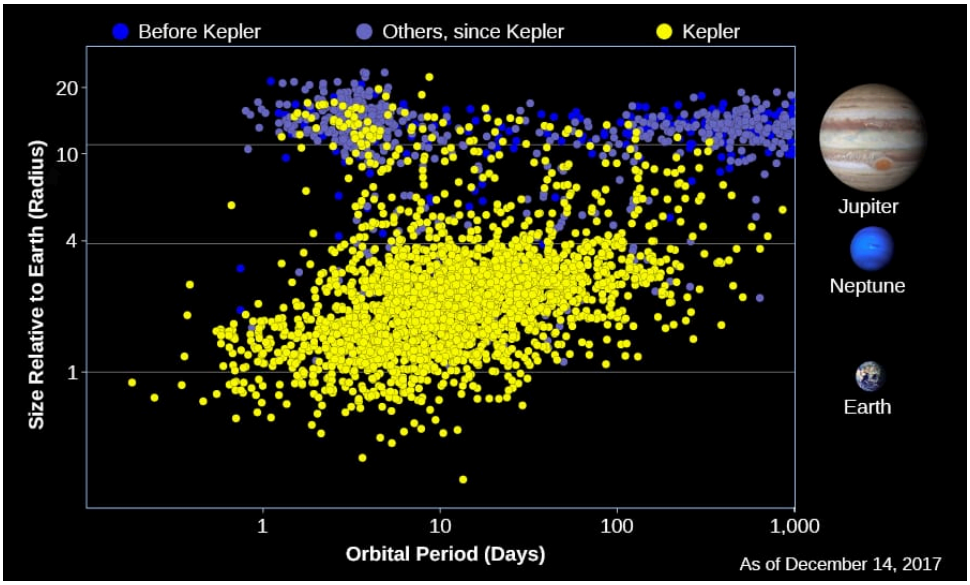


Figure 21.22 : Exoplanet Discoveries through 2017. The vertical axis shows the radius of each planet compared to Earth. Horizontal lines show the size of Earth, Neptune, and Jupiter. The horizontal axis shows the time each planet takes to make one orbit (and is given in Earth days). Recall that Mercury takes 88 days and Earth takes a little more than 365 days to orbit the Sun. The yellow dots show planets discovered by the Kepler mission using transits, and the blue dots are the discoveries by other projects and techniques. When this graph was made, astronomers had discovered just over 3,500 exoplanets; as of the start of 2022, that number has risen to roughly 5,000. (credit: modification of work by NASA/Ames Research Center/Jessie Dotson and Wendy Stenzel)

One of the primary objectives of the Kepler mission was to find out how many stars hosted planets and especially to estimate the frequency of earthlike planets. Although Kepler looked at only a very tiny fraction of the stars in the Galaxy, the sample size was

large enough to draw some interesting conclusions. While the observations apply only to the stars observed by Kepler, those stars are reasonably representative, and so astronomers can extrapolate to the entire Galaxy.

Figure 21.23 shows that the Kepler discoveries include many rocky, Earth-size planets, far more than Jupiter-size gas planets. This immediately tells us that the initial Doppler discovery of many hot Jupiters was a biased sample, in effect, finding the odd planetary systems because they were the easiest to detect. However, there is one huge difference between this observed size distribution and that of planets in our solar system. The most common planets have radii between 1.4 and 2.8 that of Earth, sizes for which we have no examples in the solar system. These have been nicknamed super-Earths, while the other large group with sizes between 2.8 and 4 that of Earth are often called mini-Neptunes.

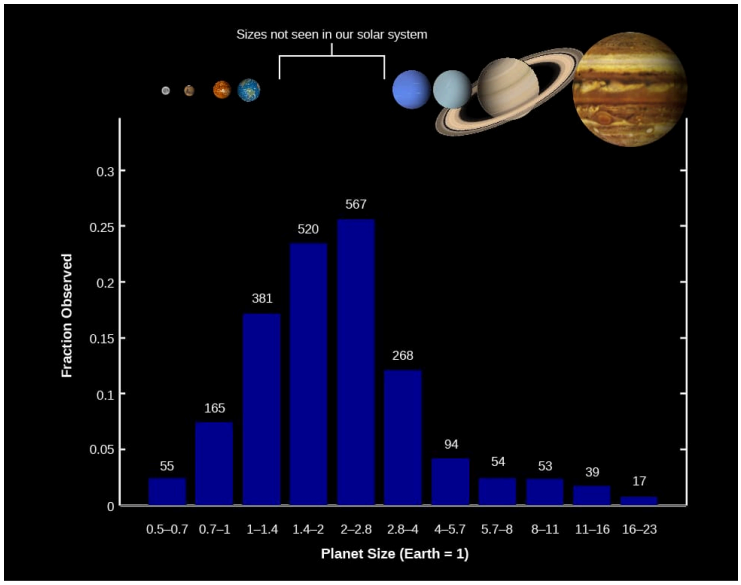


Figure 21.23 : Kepler Discoveries. This bar graph shows the number of planets of each size range found among the first 2213 Kepler planet discoveries. Sizes range from half the size of Earth to 20 times that of Earth. On the vertical axis, you can see the fraction that each size range makes up of the total. Note that planets that are between 1.4 and 4 times the size of Earth make up the largest fractions, yet this size range is not represented among the planets in our solar system. (credit: modification of work by NASA/Kepler mission)

What a remarkable discovery it is that the most common types of planets in the Galaxy are completely absent from our solar system and were unknown until Kepler's survey. However, recall that really small planets were difficult for the Kepler instruments to find. So, to estimate the frequency of Earth-size exoplanets, we need to correct for this sampling bias. The result is the corrected size distribution shown in Figure 21.24. Notice that in this graph, we have also taken the step of showing not the number of Kepler detections but the average number of planets per star for solar-type stars (spectral types F, G, and K).

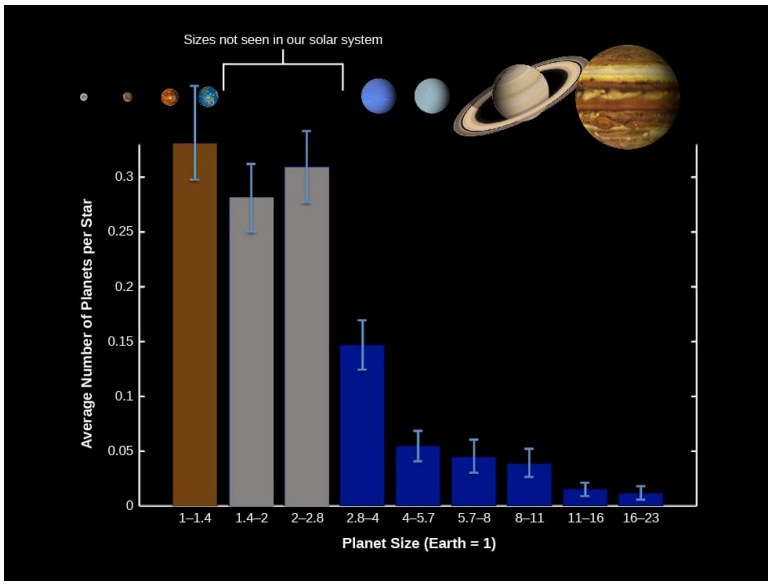


Figure 21.24 : Size Distribution of Planets for Stars Similar to the Sun. We show the average number of planets per star in each planet size range. (The average is less than one because some stars will have zero planets of that size range.) This distribution, corrected for biases in the Kepler data, shows that Earth-size planets may actually be the most common type of exoplanets. (credit: modification of work by NASA/Kepler mission)

We see that the most common planet sizes of are those with radii from 1 to 3 times that of Earth—what we have called “Earths” and “super-Earths.” Each group occurs in about one-third to one-quarter of stars. In other words, if we group these sizes together, we can conclude there is nearly one such planet per star! And remember, this census includes primarily planets with orbital periods less than 2 years. We do not yet know how many undiscovered planets might exist at larger distances from their star.

To estimate the number of Earth-size planets in our Galaxy, we need to remember that there are approximately 100 billion stars of spectral types F, G, and K. Therefore, we estimate that there are about 30 billion Earth-size planets in our Galaxy. If we include the super-Earths too, then there could be one hundred billion in the whole Galaxy. This idea—that planets of roughly Earth’s size are so numerous—is surely one of the most important discoveries of modern astronomy.

Planets with Known Densities

For several hundred exoplanets, we have been able to measure both the size of the planet from transit data and its mass from Doppler data, yielding an estimate of its density. Comparing the average density of exoplanets to the density of planets in our solar system helps us understand whether they are rocky or gaseous in nature. This has been particularly important for understanding the structure of the new categories of super-Earths and mini-Neptunes with masses between 3–10 times the mass of Earth. A key observation so far is that planets that are more than 10 times the mass of Earth have substantial gaseous envelopes (like Uranus and Neptune) whereas lower-mass planets are predominately rocky in nature (like the terrestrial planets).

Figure 21.25 compares all the exoplanets that have both mass and radius measurements. The dependence of the radius on planet mass is also shown for a few illustrative cases—hypothetical planets made of pure iron, rock, water, or hydrogen.

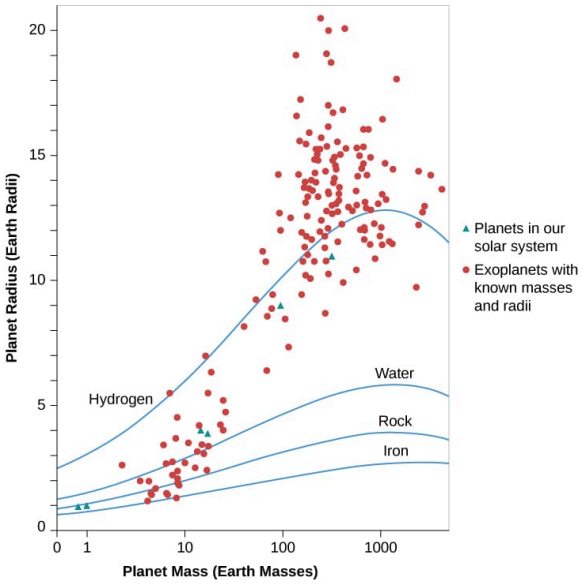


Figure 21.25 : Exoplanets with Known Densities. Exoplanets with known masses and radii (red circles) are plotted along with solid lines that show the theoretical size of pure iron, rock, water, and hydrogen planets with increasing mass. Masses are given in multiples of Earth's mass. (For comparison, Jupiter contains enough mass to make 320 Earths.) The green triangles indicate planets in our solar system.

At lower masses, notice that as the mass of these hypothetical planets increases, the radius also increases. That makes sense—if you were building a model of a planet out of clay, your toy planet would increase in size as you added more clay. However, for the highest mass planets (M > 1000 M_{Earth}) in Figure 21.25, notice that the radius stops increasing and the planets with greater mass are actually smaller. This occurs because increasing the mass also increases the gravity of the planet, so that compressible materials (even rock is compressible) will become more tightly packed, shrinking the size of the more massive planet.

In reality, planets are not pure compositions like the hypothetical water or iron planet. Earth is composed of a solid iron core, an outer liquid-iron core, a rocky mantle and crust, and a relatively thin atmospheric layer. Exoplanets are similarly likely to be differentiated into compositional layers. The theoretical lines in Figure 21.25 are simply guides that suggest a range of possible compositions.

Astronomers who work on the complex modeling of the interiors of rocky planets make the simplifying assumption that the planet consists of two or three layers. This is not perfect, but it is a reasonable approximation and another good example of how science works. Often, the first step in understanding something new is to narrow down the range of possibilities. This sets the stage for refining and deepening our knowledge. In Figure 21.25, the two green triangles with roughly 1 M_{Earth} and 1 R_{Earth} represent Venus and Earth. Notice that these planets fall between the models for a pure iron and a pure rock planet, consistent with what we would expect for the known mixed-chemical composition of Venus and Earth.

In the case of gaseous planets, the situation is more complex. Hydrogen is the lightest element in the periodic table, yet many of the detected exoplanets in Figure 21.25 with masses greater than 100 M_{Earth} have radii that suggest they are lower in density than a pure hydrogen planet. Hydrogen is the lightest element, so what is happening here? Why do some gas giant planets have inflated radii that are larger than the fictitious pure hydrogen planet? Many of these planets reside in short-period orbits close to the host star where they intercept a significant amount of radiated energy. If this energy is trapped deep in the planet atmosphere, it can cause the planet to expand.

Planets that orbit close to their host stars in slightly eccentric orbits have another source of energy: the star will raise tides in these planets that tend to circularize the orbits. This process also results in tidal dissipation of energy that can inflate the atmosphere. It would be interesting to measure the size of gas giant planets in wider orbits where the planets should be cooler—the expectation is that unless they are very young, these cooler gas giant exoplanets (sometimes called "cold Jupiters") should not be inflated. But we don't yet have data on these more distant exoplanets.

Exoplanetary Systems

As we search for exoplanets, we don't expect to find only one planet per star. Our solar system has eight major planets, half a dozen dwarf planets, and millions of smaller objects orbiting the Sun. The evidence we have of planetary systems in formation

also suggest that they are likely to produce multi-planet systems.

The first planetary system was found around the star Upsilon Andromedae in 1999 using the Doppler method, and many others have been found since then. If such exoplanetary system are common, let's consider which systems we expect to find in the Kepler transit data.

A planet will transit its star only if Earth lies in the plane of the planet's orbit. If the planets in other systems do not have orbits in the same plane, we are unlikely to see multiple transiting objects. Also, as we have noted before, Kepler was sensitive only to planets with orbital periods less than about 4 years. What we expect from Kepler data, then, is evidence of coplanar planetary systems confined to what would be the realm of the terrestrial planets in our solar system.

By 2022, astronomers gathered data on over 800 such exoplanet systems. Many have only two known planets, but a few have as many as five, and one has eight (the same number of planets as our own solar system). For the most part, these are very compact systems with most of their planets closer to their star than Mercury is to the Sun. The figure below shows one of the largest exoplanet systems: that of the star called Kepler-62 (Figure 21.26). Our solar system is shown to the same scale, for comparison (note that the Kepler-62 planets are drawn with artistic license; we have no detailed images of any exoplanets).

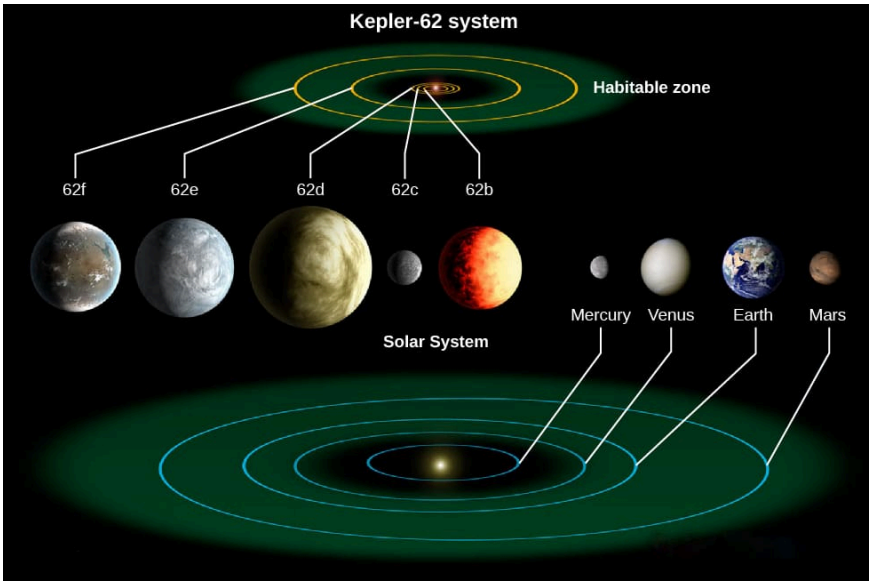


Figure 21.26 : Exoplanet System Kepler-62, with the Solar System Shown to the Same Scale. The green areas are the "habitable zones," the range of distance from the star where surface temperatures are likely to be consistent with liquid water. (credit: modification of work by NASA/ Ames/JPL-Caltech)

All but one of the planets in the K-62 system are larger than Earth. These are super-Earths, and one of them (62d) is in the size range of a mini-Neptune, where it is likely to be largely gaseous. The smallest planet in this system is about the size of Mars. The three inner planets orbit very close to their star, and only the outer two have orbits larger than Mercury in our system. The green areas represent each star's "habitable zone," which is the distance from the star where we calculate that surface temperatures would be consistent with liquid water. The Kepler-62 habitable zone is much smaller than that of the Sun because the star is intrinsically fainter.

With closely spaced systems like this, the planets can interact gravitationally with each other. The result is that the observed transits occur a few minutes earlier or later than would be predicted from simple orbits. These gravitational interactions have allowed the Kepler scientists to calculate masses for the planets, providing another way to learn about exoplanets.

Kepler has discovered some interesting and unusual planetary systems. For example, most astronomers expected planets to be limited to single stars. But we have found planets orbiting close double stars, so that the planet would see two suns in its sky, like those of the fictional planet Tatooine in the *Star Wars* films. At the opposite extreme, planets can orbit one star of a wide, double-star system without major interference from the second star.

This page titled 7.3: Exoplanets Everywhere - What we are Learning is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 21.6: Exoplanets Everywhere - What we are Learning by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

7.4: New Perspectives on Planet Formation

Learning Objectives

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

- Explain how exoplanet discoveries have revised our understanding of planet formation
- Discuss how planetary systems quite different from our solar system might have come about

Traditionally, astronomers have assumed that the planets in our solar system formed at about their current distances from the Sun and have remained there ever since. The first step in the formation of a giant planet is to build up a solid core, which happens when planetesimals collide and stick. Eventually, this core becomes massive enough to begin sweeping up gaseous material in the disk, thereby building the gas giants Jupiter and Saturn.

How to Make a Hot Jupiter

The traditional model for the formation of planets works only if the giant planets are formed far from the central star (about 5–10 AU), where the disk is cold enough to have a fairly high density of solid matter. It cannot explain the hot Jupiters, which are located very close to their stars where any rocky raw material would be completely vaporized. It also cannot explain the elliptical orbits we observe for some exoplanets because the orbit of a protoplanet, whatever its initial shape, will quickly become circular through interactions with the surrounding disk of material and will remain that way as the planet grows by sweeping up additional matter.

So we have two options: either we find a new model for forming planets close to the searing heat of the parent star, or we find a way to change the orbits of planets so that cold Jupiters can travel inward *after* they form. Most research now supports the latter explanation.

Calculations show that if a planet forms while a substantial amount of gas remains in the disk, then some of the planet's orbital angular momentum can be transferred to the disk. As it loses momentum (through a process that reminds us of the effects of friction), the planet will spiral inward. This process can transport giant planets, initially formed in cold regions of the disk, closer to the central star—thereby producing hot Jupiters. Gravitational interactions between planets in the chaotic early solar system can also cause planets to slingshot inward from large distances. But for this to work, the other planet has to carry away the angular momentum and move to a more distant orbit.

In some cases, we can use the combination of transit plus Doppler measurements to determine whether the planets orbit in the same plane and in the same direction as the star. For the first few cases, things seemed to work just as we anticipated: like the solar system, the gas giant planets orbited in their star's equatorial plane and in the same direction as the spinning star.

Then, some startling discoveries were made of gas giant planets that orbited at right angles or even in the opposite sense as the spin of the star. How could this happen? Again, there must have been interactions between planets. It's possible that before the system settled down, two planets came close together, so that one was kicked into an unusual orbit. Or perhaps a passing star perturbed the system after the planets were newly formed.

Forming Planetary Systems

When the Milky Way Galaxy was young, the stars that formed did not contain many heavy elements like iron. Several generations of star formation and star death were required to enrich the interstellar medium for subsequent generations of stars. Since planets seem to form "inside out," starting with the accretion of the materials that can make the rocky cores with which planets start, astronomers wondered when in the history of the Galaxy, planet formation would turn on.

The star Kepler-444 has shed some light on this question. This is a tightly packed system of five planets—the smallest comparable in size to Mercury and the largest similar in size to Venus. All five planets were detected with the Kepler spacecraft as they transited their parent star. All five planets orbit their host star in less than the time it takes Mercury to complete one orbit

about the Sun. Remarkably, the host star Kepler-444 is more than 11 billion years old and formed when the Milky Way was only 2 billion years old. So the heavier elements needed to make rocky planets must have already been available then. This ancient planetary system sets the clock on the beginning of rocky planet formation to be relatively soon after the formation of our Galaxy.

Kepler data demonstrate that while rocky planets inside Mercury's orbit are missing from our solar system, they are common around other stars, like Kepler-444. When the first systems packed with close-in rocky planets were discovered, we wondered why they were so different from our solar system. When many such systems were discovered, we began to wonder if it was our solar system that was different. This led to speculation that additional rocky planets might once have existed close to the Sun in our solar system.

There is some evidence from the motions in the outer solar system that Jupiter may have migrated inward long ago. If correct, then gravitational perturbations from Jupiter could have dislodged the orbits of close-in rocky planets, causing them to fall into the Sun. Consistent with this picture, astronomers now think that Uranus and Neptune probably did not form at their present distances from the Sun but rather closer to where Jupiter and Saturn are now. The reason for this idea is that density in the disk of matter surrounding the Sun at the time the planets formed was so low outside the orbit of Saturn that it would take several billion years to build up Uranus and Neptune. Yet we saw earlier in the chapter that the disks around protostars survive only a few million years.

Therefore, scientists have developed computer models demonstrating that Uranus and Neptune could have formed near the current locations of Jupiter and Saturn, and then been kicked out to larger distances through gravitational interactions with their neighbors. All these wonderful new observations illustrate how dangerous it can be to draw conclusions about a phenomenon in science (in this case, how planetary systems form and arrange themselves) when you are only working with a single example.

Exoplanets have given rise to a new picture of planetary system formation—one that is much more chaotic than we originally thought. If we think of the planets as being like skaters in a rink, our original model (with only our own solar system as a guide) assumed that the planets behaved like polite skaters, all obeying the rules of the rink and all moving in nearly the same direction, following roughly circular paths. The new picture corresponds more to a roller derby, where the skaters crash into one another, change directions, and sometimes are thrown entirely out of the rink.

Habitable Exoplanets

While thousands of exoplanets have been discovered in the past two decades, every observational technique has fallen short of finding more than a few candidates that resemble Earth (Figure 21.27). Astronomers are not sure exactly what properties would define another Earth. Do we need to find a planet that is *exactly* the same size and mass as Earth? That may be difficult and may not be important from the perspective of habitability. After all, we have no reason to think that life could not have arisen on Earth if our planet had been a little bit smaller or larger. And, remember that how habitable a planet is depends on both its distance from its star and the nature of its atmosphere. The greenhouse effect can make some planets warmer (as it did for Venus and is doing more and more for Earth).



Figure 21.27 : Many Earthlike Planets. This painting, commissioned by NASA, conveys the idea that there may be many planets resembling Earth out there as our methods for finding them improve. (credit: NASA/JPL-Caltech/R. Hurt (SSC-Caltech))

We can ask other questions to which we don't yet know the answers. Does this "twin" of Earth need to orbit a solar-type star, or can we consider as candidates the numerous exoplanets orbiting K- and M-class stars? (In the summer of 2016, astronomers reported the discovery of a planet with at least 1.3 times the mass of Earth around the nearest star, Proxima Centauri, which is spectral type M and located 4.2 light years from us.) We have a special interest in finding planets that could support life like ours, in which case, we need to find exoplanets within their star's habitable zone, where surface temperatures are consistent with liquid water on the surface. This is probably the most important characteristic defining an Earth-analog exoplanet.

The search for potentially habitable worlds is one of the prime drivers for exoplanet research in the next decade. Astronomers are beginning to develop realistic plans for new instruments that can even look for signs of life on distant worlds (examining their atmospheres for gases associated with life, for example). If we require telescopes in space to find such worlds, we need to recognize that years are required to plan, build, and launch such space observatories. The discovery of exoplanets and the knowledge that most stars have planetary systems are transforming our thinking about life beyond Earth. We are closer than ever to knowing whether habitable (and inhabited) planets are common. This work lends a new spirit of optimism to the search for life elsewhere, a subject to which we will return in Life in the Universe.

This page titled 7.4: New Perspectives on Planet Formation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 21.7: New Perspectives on Planet Formation by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

8: TERRESTRIAL PLANETS

8.1: Life, Chemical Evolution, and Climate Change
8.2: The Massive Atmosphere of Venus
8.3: Water and Life on Mars
8.4: Divergent Planetary Evolution

This page titled 8: Terrestrial Planets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.1: Life, Chemical Evolution, and Climate Change

Learning Objectives

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

- Outline the origins and subsequent diversity of life on Earth
- Explain the ways that life and geological activity have influenced the evolution of the atmosphere
- Describe the causes and effects of the atmospheric greenhouse effect and global warming
- Describe the impact of human activity on our planet's atmosphere and ecology

As far as we know, Earth seems to be the only planet in the solar system with life. The origin and development of life are an important part of our planet's story. Life arose early in Earth's history, and living organisms have been interacting with their environment for billions of years. We recognize that life-forms have evolved to adapt to the environment on Earth, and we are now beginning to realize that Earth itself has been changed in important ways by the presence of living matter. The study of the coevolution of life and our planet is one of the subjects of the modern science of *astrobiology*.

The Origin of Life

The record of the birth of life on Earth has been lost in the restless motions of the crust. According to chemical evidence, by the time the oldest surviving rocks were formed about 3.9 billion years ago, life already existed. At 3.5 billion years ago, life had achieved the sophistication to build large colonies called *stromatolites*, a form so successful that stromatolites still grow on Earth today (Figure 8.15). But, few rocks survive from these ancient times, and abundant fossils have been preserved only during the past 600 million years—less than 15% of our planet's history.



Figure 8.15 : Cross-Sections of Fossil Stromatolites. This polished cross-section of a fossilized colony of stromatolites dates to the Precambrian Era. The layered, domelike structures are mats of sediment trapped in shallow waters by large numbers of blue-green bacteria that can photosynthesize. Such colonies of microorganisms date back more than 3 billion years. (credit: James St. John)

There is little direct evidence about the actual origin of life. We know that the atmosphere of early Earth, unlike today's, contained abundant carbon dioxide and some methane, but no oxygen gas. In the absence of oxygen, many complex chemical reactions are possible that lead to the production of amino acids, proteins, and other chemical building blocks of life. Therefore, it seems likely that these chemical building blocks were available very early in Earth's history and they would have combined to make living organisms.

For tens of millions of years after Earth's formation, life (perhaps little more than large molecules, like the viruses of today) probably existed in warm, nutrient-rich seas, living off accumulated organic chemicals. When this easily accessible food became depleted, life began the long evolutionary road that led to the vast numbers of different organisms on Earth today. As it did so, life began to influence the chemical composition of the atmosphere.

In addition to the study of life's history as revealed by chemical and fossil evidence in ancient rocks, scientists use tools from the rapidly advancing fields of genetics and *genomics*—the study of the genetic code that is shared by all life on Earth. While each individual has a unique set of genes (which is why genetic "fingerprinting" is so useful for the study of crime), we also have many genetic traits in common. Your *genome*, the complete map of the DNA in your body, is identical at the 99.9% level to that of Julius Caesar or Marie Curie. At the 99% level, human and chimpanzee genomes are the same. By looking at the gene sequences of many organisms, we can determine that all life on Earth is descended from a common ancestor, and we can use the genetic variations among species as a measure of how closely different species are related.

These genetic analysis tools have allowed scientists to construct what is called the "tree of life" (Figure 8.16). This diagram illustrates the way organisms are related by examining one sequence of the nucleic acid RNA that all species have in common. This figure shows that life on Earth is dominated by microscopic creatures that you have probably never heard of. Note that the plant and animal kingdoms are just two little branches at the far right. Most of the diversity of life, and most of our evolution, has taken place at the microbial level. Indeed, it may surprise you to know that there are more microbes in a bucket of soil than there are stars in the Galaxy. You may want to keep this in mind when, later in this book, we turn to the search for life on other worlds. The "aliens" that are most likely to be out there are microbes.

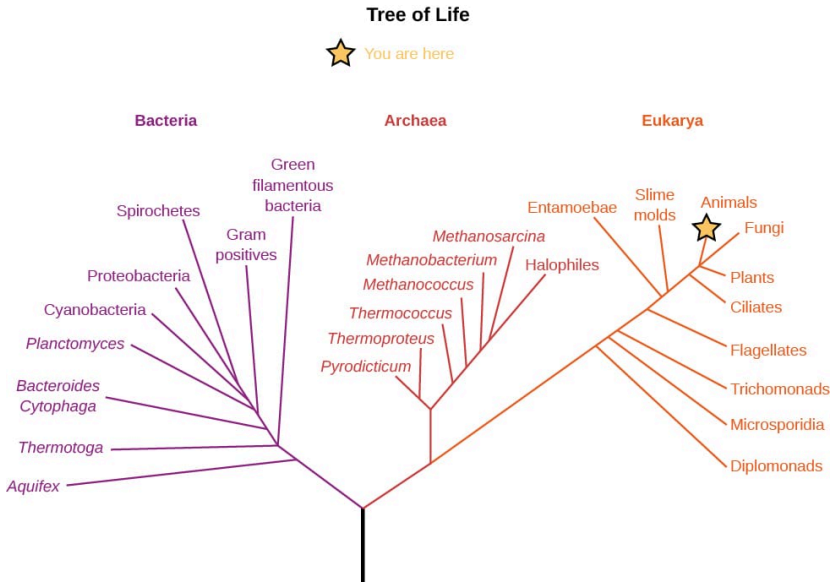


Figure 8.16 : Tree of Life. This chart shows the main subdivisions of life on Earth and how they are related. Note that the animal and plant kingdoms are just short branches on the far right, along with the fungi. The most fundamental division of Earth's living things is onto three large domains called bacteria, archaea, and eukarya. Most of the species listed are microscopic. (credit: modification of work by Eric Gaba)

Such genetic studies lead to other interesting conclusions as well. For example, it appears that the earliest surviving terrestrial life-forms were all adapted to live at high temperatures. Some biologists think that life might actually have begun in locations on our planet that were extremely hot. Yet another intriguing possibility is that life began on Mars (which cooled sooner) rather than Earth and was "seeded" onto our planet by meteorites traveling from Mars to Earth. Mars rocks are still making their way to Earth, but so far none has shown evidence of serving as a "spaceship" to carry microorganisms from Mars to Earth.

The Evolution of the Atmosphere

One of the key steps in the evolution of life on Earth was the development of blue-green algae, a very successful life-form that takes in carbon dioxide from the environment and releases oxygen as a waste product. These successful microorganisms proliferated, giving rise to all the lifeforms we call plants. Since the energy for making new plant material from chemical building blocks comes from sunlight, we call the process photosynthesis.

Studies of the chemistry of ancient rocks show that Earth's atmosphere lacked abundant free oxygen until about 2 billion years

ago, despite the presence of plants releasing oxygen by photosynthesis. Apparently, chemical reactions with Earth's crust removed the oxygen gas as quickly as it formed. Slowly, however, the increasing evolutionary sophistication of life led to a growth in the plant population and thus increased oxygen production. At the same time, it appears that increased geological activity led to heavy erosion on our planet's surface. This buried much of the plant carbon before it could recombine with oxygen to form CO_2.

Free oxygen began accumulating in the atmosphere about 2 billion years ago, and the increased amount of this gas led to the formation of Earth's ozone layer (recall that ozone is a triple molecule of oxygen, O_3), which protects the surface from deadly solar ultraviolet light. Before that, it was unthinkable for life to venture outside the protective oceans, so the landmasses of Earth were barren.

The presence of oxygen, and hence ozone, thus allowed colonization of the land. It also made possible a tremendous proliferation of animals, which lived by taking in and using the organic materials produced by plants as their own energy source.

As animals evolved in an environment increasingly rich in oxygen, they were able to develop techniques for breathing oxygen directly from the atmosphere. We humans take it for granted that plenty of free oxygen is available in Earth's atmosphere, and we use it to release energy from the food we take in. Although it may seem funny to think of it this way, we are lifeforms that have evolved to breathe in the waste product of plants. It is plants and related microbes that are the primary producers, using sunlight to create energy-rich "food" for the rest of us.

On a planetary scale, one of the consequences of life has been a decrease in atmospheric carbon dioxide. In the absence of life, Earth would probably have an atmosphere dominated by CO_2, like Mars or Venus. But living things, in combination with high levels of geological activity, have effectively stripped our atmosphere of most of this gas.

The Greenhouse Effect and Global Warming

We have a special interest in the carbon dioxide content of the atmosphere because of the key role this gas plays in retaining heat from the Sun through a process called the greenhouse effect. To understand how the greenhouse effect works, consider the fate of sunlight that strikes the surface of Earth. The light penetrates our atmosphere, is absorbed by the ground, and heats the surface layers. At the temperature of Earth's surface, that energy is then reemitted as infrared or heat radiation (Figure 8.17). However, the molecules of our atmosphere, which allow visible light through, are good at absorbing infrared energy. As a result, CO_2 (along with methane and water vapor) acts like a blanket, trapping heat in the atmosphere and impeding its flow back to space. To maintain an energy balance, the temperature of the surface and lower atmosphere must increase until the total energy radiated by Earth to space equals the energy received from the Sun. The more CO_2 there is in our atmosphere, the higher the temperature at which Earth's surface reaches a new balance.

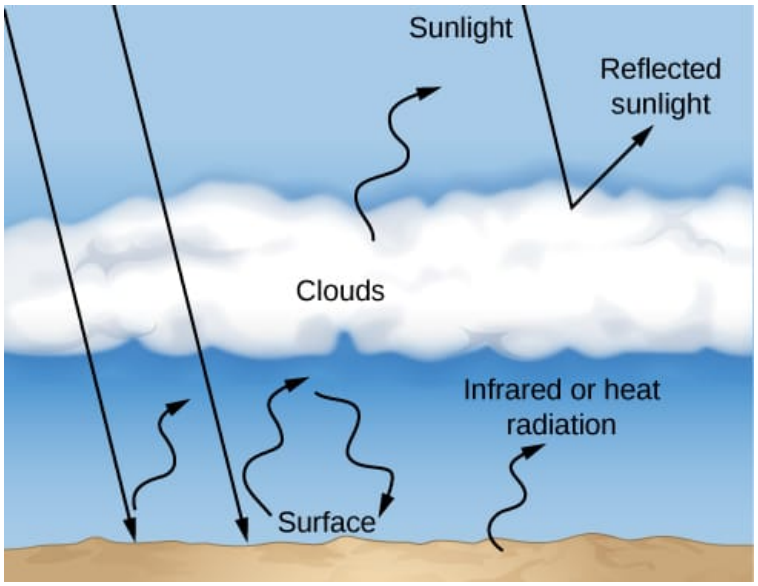


Figure 8.17 : How the Greenhouse Effect Works. Sunlight that penetrates to Earth's lower atmosphere and surface is reradiated as infrared or heat radiation, which is trapped by greenhouse gases such as water vapor, methane, and CO_2 in the atmosphere. The result is a higher surface temperature for our planet.

The greenhouse effect in a planetary atmosphere is similar to the heating of a gardener's greenhouse or the inside of a car left out in the Sun with the windows rolled up. In these examples, the window glass plays the role of greenhouse gases, letting sunlight in but reducing the outward flow of heat radiation. As a result, a greenhouse or car interior winds up much hotter than would be expected from the heating of sunlight alone. On Earth, the current greenhouse effect elevates the surface temperature by about 23 °C. Without this greenhouse effect, the average surface temperature would be well below freezing and Earth would be locked in a global ice age.

That's the good news; the bad news is that the heating due to the greenhouse effect is increasing. Modern industrial society depends on energy extracted from burning fossil fuels. In effect, we are exploiting the energy-rich material created by photosynthesis tens of millions of years ago. As these ancient coal and oil deposits are oxidized (burned using oxygen), large quantities of carbon dioxide are released into the atmosphere. The problem is exacerbated by the widespread destruction of tropical forests, which we depend on to extract CO_2 from the atmosphere and replenish our supply of oxygen. In the past century of increased industrial and agricultural development, the amount of CO_2 in the atmosphere increased by about 30% and continues to rise at more than 0.5% per year.

Before the end of the present century, Earth's CO_2 level is predicted to reach twice the value it had before the industrial revolution (Figure 8.18). The consequences of such an increase for Earth's surface and atmosphere (and the creatures who live there) are likely to be complex changes in climate, and may be catastrophic for many species. Many groups of scientists are now studying the effects of such global warming with elaborate computer models, and climate change has emerged as the greatest known threat (barring nuclear war) to both industrial civilization and the ecology of our planet.

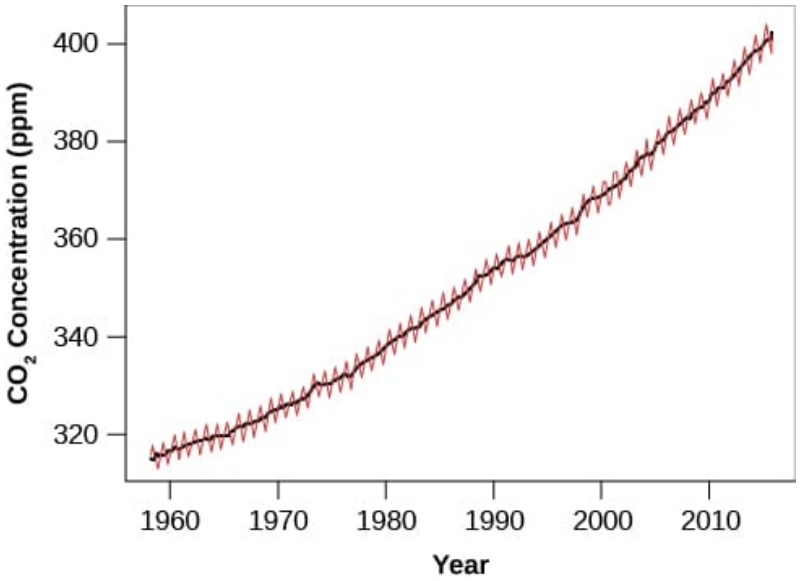


Figure 8.18 : Increase of Atmospheric Carbon Dioxide over Time. Scientists expect that the amount of CO_2 will double its preindustrial level before the end of the twenty-first century. Measurements of the isotopic signatures of this added CO_2 demonstrate that it is mostly coming from burning fossil fuels. (credit: modification of work by NOAA)

Link to Learning

This short PBS video explains the physics of the greenhouse effect.

Already climate change is widely apparent. Around the world, temperature records are constantly set and broken; all but one of the hottest recorded years have taken place since 2000. Glaciers are retreating, and the Arctic Sea ice is now much thinner than when it was first explored with nuclear submarines in the 1950s. Rising sea levels (from both melting glaciers and expansion of the water as its temperature rises) pose one of the most immediate threats, and many coastal cities have plans to build dikes or seawalls to hold back the expected flooding. The rate of temperature increase is without historical precedent, and we are rapidly entering "unknown territory" where human activities are leading to the highest temperatures on Earth in more than 50 million years.

Human Impacts on Our Planet

Earth is so large and has been here for so long that some people have trouble accepting that humans are really changing the planet, its atmosphere, and its climate. They are surprised to learn, for example, that the carbon dioxide released from burning fossil fuels is 100 times greater than that emitted by volcanoes. But, the data clearly tell the story that our climate is changing rapidly, and that almost all of the change is a result of human activity.

This is not the first time that humans have altered our environment dramatically. Some of the greatest changes were caused by our ancestors, before the development of modern industrial society. If aliens had visited Earth 50,000 years ago, they would have seen much of the planet supporting large animals of the sort that now survive only in Africa. The plains of Australia were occupied by giant marsupials such as diprotodon and zygomaturus (the size of our elephants today), and a species of kangaroo that stood 10 feet high. North America and North Asia hosted mammoths, saber tooth cats, mastodons, giant sloths, and even camels. The Islands of the Pacific teemed with large birds, and vast forests covered what are now the farms of Europe and China. Early human hunters killed many large mammals and marsupials, early farmers cut down most of the forests, and the Polynesian expansion across the Pacific doomed the population of large birds.

An even greater mass extinction is underway as a result of rapid climate change. In recognition of our impact on the environment, scientists have proposed giving a new name to the current epoch, the *anthropocene*, when human activity started to have a significant global impact. Although not an officially approved name, the concept of “anthropocene” is useful for recognizing that we humans now represent the dominant influence on our planet’s atmosphere and ecology, for better or for worse.

This page titled 8.1: Life, Chemical Evolution, and Climate Change is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 8.5: Life, Chemical Evolution, and Climate Change by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

8.2: The Massive Atmosphere of Venus

Learning Objectives

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

- Describe the general composition and structure of the atmosphere on Venus
- Explain how the greenhouse effect has led to high temperatures on Venus

The thick atmosphere of Venus produces the high surface temperature and shrouds the surface in a perpetual red twilight. Sunlight does not penetrate directly through the heavy clouds, but the surface is fairly well lit by diffuse light (about the same as the light on Earth under a heavy overcast). The weather at the bottom of this deep atmosphere remains perpetually hot and dry, with calm winds. Because of the heavy blanket of clouds and atmosphere, one spot on the surface of Venus is similar to any other as far as weather is concerned.

Composition and Structure of the Atmosphere

The most abundant gas on Venus is carbon dioxide (CO_2), which accounts for 96% of the atmosphere. The second most common gas is nitrogen. The predominance of carbon dioxide over nitrogen is not surprising when you recall that Earth's atmosphere would also be mostly carbon dioxide if this gas were not locked up in marine sediments (see the discussion of Earth's atmosphere in Earth as a Planet).

Table 10.2 compares the compositions of the atmospheres of Venus, Mars, and Earth. Expressed in this way, as percentages, the proportions of the major gases are very similar for Venus and Mars, but in total quantity, their atmospheres are dramatically different. With its surface pressure of 90 bars, the venusian atmosphere is more than 10,000 times more massive than its martian counterpart. Overall, the atmosphere of Venus is very dry; the absence of water is one of the important ways that Venus differs from Earth.

Table 10.2: Atmospheric Composition of Earth, Venus, and Mars

Gas	Earth	Venus	Mars
Carbon dioxide (CO_2)	0.03%	96%	95.3%
Nitrogen (N_2)	78.1%	3.5%	2.7%
Argon (Ar)	0.93%	0.006%	1.6%
Oxygen (O_2)	21.0%	0.003%	0.15%
Neon (Ne)	0.002%	0.001%	0.0003%

The atmosphere of Venus has a huge troposphere (region of convection) that extends up to at least 50 kilometers above the surface (Figure 10.12). Within the troposphere, the gas is heated from below and circulates slowly, rising near the equator and descending over the poles. Being at the base of the atmosphere of Venus is something like being a kilometer or more below the ocean surface on Earth. There, the mass of water evens out temperature variations and results in a uniform environment—the same effect the thick atmosphere has on Venus.

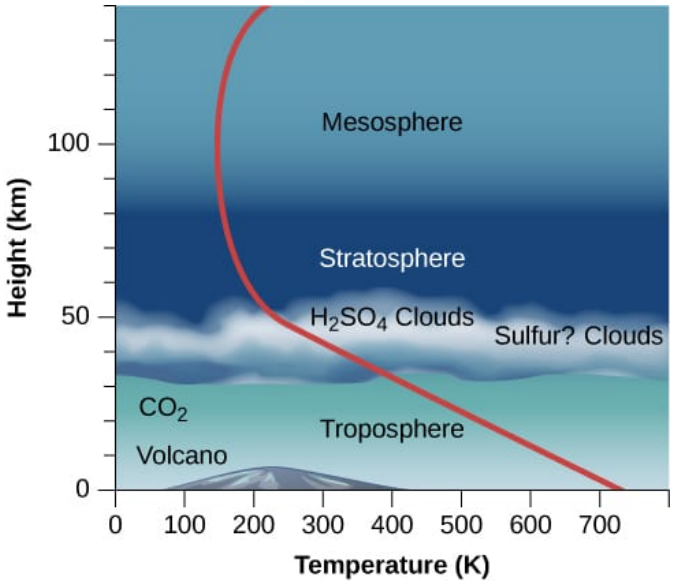


Figure 10.12 : Venus' Atmosphere. The layers of the massive atmosphere of Venus shown here are based on data from the Pioneer and Venera entry probes. Height is measured along the left axis, the bottom scale shows temperature, and the red line allows you to read off the temperature at each height. Notice how steeply the temperature rises below the clouds, thanks to the planet's huge greenhouse effect.

In the upper troposphere, between 30 and 60 kilometers above the surface, a thick cloud layer is composed primarily of sulfuric acid droplets. Sulfuric acid (H_2SO_4) is formed from the chemical combination of sulfur dioxide (SO_2) and water (H_2O). In the atmosphere of Earth, sulfur dioxide is one of the primary gases emitted by volcanoes, but it is quickly diluted and washed out by rainfall. In the dry atmosphere of Venus, this unpleasant substance is apparently stable. Below 30 kilometers, the Venus atmosphere is clear of clouds.

Surface Temperature on Venus

The high surface temperature of Venus was discovered by radio astronomers in the late 1950s and confirmed by the Mariner and Venera probes. How can our neighbor planet be so hot? Although Venus is somewhat closer to the Sun than is Earth, its surface is hundreds of degrees hotter than you would expect from the extra sunlight it receives. Scientists wondered what could be heating the surface of Venus to a temperature above 700 K. The answer turned out to be the *greenhouse effect*.

The greenhouse effect works on Venus just as it does on Earth, but since Venus has so much more CO_2—almost a million times more—the effect is much stronger. The thick CO_2 acts as a blanket, making it very difficult for the infrared (heat) radiation from the ground to get back into space. As a result, the surface heats up. The energy balance is only restored when the planet is radiating as much energy as it receives from the Sun, but this can happen only when the temperature of the lower atmosphere is very high. One way of thinking of greenhouse heating is that it must raise the surface temperature of Venus until this energy balance is achieved.

Has Venus always had such a massive atmosphere and high surface temperature, or might it have evolved to such conditions from a climate that was once more nearly earthlike? The answer to this question is of particular interest to us as we look at the increasing levels of CO_2 in Earth's atmosphere. As the greenhouse effect becomes stronger on Earth, are we in any danger of transforming our own planet into a hellish place like Venus?

Let us try to reconstruct the possible evolution of Venus from an earthlike beginning to its present state. Venus may once have had a climate similar to that of Earth, with moderate temperatures, water oceans, and much of its CO_2 dissolved in the ocean or chemically combined with the surface rocks. Then we allow for modest additional heating—by gradual increase in the energy output of the Sun, for example. When we calculate how Venus' atmosphere would respond to such effects, it turns out that even a small amount of extra heat can lead to increased evaporation of water from the oceans and the release of gas from surface rocks.

This in turn means a further increase in the atmospheric CO_2 and H_2O, gases that would amplify the greenhouse effect in Venus' atmosphere. That would lead to still more heat near Venus' surface and the release of further CO_2 and H_2O. Unless some other processes intervene, the temperature thus continues to rise. Such a situation is called the runaway greenhouse effect.

We want to emphasize that the runaway greenhouse effect is not just a large greenhouse effect; it is an evolutionary *process*. The atmosphere evolves from having a small greenhouse effect, such as on Earth, to a situation where greenhouse warming is a major factor, as we see today on Venus. Once the large greenhouse conditions develop, the planet establishes a new, much hotter equilibrium near its surface.

Reversing the situation is difficult because of the role water plays. On Earth, most of the CO_2 is either chemically bound in the rocks of our crust or dissolved by the water in our oceans. As Venus got hotter and hotter, its oceans evaporated, eliminating that safety valve. But the water vapor in the planet's atmosphere will not last forever in the presence of ultraviolet light from the Sun. The light element hydrogen can escape from the atmosphere, leaving the oxygen behind to combine chemically with surface rock. The loss of water is therefore an irreversible process: once the water is gone, it cannot be restored. There is evidence that this is just what happened to the water once present on Venus.

We don't know if the same runaway greenhouse effect could one day happen on Earth. Although we are uncertain about the point at which a stable greenhouse effect breaks down and turns into a runaway greenhouse effect, Venus stands as clear testament to the fact that a planet cannot continue heating indefinitely without a major change in its oceans and atmosphere. It is a conclusion that we and our descendants will surely want to pay close attention to.

This page titled 8.2: The Massive Atmosphere of Venus is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 10.4: The Massive Atmosphere of Venus by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

8.3: Water and Life on Mars

Learning Objectives

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

- Describe the general composition of the atmosphere on Mars
- Explain what we know about the polar ice caps on Mars and how we know it
- Describe the evidence for the presence of water in the past history of Mars
- Summarize the evidence for and against the possibility of life on Mars

Of all the planets and moons in the solar system, Mars seems to be the most promising place to look for life, both fossil microbes and (we hope) some forms of life deeper underground that still survive today. But where (and how) should we look for life? We know that the one requirement shared by all life on Earth is liquid water. Therefore, the guiding principle in assessing habitability on Mars and elsewhere has been to "follow the water." That is the perspective we take in this section, to follow the water on the red planet and hope it will lead us to life.

Atmosphere and Clouds on Mars

The atmosphere of Mars today has an average surface pressure of only 0.007 bar, less than 1% that of Earth. (This is how thin the air is about 30 kilometers above Earth's surface.) Martian air is composed primarily of carbon dioxide (95%), with about 3% nitrogen and 2% argon. The proportions of different gases are similar to those in the atmosphere of Venus (see Table 10.2), but a lot less of each gas is found in the thin air on Mars.

While winds on Mars can reach high speeds, they exert much less force than wind of the same velocity would on Earth because the atmosphere is so thin. The wind is able, however, to loft very fine dust particles, which can sometimes develop planet-wide dust storms. It is this fine dust that coats almost all the surface, giving Mars its distinctive red color. In the absence of surface water, wind erosion plays a major role in sculpting the martian surface (Figure 10.24).

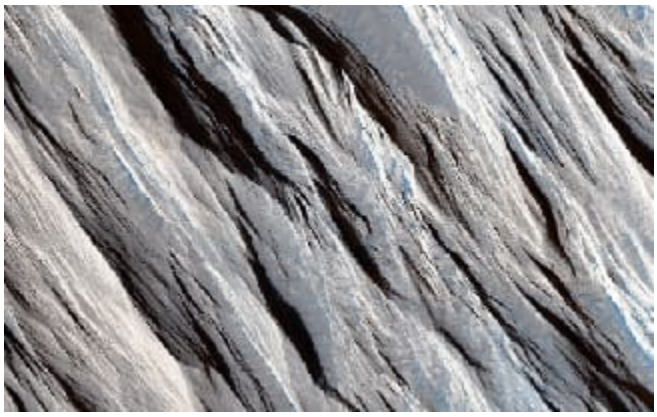


Figure 10.24 : Wind Erosion on Mars. These long straight ridges, called yardangs, are aligned with the dominant wind direction. This is a high-resolution image from the *Mars Reconnaissance Orbiter* and is about 1 kilometer wide. (credit: NASA/JPL-Caltech/University of Arizona)

Link to Learning

The issue of how strong the winds on Mars can be plays a big role in the 2015 hit movie *The Martian* in which the main character is stranded on Mars after being buried in the sand in a windstorm so great that his fellow astronauts have to leave the planet so their ship is not damaged. Astronomers have noted that the martian winds could not possibly be as forceful as depicted in the film because the air pressure is so low. In most ways, however, the depiction of Mars in this movie is remarkably accurate.

Although the atmosphere contains small amounts of water vapor and occasional clouds of water ice, liquid water is not stable under present conditions on Mars. Part of the problem is the low temperatures on the planet. But even if the temperature on a sunny summer day rises above the freezing point, the low pressure means that liquid water still cannot exist on the surface, except at the lowest elevations. At a pressure of less than 0.006 bar, the boiling point is as low or lower than the freezing point, and water changes directly from solid to vapor without an intermediate liquid state (as does "dry ice," carbon dioxide, on Earth). However, salts dissolved in water lower its freezing point, as we know from the way salt is used to thaw roads after snow and ice forms during winter on Earth. Salty water is therefore sometimes able to exist in liquid form on the martian surface, under the right conditions.

Several types of clouds can form in the martian atmosphere. First there are dust clouds, discussed above. Second are water-ice clouds similar to those on Earth. These often form around mountains, just as happens on our planet. Finally, the CO_2 of the atmosphere can itself condense at high altitudes to form hazes of dry ice crystals. The CO_2 clouds have no counterpart on Earth, since on our planet temperatures never drop low enough (down to about 150 K or about –125 °C) for this gas to condense.

The Polar Caps

Through a telescope, the most prominent surface features on Mars are the bright polar caps, which change with the seasons, similar to the seasonal snow cover on Earth. We do not usually think of the winter snow in northern latitudes as a part of our polar caps, but seen from space, the thin winter snow merges with Earth's thick, permanent ice caps to create an impression much like that seen on Mars (Figure 10.25).

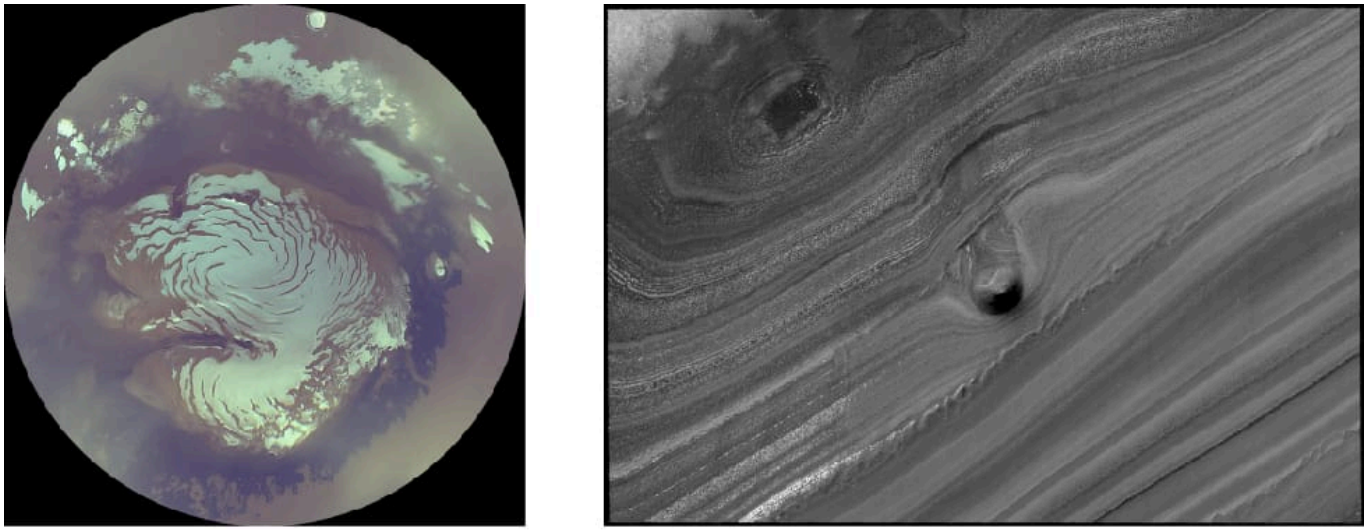


Figure 10.25 : Martian North Polar Cap. (a) This is a composite image of the north pole in summer, obtained in October 2006 by the *Mars Reconnaissance Orbiter*. It shows the mostly water-ice residual cap sitting atop light, tan-colored, layered sediments. Note that although the border of this photo is circular, it shows only a small part of the planet. (b) Here we see a small section of the layered terrain near the martian north pole. There is a mound about 40 meters high that is sticking out of a trough in the center of the picture. (credit a: modification of work by NASA/JPL/MSSS; credit b: modification of work by NASA/JPL-Caltech/University of Arizona)

The *seasonal caps* on Mars are composed not of ordinary snow but of frozen CO_2 (dry ice). These deposits condense directly from the atmosphere when the surface temperature drops below about 150 K. The caps develop during the cold martian winters and extend down to about 50° latitude by the start of spring.

Quite distinct from these thin seasonal caps of CO_2 are the *permanent* or *residual caps* that are always present near the poles. The southern permanent cap has a diameter of 350 kilometers and is composed of frozen CO_2 deposits together with a great deal of water ice. Throughout the southern summer, it remains at the freezing point of CO_2, 150 K, and this cold reservoir is thick enough to survive the summer heat intact.

The northern permanent cap is different. It is much larger, never shrinking to a diameter less than 1000 kilometers, and is composed of water ice. Summer temperatures in the north are too high for the frozen CO_2 to be retained. Measurements from the *Mars Global Surveyor* have established the exact elevations in the north polar region of Mars, showing that it is a large basin

about the size of our own Arctic Ocean basin. The ice cap itself is about 3 kilometers thick, with a total volume of about 10 million km^3 (similar to that of Earth's Mediterranean Sea). If Mars ever had extensive liquid water, this north polar basin would have contained a shallow sea. There is some indication of ancient shorelines visible, but better images will be required to verify this suggestion.

Images taken from orbit also show a distinctive type of terrain surrounding the permanent polar caps, as shown in Figure 10.25. At latitudes above 80° in both hemispheres, the surface consists of recent layered deposits that cover the older cratered ground below. Individual layers are typically ten to a few tens of meters thick, marked by alternating light and dark bands of sediment. Probably the material in the polar deposits includes dust carried by wind from the equatorial regions of Mars.

What do these terraced layers tell us about Mars? Some cyclic process is depositing dust and ice over periods of time. The time scales represented by the polar layers are tens of thousands of years. Apparently the martian climate experiences periodic changes at intervals similar to those between ice ages on Earth. Calculations indicate that the causes are probably also similar: the gravitational pull of the other planets produces variations in Mars' orbit and tilt as the great clockwork of the solar system goes through its paces.

The *Phoenix* spacecraft landed near the north polar cap in summer (Figure 10.26). Controllers knew that it would not be able to survive a polar winter, but directly measuring the characteristics of the polar region was deemed important enough to send a dedicated mission. The most exciting discovery came when the spacecraft tried to dig a shallow trench under the spacecraft. When the overlying dust was stripped off, they saw bright white material, apparently some kind of ice. From the way this ice sublimated over the next few days, it was clear that it was frozen water.

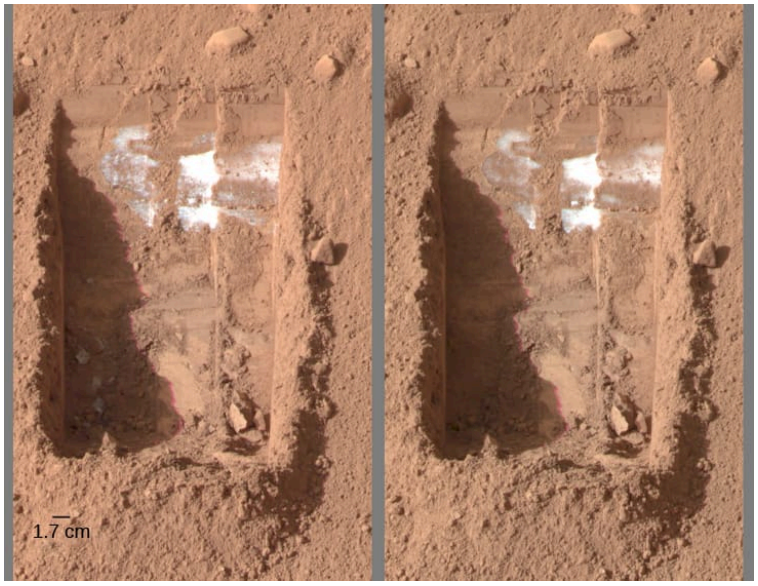


Figure 10.26 : Evaporating Ice on Mars. We see a trench dug by the *Phoenix* lander in the north polar region four martian days apart in June 2008. If you look at the shadowed region in the bottom left of the trench, you can see three spots of ice in the left image which have sublimated away in the right image. (credit: modification of work by NASA/JPL-Caltech/University of Arizona/Texas A&M University)

Example 10.1: Comparing the Amount of Water on Mars and Earth

It is interesting to estimate the amount of water (in the form of ice) on Mars and to compare this with the amount of water on Earth. In each case, we can find the total volume of a layer on a sphere by multiplying the area of the sphere Math input error

Solution

The volume of Earth's water is therefore the area $4\pi R^2$

$$4\pi\left(6.378\ \times\ 10^6\mathrm{m}\right)^2 = 5.1\ \times\ 10^{14}\ \mathrm{m}^2$$

multiplied by the thickness of 3000 m:

$$5.1 \times 10^{14}\,\mathrm{m}^2 \times 3000\,\mathrm{m} = 1.5 \times 10^{18}\,\mathrm{m}^3$$

This gives $1.5 \times 10^{18}\mathrm{m}^3$ of water. Since water has a density of 1 ton per cubic meter (1000 kg/m^3), we can calculate the mass:

$$1.5 \times 10^{18}\,\mathrm{m}^3 \times 1\,\mathrm{ton/m}^3 = 1.5 \times 10^{18}\,\mathrm{tons}$$

For Mars, the ice doesn't cover the whole planet, only the caps; the polar cap area is

$$\pi R^2 = \pi(4 \times 10^5\,\mathrm{m})^2 = 5 \times 10^{11}\,\mathrm{m}^2$$

(Note that we converted kilometers to meters.)

The volume = area × height, so we have:

$$(2 \times 10^3\,\mathrm{m})(5 \times 10^{11}\,\mathrm{m}^2) = 1 \times 10^{15}\,\mathrm{m}^3 = 10^{15}\,\mathrm{m}^3$$

Therefore, the mass is:

$$10^{15}\,\mathrm{m}^3 \times 1\,\mathrm{ton/m}^3 = 10^{15}\,\mathrm{tons}$$

This is about 0.1% that of Earth's oceans.

Exercise 8.3.1

A better comparison might be to compare the amount of ice in the Mars polar ice caps to the amount of ice in the Greenland ice sheet on Earth, which has been estimated as 2.85 × 10^{15} m^3. How does this compare with the ice on Mars?

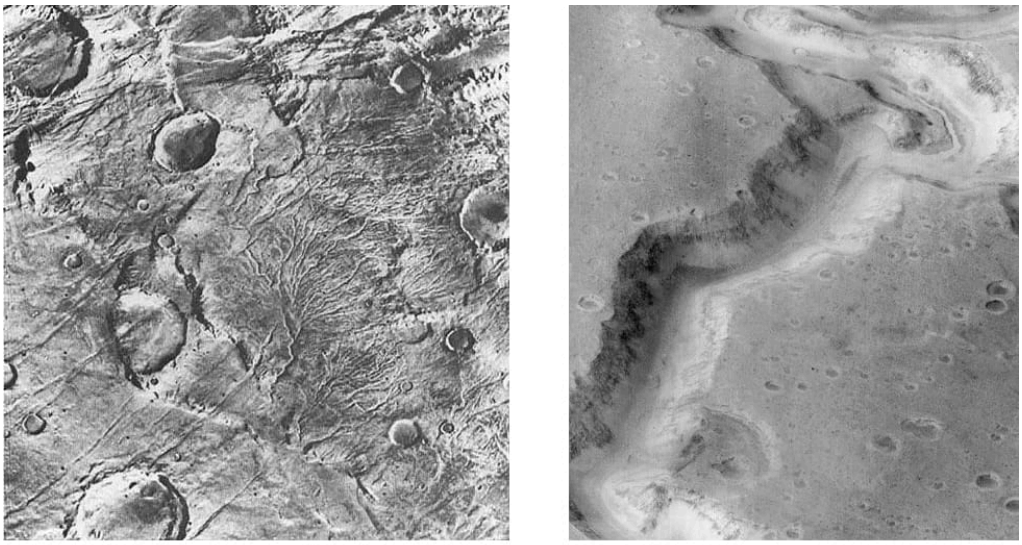
Answer available at the end of this section.

Channels and Gullies on Mars

Although no bodies of liquid water exist on Mars today, evidence has accumulated that rivers flowed on the red planet long ago. Two kinds of geological features appear to be remnants of ancient watercourses, while a third class—smaller gullies—suggests intermittent outbreaks of liquid water even today. We will examine each of these features in turn.

In the highland equatorial plains, there are multitudes of small, sinuous (twisting) channels—typically a few meters deep, some tens of meters wide, and perhaps 10 or 20 kilometers long (Figure 10.27). They are called runoff channels because they look like what geologists would expect from the surface runoff of ancient rain storms. These runoff channels seem to be telling us that the planet had a very different climate long ago. To estimate the age of these channels, we look at the cratering record. Crater counts show that this part of the planet is more cratered than the lunar maria but less cratered than the lunar highlands. Thus, the runoff channels are probably older than the lunar maria, presumably about 4 billion years old.

The second set of water-related features we see are *outflow channels* (Figure 10.27) are much larger than the runoff channels. The largest of these, which drain into the Chryse basin where Pathfinder landed, are 10 kilometers or more wide and hundreds of kilometers long. Many features of these outflow channels have convinced geologists that they were carved by huge volumes of running water, far too great to be produced by ordinary rainfall. Where could such floodwater have come from on Mars?



(a) (b)

Figure 10.27 : Runoff and Outflow Channels. (a) These runoff channels in the old martian highlands are interpreted as the valleys of ancient rivers fed by either rain or underground springs. The width of this image is about 200 kilometers. (b) This intriguing channel, called Nanedi Valles, resembles Earth riverbeds in some (but not all) ways. The tight curves and terraces seen in the channel certainly suggest the sustained flow of a fluid like water. The channel is about 2.5 kilometers across. (credit a: modification of work by Jim Secosky/NASA; credit b: modification of work by Jim Secosky/NASA)

As far we can tell, the regions where the outflow channels originate contained abundant water frozen in the soil as permafrost. Some local source of heating must have released this water, leading to a period of rapid and catastrophic flooding. Perhaps this heating was associated with the formation of the volcanic plains on Mars, which date back to roughly the same time as the outflow channels.

Note that neither the runoff channels nor the outflow channels are wide enough to be visible from Earth, nor do they follow straight lines. They could not have been the “canals” Percival Lowell imagined seeing on the red planet.

The third type of water feature, the smaller *gullies*, was discovered by the *Mars Global Surveyor* (Figure 10.28). The *Mars Global Surveyor*’s camera images achieved a resolution of a few meters, good enough to see something as small as a truck or bus on the surface. On the steep walls of valleys and craters at high latitudes, there are many erosional features that look like gullies carved by flowing water. These gullies are very young: not only are there no superimposed impact craters, but in some instances, the gullies seem to cut across recent wind-deposited dunes. Perhaps there is liquid water underground that can occasionally break out to produce short-lived surface flows before the water can freeze or evaporate.

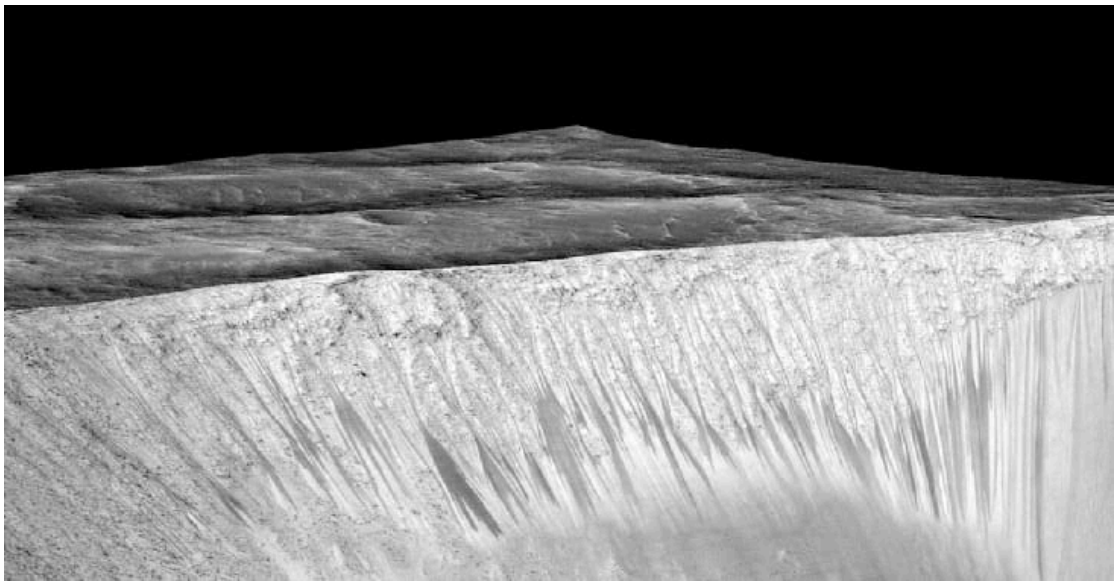


Figure 10.28 : Gullies on the Wall of Garni Crater. This high-resolution image is from the *Mars Reconnaissance Orbiter*. The dark streaks, which are each several hundred meters long, change in a seasonal pattern that suggests they are caused by the temporary flow of surface water. (credit: NASA/JPL-Caltech/University of Arizona)

The gullies also have the remarkable property of changing regularly with the martian seasons. Many of the dark streaks (visible in Figure 10.28) elongate within a period of a few days, indicating that something is flowing downhill—either water or dark sediment. If it is water, it requires a continuing source, either from the atmosphere or from springs that tap underground water layers (aquifers.) Underground water would be the most exciting possibility, but this explanation seems inconsistent with the fact that many of the dark streaks start at high elevations on the walls of craters.

Additional evidence that the dark streaks (called by the scientists *recurring slope lineae*) are caused by water was found in 2015 when spectra were obtained of the dark streaks (Figure 10.29). These showed the presence of hydrated salts produced by the evaporation of salty water. If the water is salty, it could remain liquid long enough to flow downstream for distances of a hundred meters or more, before it either evaporates or soaks into the ground. However, this discovery still does not identify the ultimate source of the water.

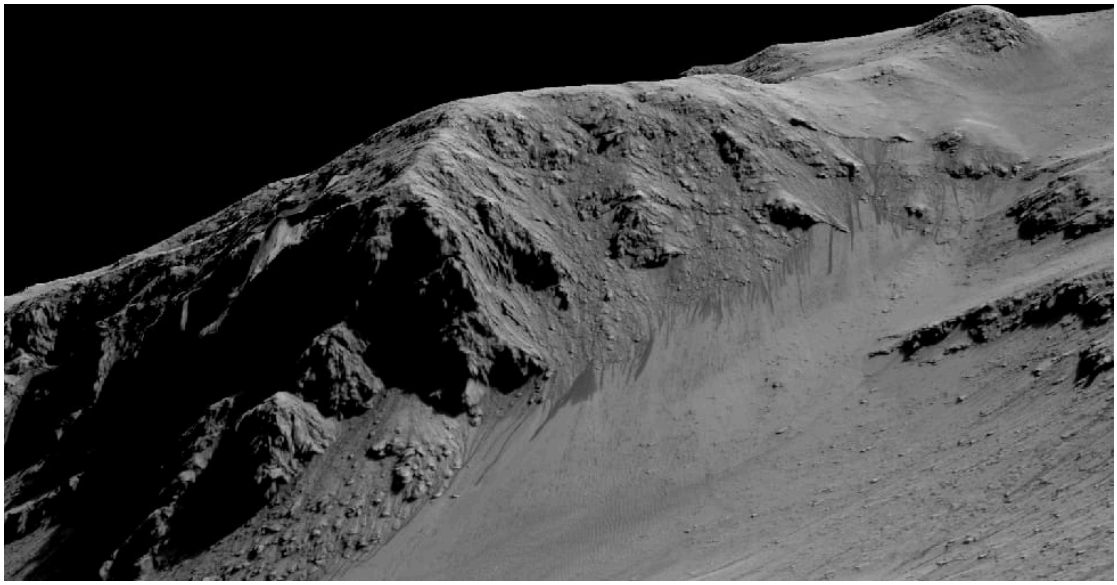


Figure 10.29 : Evidence for Liquid Water on Mars. The dark streaks in Horowitz crater, which move downslope, have been called recurring slope lineae. The streaks in the center of the image go down the wall of the crater for about a distance of 100 meters. Spectra taken of this region indicate that these are locations where salty liquid water flows on or just below the surface of Mars. (The vertical dimension is exaggerated by a factor of 1.5 compared to horizontal dimensions.) (credit: NASA/JPL-Caltech/University of Arizona)

Ancient Lakes and Glaciers

The rovers (*Spirit*, *Opportunity*, *Curiosity*, and *Perseverance,*, that have operated on the surface of Mars have been used to hunt for additional evidence of water. They could not reach the most interesting sites, such as the gullies, which are located on steep slopes. Instead, they explored sites that might be dried-out lake beds, dating back to a time when the climate on Mars was warmer and the atmosphere thicker—allowing water to be liquid on the surface.

Spirit was specifically targeted to explore what looked like an ancient lake-bed in Gusev crater, with an outflow channel emptying into it. However, when the spacecraft landed, it found that the former lakebed had been covered by thin lava flows, blocking the rover from access to the sedimentary rocks it had hoped to find. However, *Opportunity* had better luck. Peering at the walls of a small crater, it detected layered sedimentary rock. These rocks contained chemical evidence of evaporation, suggesting there had been a shallow salty lake in that location. In these sedimentary rocks were also small spheres that were rich in the mineral hematite, which forms only in watery environments. Apparently this very large basin had once been underwater.

Link to Learning

The small spherical rocks were nicknamed “blueberries” by the science team and the discovery of a whole “berry-bowl” of them was announced in this interesting news release from NASA.

The *Curiosity* rover landed inside Gale crater, where photos taken from orbit also suggested past water erosion. It discovered numerous sedimentary rocks, some in the form of mudstones from an ancient lakebed; it also found indications of rocks formed by the action of shallow water at the time the sediment formed (Figure 10.30).

Even today there is evidence of large quantities of ice just below the surface of Mars. In the mid-latitudes, high-resolution photos from orbit have revealed glaciers covered with dirt and dust. In some cliffs, the ice is observed directly (see Figure 10.30). *Perseverance* is exploring the 45-km-wide crater Jezero, which is a dusty former lakebed and river delta with evidence of large rocks deposited by ancient floods. These glaciers are thought to have formed during warm periods, when the atmospheric pressure was greater and snow and ice could precipitate. They also suggest readily available frozen water that could support future human exploration of the planet.

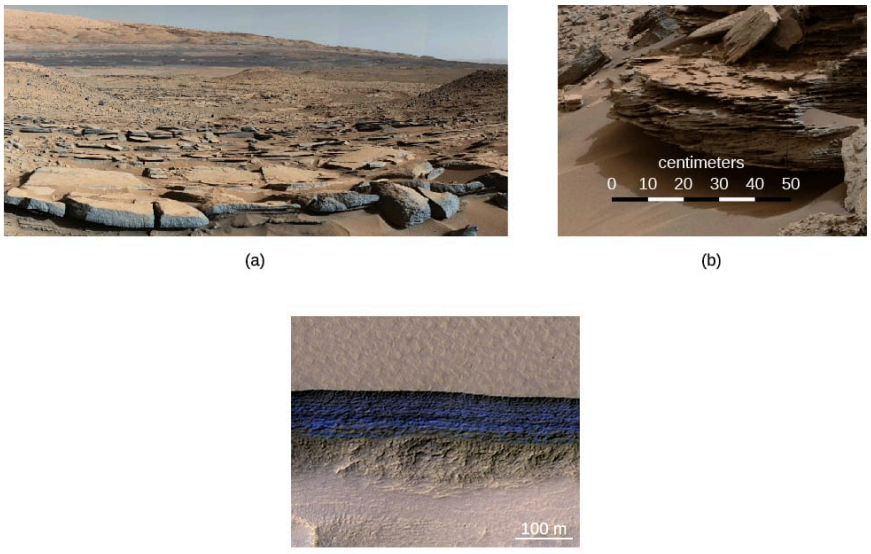


Figure 10.30 : Gale Crater and Underground Ice Deposits. (a) This scene, photographed by the *Curiosity* rover, shows an ancient lakebed of cracked mudstones. (b) Geologists working with the *Curiosity* rover interpret this image of cross-bedded sandstone in Gale crater as evidence of liquid water passing over a loose bed of sediment at the time this rock formed. (c) Ice bands a hundred meters tall are visible in blue in a cliff-face on Mars, suggesting large deposits of frozen water buried just a few meters below the surface. Note that the blue color has been exaggerated in this photo, taken by the Mars Reconnaissance Orbiter spacecraft. (credit a: modification of work by NASA/JPL-Caltech/MSSS; credit b: modification of work by NASA/JPL-Caltech/MSSS; credit c: modification of work by NASA/JPL-Caltech/UA/USGS)

Making Connections: Astronomy and Pseudoscience: The “Face on Mars”

People like human faces. We humans have developed great skill in recognizing people and interpreting facial expressions. We also have a tendency to see faces in many natural formations, from clouds to the man in the Moon. One of the curiosities that emerged from the Viking orbiters’ global mapping of Mars was the discovery of a strangely shaped mesa in the Cydonia region that resembled a human face. Despite later rumors of a cover-up, the “Face on Mars” was, in fact, recognized by Viking scientists and included in one of the early mission press releases. At the low resolution and oblique lighting under which the Viking image was obtained, the mile-wide mesa had something of a Sphinx-like appearance.

Unfortunately, a small band of individuals decided that this formation was an artificial, carved sculpture of a human face placed on Mars by an ancient civilization that thrived there hundreds of thousands of years ago. A band of “true believers” grew around the face and tried to deduce the nature of the “sculptors” who made it. This group also linked the face to a variety of other pseudoscientific phenomena such as crop circles (patterns in fields of grain, mostly in Britain, now known to be the work of pranksters).

Members of this group accused NASA of covering up evidence of intelligent life on Mars, and they received a great deal of help in publicizing their perspective from tabloid media. Some of the believers picketed the Jet Propulsion Laboratory at the time of the failure of the *Mars Observer* spacecraft, circulating stories that the “failure” of the *Mars Observer* was itself a fake, and that its true (secret) mission was to photograph the face.

The high-resolution *Mars Observer* camera (MOC) was reflown on the *Mars Global Surveyor* mission, which arrived at Mars in 1997. On April 5, 1998, in Orbit 220, the MOC obtained an oblique image of the face at a resolution of 4 meters per pixel, a factor-of-10 improvement in resolution over the Viking image. Another image in 2001 had even higher resolution. Immediately released by NASA, the new images showed a low mesa-like hill cut crossways by several roughly linear ridges and depressions, which were misidentified in the 1976 photo as the eyes and mouth of a face. Only with an enormous dose of imagination can any resemblance to a face be seen in the new images, demonstrating how dramatically our interpretation of geology can change with large improvements in resolution. The original and the higher resolution images can be seen in Figure 10.31.

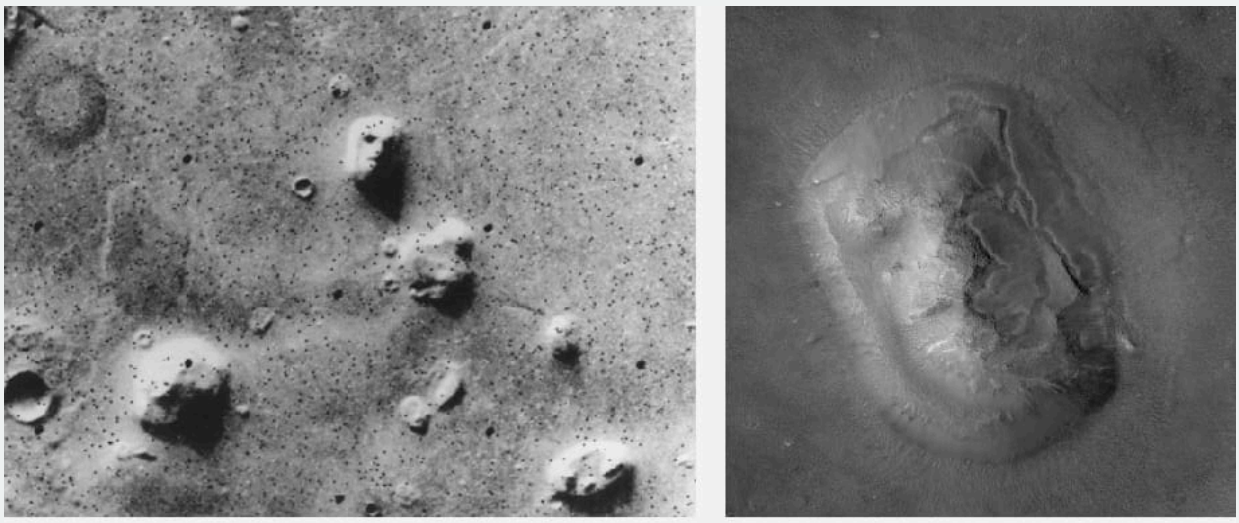


Figure 10.31 : Face on Mars. The so-called “Face on Mars” is seen (a) in low resolution from Viking (the “face” is in the upper part of the picture) and (b) with 20 times better resolution from the *Mars Global Surveyor*. (credit a: modification of work NASA/JPL; credit b: modification of work by NASA/JPL/MSSS)

After 20 years of promoting pseudoscientific interpretations and various conspiracy theories, can the “Face on Mars” believers now accept reality? Unfortunately, it does not seem so. They have accused NASA of faking the new picture. They also suggest that the secret mission of the *Mars Observer* included a nuclear bomb used to destroy the face before it could be photographed in greater detail by the *Mars Global Surveyor*.

Space scientists find these suggestions incredible. NASA is spending increasing sums for research on life in the universe, and a major objective of current and upcoming Mars missions is to search for evidence of past microbial life on Mars. Conclusive evidence of extraterrestrial life would be one of the great discoveries of science and incidentally might well lead to increased funding for NASA. The idea that NASA or other government agencies would (or could) mount a conspiracy to suppress such welcome evidence is truly bizarre.

Alas, the "Face on Mars" story is only one example of a whole series of conspiracy theories that are kept before the public by dedicated believers, by people out to make a fast buck, and by irresponsible media attention. Others include the "urban legend" that the Air Force has the bodies of extraterrestrials at a secret base, the widely circulated report that UFOs crashed near Roswell, New Mexico (actually it was a balloon carrying scientific instruments to find evidence of Soviet nuclear tests), or the notion that alien astronauts helped build the Egyptian pyramids and many other ancient monuments because our ancestors were too stupid to do it alone.

In response to the increase in publicity given to these "fiction science" ideas, a group of scientists, educators, scholars, and magicians (who know a good hoax when they see one) have formed the Committee for Skeptical Inquiry. Two of the original authors of your book are active on the committee. For more information about its work delving into the rational explanations for paranormal claims, see their excellent magazine, *The Skeptical Inquirer*, or check out their website at https://skepticalinquirer.org/.

Climate Change on Mars

The evidence about ancient rivers and lakes of water on Mars discussed so far suggests that, billions of years ago, martian temperatures must have been warmer and the atmosphere must have been more substantial than it is today. But what could have changed the climate on Mars so dramatically?

We presume that, like Earth and Venus, Mars probably formed with a higher surface temperature thanks to the greenhouse effect. But Mars is a smaller planet, and its lower gravity means that atmospheric gases could escape more easily than from Earth and Venus. As more and more of the atmosphere escaped into space, the temperature on the surface gradually fell.

Eventually Mars became so cold that most of the water froze out of the atmosphere, further reducing its ability to retain heat. The planet experienced a sort of *runaway refrigerator effect*, just the opposite of the runaway greenhouse effect that occurred on Venus. Probably, this loss of atmosphere took place within less than a billion years after Mars formed. The result is the cold, dry Mars we see today.

Conditions a few meters below the martian surface, however, may be much different. There, liquid water (especially salty water) might persist, kept warm by the internal heat of Mars or the insulating layers solid and rock. Even on the surface, there may be ways to change the martian atmosphere temporarily.

Mars is likely to experience long-term climate cycles, which may be caused by the changing orbit and tilt of the planet. At times, one or both of the polar caps might melt, releasing a great deal of water vapor into the atmosphere. Perhaps an occasional impact by a comet might produce a temporary atmosphere that is thick enough to permit liquid water on the surface for a few weeks or months. Some have even suggested that future technology might allow us to *terraform* Mars—that is, to engineer its atmosphere and climate in ways that might make the planet more hospitable for long-term human habitation.

The Search for Life on Mars

If there was running water on Mars in the past, perhaps there was life as well. Could life, in some form, remain in the martian soil today? Testing this possibility, however unlikely, was one of the primary objectives of the Viking landers in 1976. These landers carried miniature biological laboratories to test for microorganisms in the martian soil. Martian soil was scooped up by the spacecraft's long arm and placed into the experimental chambers, where it was isolated and incubated in contact with a variety of gases, radioactive isotopes, and nutrients to see what would happen. The experiments looked for evidence of *respiration* by living animals, *absorption* of *nutrients* offered to organisms that might be present, and an *exchange of gases* between the soil and its surroundings for any reason whatsoever. A fourth instrument pulverized the soil and analyzed it carefully to determine what organic (carbon-bearing) material it contained.

The Viking experiments were so sensitive that, had one of the spacecraft landed anywhere on Earth (with the possible exception

of Antarctica), it would easily have detected life. But, to the disappointment of many scientists and members of the public, no life was detected on Mars. The soil tests for absorption of nutrients and gas exchange did show some activity, but this was most likely caused by chemical reactions that began as water was added to the soil and had nothing to do with life. In fact, these experiments showed that martian soil seems much more chemically active than terrestrial soils because of its exposure to solar ultraviolet radiation (since Mars has no ozone layer).

The organic chemistry experiment showed no trace of organic material, which is apparently destroyed on the martian surface by the sterilizing effect of this ultraviolet light. While the possibility of life on the surface has not been eliminated, most experts consider it negligible. Although Mars has the most earthlike environment of any planet in the solar system, the sad fact is that nobody seems to be home today, at least on the surface.

However, there is no reason to think that life could not have begun on Mars about 4 billion years ago, at the same time it started on Earth. The two planets had very similar surface conditions then. Thus, the attention of scientists has shifted to the search for *fossil* life on Mars. One of the primary questions to be addressed by future spacecraft is whether Mars once supported its own life forms and, if so, how this martian life compared with that on our own planet. Future missions will include the return of martian samples selected from sedimentary rocks at sites that once held water and thus perhaps ancient life. The most powerful searches for martian life (past or present) will thus be carried out in our laboratories here on Earth.

Making Connections: Planetary Protection

When scientists begin to search for life on another planet, they must make sure that we do not contaminate the other world with life carried from Earth. At the very beginning of spacecraft exploration on Mars, an international agreement specified that all landers were to be carefully sterilized to avoid accidentally transplanting terrestrial microbes to Mars. In the case of Viking, we know the sterilization was successful. Viking's failure to detect martian organisms also implies that these experiments did not detect hitchhiking terrestrial microbes.

As we have learned more about the harsh conditions on the martian surface, the sterilization requirements have been somewhat relaxed. It is evident that no terrestrial microbes could grow on the martian surface, with its low temperature, absence of water, and intense ultraviolet radiation. Microbes from Earth might survive in a dormant, dried state, but they cannot grow and proliferate on Mars.

The problem of contaminating Mars will become more serious, however, as we begin to search for life below the surface, where temperatures are higher and no ultraviolet light penetrates. The situation will be even more daunting if we consider human flights to Mars. Any humans will carry with them a multitude of terrestrial microbes of all kinds, and it is hard to imagine how we can effectively keep the two biospheres isolated from each other if Mars has indigenous life. Perhaps the best situation could be one in which the two life-forms are so different that each is effectively invisible to the other—not recognized on a chemical level as living or as potential food.

The most immediate issue of public concern is not with the contamination of Mars but with any dangers associated with returning Mars samples to Earth. NASA is committed to the complete biological isolation of returned samples until they are demonstrated to be safe. Even though the chances of contamination are extremely low, it is better to be safe than sorry.

Most likely there is no danger, even if there is life on Mars and alien microbes hitch a ride to Earth inside some of the returned samples. In fact, Mars is sending samples to Earth all the time in the form of the Mars meteorites. Since some of these microbes (if they exist) could probably survive the trip to Earth inside their rocky home, we may have been exposed many times over to martian microbes. Either they do not interact with our terrestrial life, or in effect our planet has already been inoculated against such alien bugs.

Link to Learning

More than any other planet, Mars has inspired science fiction writers over the years. You can find scientifically reasonable stories about Mars in a subject index of such stories online. If you click on Mars as a topic, you will find stories by a number of space scientists, including William Hartmann, Geoffrey Landis, and Ludek Pesek.

Answers

Exercise 8.3.1

The Greenland ice sheet has about 2.85 times as much ice as in the polar ice caps on Mars. They are about the same to the nearest power of 10.

This page titled 8.3: Water and Life on Mars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 10.6: Water and Life on Mars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

8.4: Divergent Planetary Evolution

Learning Objectives

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

- Compare the planetary evolution of Venus, Earth, and Mars

Venus, Mars, and our own planet Earth form a remarkably diverse triad of worlds. Although all three orbit in roughly the same inner zone around the Sun and all apparently started with about the same chemical mix of silicates and metals, their evolutionary paths have diverged. As a result, Venus became hot and dry, Mars became cold and dry, and only Earth ended up with what we consider a hospitable climate.

We have discussed the runaway greenhouse effect on Venus and the runaway refrigerator effect on Mars, but we do not understand exactly what started these two planets down these separate evolutionary paths. Was Earth ever in danger of a similar fate? Or might it still be diverted onto one of these paths, perhaps due to stress on the atmosphere generated by human pollutants? One of the reasons for studying Venus and Mars is to seek insight into these questions.

Some people have even suggested that if we understood the evolution of Mars and Venus better, we could possibly reverse their evolution and restore more earthlike environments. While it seems unlikely that humans could ever make either Mars or Venus into a replica of Earth, considering such possibilities is a useful part of our more general quest to understand the delicate environmental balance that distinguishes our planet from its two neighbors. In Cosmic Samples and the Origin of the Solar System, we return to the comparative study of the terrestrial planets and their divergent evolutionary histories.

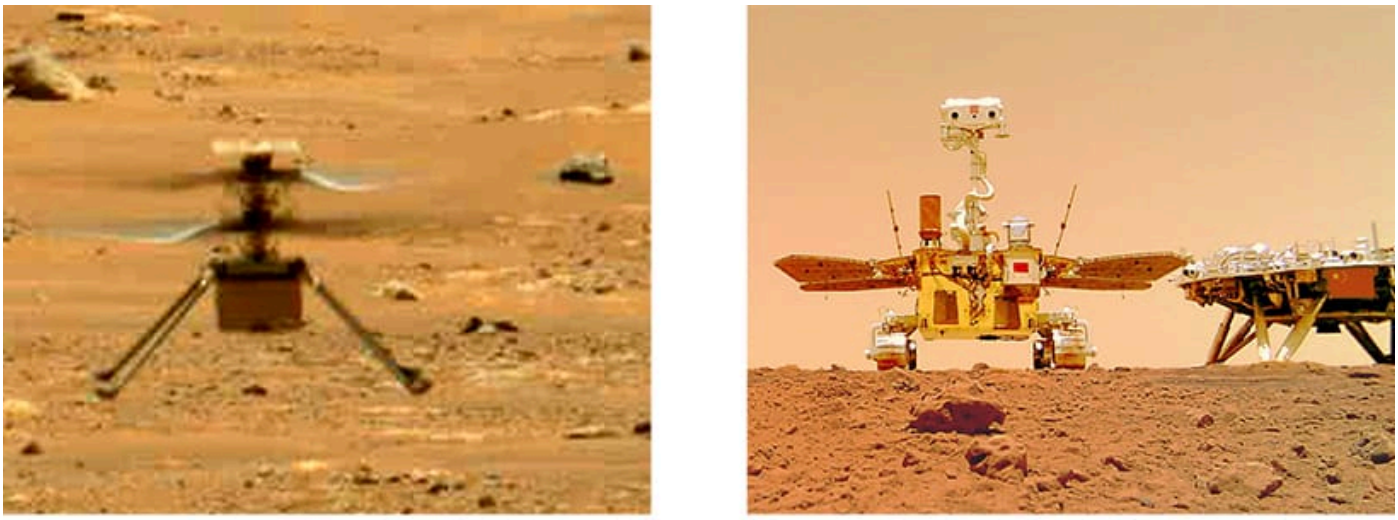


Figure 10.32 : New Vehicles on Mars. (a) The *Ingenuity* Helicopter flies above the martian surface. (b) The Chinese *Zhurong* rover, deployed in June 2021, is next to the *Tianwen-1* lander; the “group portrait” was taken by a wireless camera placed by the rover. (credit a: modification of work by NASA/JPL; credit b: modification of work by the China National Space Administration)

This page titled 8.4: Divergent Planetary Evolution is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 10.7: Divergent Planetary Evolution by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

9: JOVIAN PLANETS

9.1: Exploring the Outer Planets
9.2: The Giant Planets
9.3: The Gailean Moons of Jupiter
9.4: Titan and Triton

This page titled 9: Jovian Planets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

9.1: Exploring the Outer Planets

Learning Objectives

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

- Provide an overview of the composition of the giant planets
- Chronicle the robotic exploration of the outer solar system
- Summarize the missions sent to orbit the gas giants

The giant planets hold most of the mass in our planetary system. Jupiter alone exceeds the mass of all the other planets combined (Figure 11.2). The material available to build these planets can be divided into three classes by what they are made of: “gases,” “ices,” and “rocks” (see Table 11.1). The “gases” are primarily hydrogen and helium, the most abundant elements in the universe. The way it is used here, the term “ices” refers to composition only and not whether a substance is actually in a solid state. “Ices” means compounds that form from the next most abundant elements: oxygen, carbon, and nitrogen. Common ices are water, methane, and ammonia, but ices may also include carbon monoxide, carbon dioxide, and others. “Rocks” are even less abundant than ices, and include everything else: magnesium, silicon, iron, and so on.

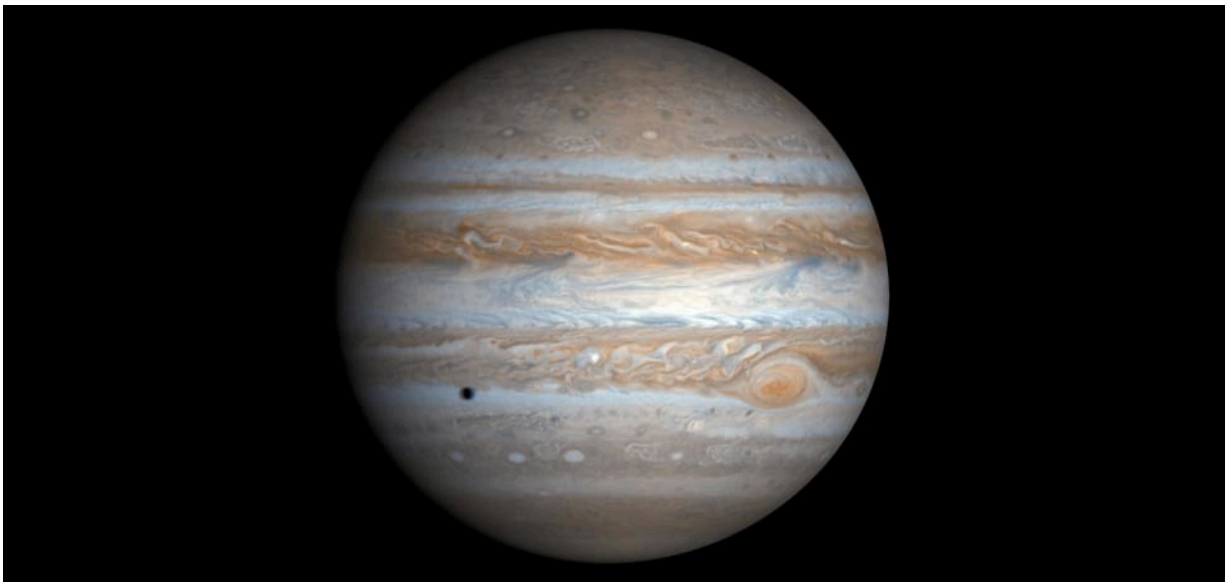


Figure 11.2 : Jupiter. The Cassini spacecraft imaged Jupiter on its way to Saturn in 2012. The giant storm system called the Great Red Spot is visible to the lower right. The dark spot to the lower left is the shadow of Jupiter’s moon Europa. (credit: modification of work by NASA/JPL)

Table 11.1: Abundances in the Outer Solar System

Type of Material	Name	Approximate % (by Mass)
Gas	Hydrogen (H_2)	75
Gas	Helium (He)	24
Ice	Water (H_2O)	0.6
Ice	Methane (CH_4)	0.4
Ice	Ammonia (NH_3)	0.1
Rock	Magnesium (Mg), iron (Fe), silicon (Si)	0.3

In the outer solar system, gases dominate the two largest planets, Jupiter and Saturn, hence their nickname “gas giants.” Uranus

and Neptune are sometimes called “ice giants” because their interiors contain far more of the “ice” component than their larger cousins. The chemistry for all four giant planet atmospheres is dominated by hydrogen. This hydrogen caused the chemistry of the outer solar system to become *reducing,* meaning that other elements tend to combine with hydrogen first. In the early solar system, most of the oxygen combined with hydrogen to make H_2O and was thus unavailable to form the kinds of oxidized compounds with other elements that are more familiar to us in the inner solar system (such as CO_2). As a result, the compounds detected in the atmosphere of the giant planets are mostly hydrogen-based gases such as methane (CH_4) and ammonia (NH_3), or more complex hydrocarbons (combinations of hydrogen and carbon) such as ethane (C_2H_6) and acetylene (C_2H_2).

Exploration of the Outer Solar System So Far

Eight spacecraft, seven from the United States and one from Europe, have penetrated beyond the asteroid belt into the realm of the giants. Table 11.2 summarizes the spacecraft missions to the outer solar system.

Table 11.2: Missions to the Giant Planets

Planet	Spacecraft[1]	Encounter Date	Type
Jupiter	Pioneer 10	December 1973	Flyby
	Pioneer 11	December 1974	Flyby
	Voyager 1	March 1979	Flyby
	Voyager 2	July 1979	Flyby
	Ulysses	February 1992	Flyby during gravity assist
	Galileo	December 1995	Orbiter and probe
	Cassini	December 2002	Flyby
	New Horizons	February 2007	Flyby during gravity assist
	Juno	July 2016	Orbiter
Saturn	Pioneer 11	September 1979	Flyby
	Voyager 1	November 1980	Flyby
	Voyager 2	August 1981	Flyby
	Cassini	July 2004	Orbiter
Uranus	Voyager 2	January 1986	Flyby
Neptune	Voyager 2	August 1989	Flyby

The challenges of exploring so far away from Earth are considerable. Flight times to the giant planets are measured in years to decades, rather than the months required to reach Venus or Mars. Even at the speed of light, messages take hours to pass between Earth and the spacecraft. If a problem develops near Saturn, for example, a wait of hours for the alarm to reach Earth and for instructions to be routed back to the spacecraft could spell disaster. Spacecraft to the outer solar system must therefore be highly reliable and capable of a greater degree of independence and autonomy. Outer solar system missions also must carry their own power sources since the Sun is too far away to provide enough energy, or else they must have very large arrays of solar cells. Heaters are required to keep instruments at proper operating temperatures, and spacecraft must have radio transmitters powerful enough to send their data to receivers on distant Earth.

The first spacecraft to investigate the regions past Mars were the NASA Pioneers 10 and 11, launched in 1972 and 1973 as pathfinders to Jupiter. One of their main objectives was simply to determine whether a spacecraft could actually navigate through the belt of asteroids that lies beyond Mars without getting destroyed by collisions with asteroidal dust. Another objective was to measure the radiation hazards in the *magnetosphere* (or zone of magnetic influence) of Jupiter. Both spacecraft passed through the asteroid belt without incident, but the energetic particles in Jupiter’s magnetic field nearly wiped out their electronics, providing information necessary for the safe design of subsequent missions.

Pioneer 10 flew past Jupiter in 1973, after which it sped outward toward the limits of the solar system. Pioneer 11 undertook a more

ambitious program, using the gravity of Jupiter to aim for Saturn, which it reached in 1979. The twin Voyager spacecraft launched the next wave of outer planet exploration in 1977. Voyagers 1 and 2 each carried 11 scientific instruments, including cameras and spectrometers, as well as devices to measure the characteristics of planetary magnetospheres. Since they kept going outward after their planetary encounters, these are now the most distant spacecraft ever launched by humanity.

Voyager 1 reached Jupiter in 1979 and used a gravity assist from that planet to take it on to Saturn in 1980. Voyager 2 arrived at Jupiter four months later, but then followed a different path to visit all the outer planets, reaching Saturn in 1981, Uranus in 1986, and Neptune in 1989. This trajectory was made possible by the approximate alignment of the four giant planets on the same side of the Sun. About once every 175 years, these planets are in such a position, and it allows a single spacecraft to visit them all by using gravity-assisted flybys to adjust course for each subsequent encounter; such a maneuver has been nicknamed a “Grand Tour” by astronomers.

Link to Learning

The Jet Propulsion Laboratory has a nice video called Voyager: The Grand Tour that describes the Voyager mission and what it found.

Making Connections: Engineering and Space Science: Teaching an Old Spacecraft New Tricks

By the time Voyager 2 arrived at Neptune in 1989, 12 years after its launch, the spacecraft was beginning to show signs of old age. The arm on which the camera and other instruments were located was “arthritic”: it could no longer move easily in all directions. The communications system was “hard of hearing”: part of its radio receiver had stopped working. The “brains” had significant “memory loss”: some of the onboard computer memory had failed. And the whole spacecraft was beginning to run out of energy: its generators had begun showing serious signs of wear.

To make things even more of a challenge, Voyager’s mission at Neptune was in many ways the most difficult of all four flybys. For example, since sunlight at Neptune is 900 times weaker than at Earth, the onboard camera had to take much longer exposures in this light-starved environment. This was a nontrivial requirement, given that the spacecraft was hurtling by Neptune at ten times the speed of a rifle bullet.

The solution was to swivel the camera backward at exactly the rate that would compensate for the forward motion of the spacecraft. Engineers had to preprogram the ship’s computer to execute an incredibly complex series of maneuvers for each image. The beautiful Voyager images of Neptune are a testament to the ingenuity of spacecraft engineers.

The sheer distance of the craft from its controllers on Earth was yet another challenge. Voyager 2 received instructions and sent back its data via on-board radio transmitter. The distance from Earth to Neptune is about 4.8 billion kilometers. Over this vast distance, the power that reached us from Voyager 2 at Neptune was approximately 10^{-16} watts, or 20 billion times less power than it takes to operate a digital watch. Thirty-eight different antennas on four continents were used by NASA to collect the faint signals from the spacecraft and decode the precious information about Neptune that they contained.

Enter the Orbiters: Galileo, Cassini, and Juno

The Pioneer and Voyager missions were flybys of the giant planets: they each produced only quick looks before the spacecraft sped onward. For more detailed studies of these worlds, we require spacecraft that can go into orbit around a planet. For Jupiter and Saturn, these orbiters were the Galileo, Cassini, and Juno spacecraft. To date, no orbiter missions have been started for Uranus and Neptune, although planetary scientists have expressed keen interest.

The Galileo spacecraft was launched toward Jupiter in 1989 and arrived in 1995. Galileo began its investigations by deploying an entry probe into Jupiter, for the first direct studies of the planet’s outer atmospheric layers.

The probe plunged at a shallow angle into Jupiter’s atmosphere, traveling at a speed of 50 kilometers *per second*—that’s fast enough to fly from New York to San Francisco in 100 seconds! This was the highest speed at which any probe has so far entered the atmosphere of a planet, and it put great demands on the heat shield protecting it. The high entry speed was a result of acceleration by the strong gravitational attraction of Jupiter.

Atmospheric friction slowed the probe within 2 minutes, producing temperatures at the front of its heat shield as high as 15,000 °C. As the probe's speed dropped to 2500 kilometers per hour, the remains of the glowing heat shield were jettisoned, and a parachute was deployed to lower the instrumented probe spacecraft more gently into the atmosphere (Figure 11.3). The data from the probe instruments were relayed to Earth via the main Galileo spacecraft.



(a) (b)

Figure 11.3 : Galileo Probe Falling into Jupiter and Juno Image of Jupiter's South Pole. (a) This artist's depiction shows the Galileo probe descending into the clouds via parachute just after the protective heat shield separated. The probe made its measurements of Jupiter's atmosphere on December 7, 1995. (b) This Juno image, taken in 2017 from about 100,000 kilometers above the cloudtops, shows the south polar region of Jupiter with its dramatic complex of storms and clouds. The enhanced-color image was processed for NASA/JPL by citizen scientist John Landino. (credit a: modification of work by NASA/Ames Research Center; credit b: modification of work by NASA/JPL-Caltech/SwRI/MSSS/John Landino)

The probe continued to operate for an hour, descending 200 kilometers into the atmosphere. A few minutes later the polyester parachute melted, and within a few hours the main aluminum and titanium structure of the probe vaporized to become a part of Jupiter itself. About 2 hours after receipt of the final probe data, the main spacecraft fired its retro-rockets so it could be captured into orbit around the planet, where its primary objectives were to study Jupiter's large and often puzzling moons.

The Cassini mission to Saturn (Figure 11.4), a cooperative venture between NASA and the European Space Agency, was similar to Galileo in its two-fold approach. Launched in 1997, Cassini arrived in 2004 and went into orbit around Saturn, beginning extensive studies of its rings and moons, as well as the planet itself. In January 2005, Cassini deployed an entry probe into the atmosphere of Saturn's large moon, Titan, where it successfully landed on the surface. (We'll discuss the probe and what it found in the chapter on Rings, Moons, and Pluto.)

The Voyager and Galileo missions to Jupiter were primarily designed to study the moons and the atmosphere of the planet. The next NASA mission, an orbiter called Juno, arrived at Jupiter in July 2016. In order to meet its objectives of studying the jovian magnetosphere, it has a very elongated (eccentric) 55-day orbit, that takes it from 4 thousand kilometers above the cloud tops out to 76 thousand kilometers. The orbit takes the craft over Jupiter's poles, giving us remarkable close-ups of the polar regions (previous spacecraft viewed the planet from lower latitudes).

Juno was originally designed without a camera, but fortunately scientists rectified this omission, adding a simple downward-looking color camera to use during close passes by Jupiter. Recognizing the value of such images, both scientific and artistic, it was decided to post the raw images and encourage "citizen scientists" to process them. The product has been many dramatic, brightly colored views of Jupiter, such as Figure 11.3.



Figure 11.4 : Earth as Seen from Saturn. This popular Cassini image shows Earth as a tiny dot (marked with an arrow) seen below Saturn's rings. It was taken in July 2013, when Saturn was 1.4 billion kilometers from Earth. (credit: modification of work by NASA/JPL-Caltech/Space Science Institute)

Footnotes

- 1Both the Ulysses and the New Horizons spacecraft (designed to study the Sun and Pluto, respectively) flew past Jupiter for a gravity boost (gaining energy by "stealing" a little bit from the giant planet's rotation).

This page titled 9.1: Exploring the Outer Planets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 11.2: Exploring the Outer Planets by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

9.2: The Giant Planets

Learning Objectives

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

- Describe the basic physical characteristics, general appearance, and rotation of the giant planets
- Describe the composition and structure of Jupiter, Saturn, Uranus, and Neptune
- Compare and contrast the internal heat sources of the giant planets
- Describe the discovery and characteristics of the giant planets' magnetic fields

Let us now examine the four giant (or *jovian*) planets in more detail. Our approach is not just to catalog their characteristics, but to compare them with each other, noting their similarities and differences and attempting to relate their properties to their differing masses and distances from the Sun.

Basic Characteristics

The giant planets are very far from the Sun. Jupiter is more than five times farther from the Sun than Earth's distance (5 AU), and takes just under 12 years to circle the Sun. Saturn is about twice as far away as Jupiter (almost 10 AU) and takes nearly 30 years to complete one orbit. Uranus orbits at 19 AU with a period of 84 years, while Neptune, at 30 AU, requires 165 years for each circuit of the Sun. These long timescales make it difficult for us short-lived humans to study seasonal change on the outer planets.

Jupiter and Saturn have many similarities in composition and internal structure, although Jupiter is nearly four times more massive. Uranus and Neptune are smaller and differ in composition and internal structure from their large siblings. Some of the main properties of these four planets are summarized in Table 11.3.

Table 11.3: Basic Properties of the Jovian Planets

Planet	Distance (AU)	Period (years)	Diameter (km)	Mass (Earth = 1)	Density (g/cm^3)	Rotation (hours)
Jupiter	5.2	11.9	142,800	318	1.3	9.9
Saturn	9.5	29.5	120,540	95	0.7	10.7
Uranus	19.2	84.1	51,200	14	1.3	17.2
Neptune	30.0	164.8	49,500	17	1.6	16.1

Jupiter, the giant among giants, has enough mass to make 318 Earths. Its diameter is about 11 times that of Earth (and about one tenth that of the Sun). Jupiter's average density is 1.3 g/cm^3, much lower than that of any of the terrestrial planets. (Recall that water has a density of 1 g/cm^3.) Jupiter's material is spread out over a volume so large that about 1,300 Earths could fit within it.

Saturn's mass is 95 times that of Earth, and its average density is only 0.7 g/cm^3—the lowest of any planet. Since this is less than the density of water, Saturn would be light enough to float.

Uranus and Neptune each have a mass about 15 times that of Earth and, hence, are only 5% as massive as Jupiter. Their densities of 1.3 g/cm^3 and 1.6 g/cm^3, respectively, are much higher than that of Saturn. This is one piece of evidence that tells us that their composition must differ fundamentally from the gas giants. When astronomers began to discover other planetary systems (exoplanets), we found that planets the size of Uranus and Neptune are common, and that there are even more exoplanets intermediate in size between Earth and these ice giants, a type of planet not found in our solar system.

Appearance and Rotation

When we look at the planets, we see only their atmospheres, composed primarily of hydrogen and helium gas (see Figure 11.1). The uppermost clouds of Jupiter and Saturn, the part we see when looking down at these planets from above, are composed of ammonia crystals. On Neptune, the upper clouds are made of methane. On Uranus, we see no obvious cloud layer at all, but only a deep and featureless haze.

Seen through a telescope, Jupiter is a colorful and dynamic planet. Distinct details in its cloud patterns allow us to determine the rotation rate of its atmosphere at the cloud level, although such atmosphere rotation may have little to do with the spin of the underlying planet. Much more fundamental is the rotation of the mantle and core; these can be determined by periodic variations in radio waves coming from Jupiter, which are controlled by its magnetic field. Since the magnetic field (which we will discuss below) originates deep inside the planet, it shares the rotation of the interior. The rotation period we measure in this way is 9 hours 56 minutes, which gives Jupiter the shortest "day" of any planet. In the same way, we can measure that the underlying rotation period of Saturn is 10 hours 40 minutes. Uranus and Neptune have slightly longer rotation periods of about 17 hours, also determined from the rotation of their magnetic fields.

Link to Learning

A brief video made from Hubble Space Telescope photos shows the rotation of Jupiter with its many atmospheric features.

Remember that Earth and Mars have seasons because their spin axes, instead of "standing up straight," are tilted relative to the orbital plane of the solar system. This means that as Earth revolves around the Sun, sometimes one hemisphere and sometimes the other "leans into" the Sun.

What are the seasons like for the giant planets? The spin axis of Jupiter is tilted by only 3°, so there are no seasons to speak of. Saturn, however, does have seasons, since its spin axis is inclined at 27° to the perpendicular to its orbit. Neptune has about the same tilt as Saturn (29°); therefore, it experiences similar seasons (only more slowly). The strangest seasons of all are on Uranus, which has a spin axis tilted by 98° with respect to the north direction. Practically speaking, we can say that Uranus orbits on its side, and its ring and moon system follow along, orbiting about Uranus' equator (Figure 11.5).

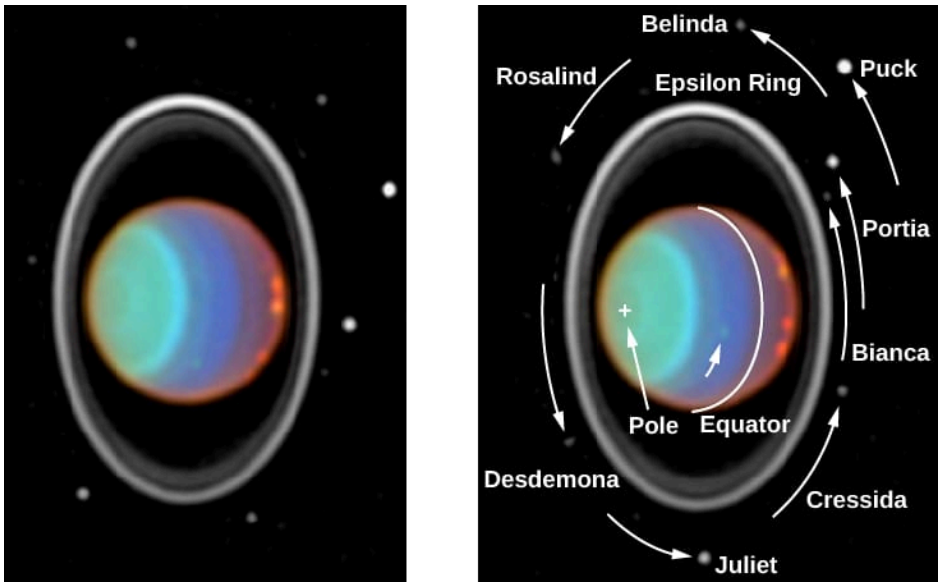


Figure 11.5 : Infrared Image of Uranus. The infrared camera on the Hubble Space Telescope took these false-color images of the planet Uranus, its ring system, and moons in 1997. The south pole of the planet (marked with a "+" on the right image) faces the Sun; its green color shows a strong local haze. The two images were taken 90 minutes apart, and during that time the five reddish clouds can be seen to rotate around the parallel to the equator. The rings (which are very faint in the visible light, but prominent in infrared) and eight moons can be seen around the equator. This was the "bull's eye" arrangement that Voyager saw as it approached Uranus in 1986. (credit: modification of work by Erich Karkoschka (University of Arizona), and NASA/ESA)

We don't know what caused Uranus to be tipped over like this, but one possibility is a collision with a large planetary body when

our system was first forming. Whatever the cause, this unusual tilt creates dramatic seasons. When Voyager 2 arrived at Uranus, its south pole was facing directly into the Sun. The southern hemisphere was experiencing a 21-year sunlit summer, while during that same period the northern hemisphere was plunged into darkness. For the next 21-year season, the Sun shines on Uranus' equator, and both hemispheres go through cycles of light and dark as the planet rotates (Figure 11.6). Then there are 21 years of an illuminated northern hemisphere and a dark southern hemisphere. After that the pattern of alternating day and night repeats.

Just as on Earth, the seasons are even more extreme at the poles. If you were to install a floating platform at the south pole of Uranus, for example, it would experience 42 years of light and 42 years of darkness. Any future astronauts crazy enough to set up camp there could spend most of their lives without ever seeing the Sun.

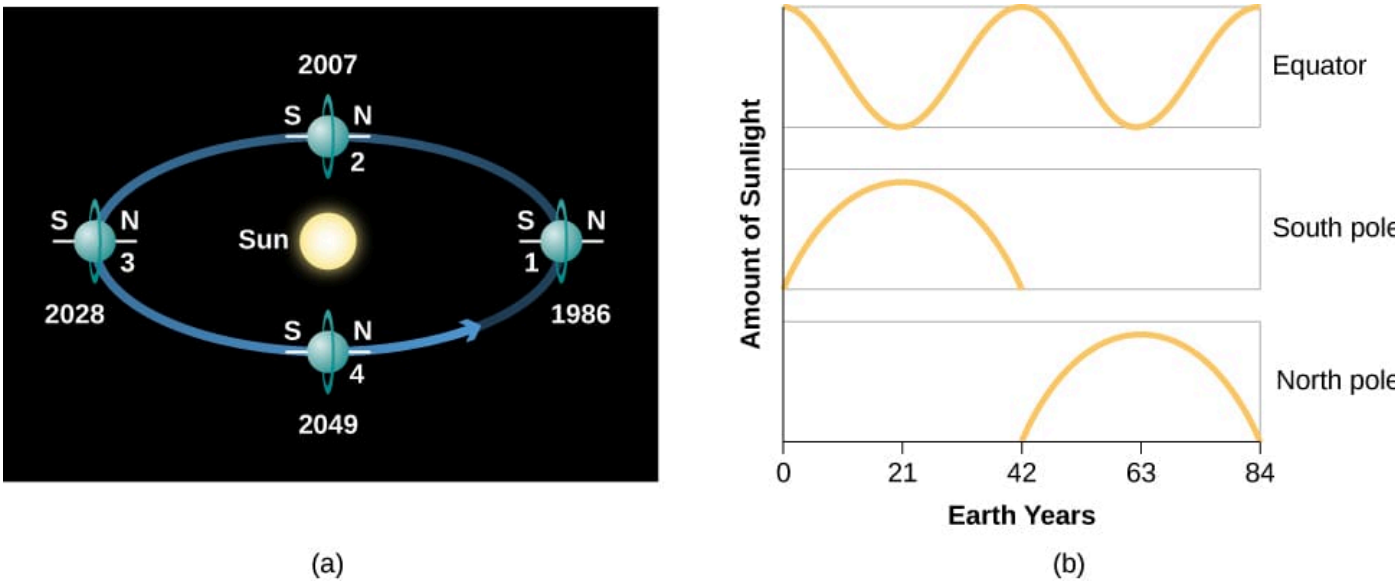


Figure 11.6 : Strange Seasons on Uranus. (a) This diagram shows the orbit of Uranus as seen from above. At the time Voyager 2 arrived (position 1), the South Pole was facing the Sun. As we move counterclockwise in the diagram, we see the planet 21 years later at each step. (b) This graph compares the amount of sunlight seen at the poles and the equator of Uranus over the course of its 84-year revolution around the Sun.

Composition and Structure

Although we cannot see into these planets, astronomers are confident that the interiors of Jupiter and Saturn are composed primarily of hydrogen and helium. Of course, these gases have been measured only in their atmosphere, but calculations first carried out more than 50 years ago showed that these two light gases are the only possible materials out of which a planet with the observed masses and densities of Jupiter and Saturn could be constructed.

The deep internal structures of these two planets are difficult to predict. This is mainly because these planets are so big that the hydrogen and helium in their centers become tremendously compressed and behave in ways that these gases can never behave on Earth. The best theoretical models we have of Jupiter's structure predict a central pressure greater than 100 million bars and a central density of about 31 g/cm^3. (By contrast, Earth's core has a central pressure of 4 million bars and a central density of 17 g/cm^3.)

At the pressures inside the giant planets, familiar materials can take on strange forms. A few thousand kilometers below the visible clouds of Jupiter and Saturn, pressures become so great that hydrogen changes from a gaseous to a liquid state. Still deeper, this liquid hydrogen is further compressed and begins to act like a metal, something it never does on Earth. (In a metal, electrons are not firmly attached to their parent nuclei but can wander around. This is why metals are such good conductors of electricity.) On Jupiter, the greater part of the interior is liquid metallic hydrogen.

Because Saturn is less massive, it has only a small volume of metallic hydrogen, but most of its interior is liquid. Uranus and Neptune are too small to reach internal pressures sufficient to liquefy hydrogen. We will return to the discussion of the metallic hydrogen layers when we examine the magnetic fields of the giant planets.

Each of these planets has a core composed of heavier materials, as demonstrated by detailed analyses of their gravitational fields.

Presumably these cores are the original rock-and-ice bodies that formed before the capture of gas from the surrounding nebula. The cores exist at pressures of tens of millions of bars. While scientists speak of the giant planet cores being composed of rock and ice, we can be sure that neither rock nor ice assumes any familiar forms at such pressures and temperatures. Remember that what is really meant by “rock” is any material made up primarily of iron, silicon, and oxygen, while the term “ice” in this chapter denotes materials composed primarily of the elements carbon, nitrogen, and oxygen in combination with hydrogen.

Figure 11.7 illustrates the likely interior structures of the four jovian planets. It appears that all four have similar cores of rock and ice. On Jupiter and Saturn, the cores constitute only a few percent of the total mass, consistent with the initial composition of raw materials shown in Table 11.1. However, most of the mass of Uranus and Neptune resides in these cores, demonstrating that the two outer planets were unable to attract massive quantities of hydrogen and helium when they were first forming.

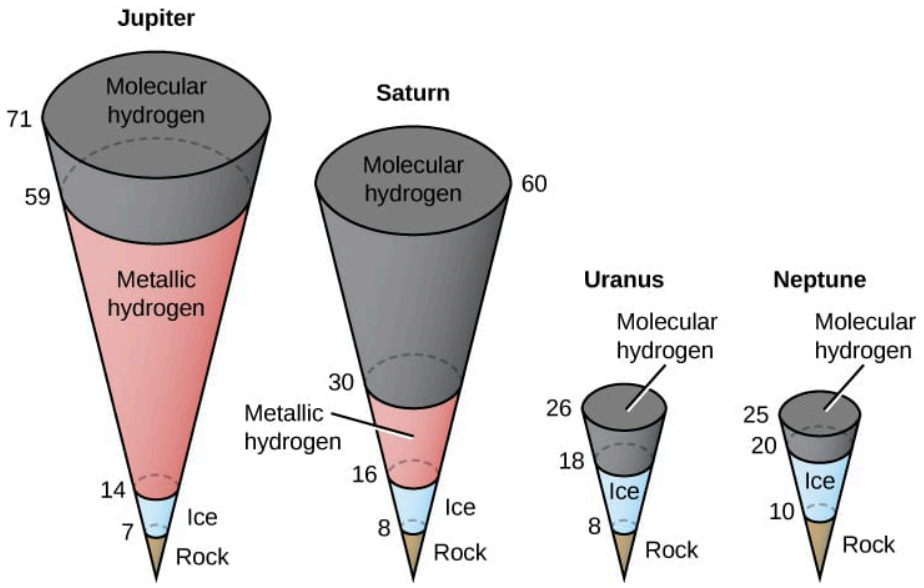


Figure 11.7 : Internal Structures of the Jovian Planets. Jupiter and Saturn are composed primarily of hydrogen and helium (but hydrogen dominates), but Uranus and Neptune consist in large part of compounds of carbon, nitrogen, and oxygen. (The diagrams are drawn to scale; numbers show radii in thousands of kilometers.)

Internal Heat Sources

Because of their large sizes, all the giant planets were strongly heated during their formation by the collapse of surrounding material onto their cores. Jupiter, being the largest, was the hottest. Some of this primordial heat can still remain inside such large planets. In addition, it is possible for giant, largely gaseous planets to generate heat after formation by slowly contracting. (With so large a mass, even a minuscule amount of shrinking can generate significant heat.) The effect of these internal energy sources is to raise the temperatures in the interiors and atmospheres of the planets higher than we would expect from the heating effect of the Sun alone.

Jupiter has the largest internal energy source, amounting to 4×10^{17} watts; that is, it is heated from inside with energy equivalent to 4 million billion 100-watt lightbulbs. This energy is about the same as the total solar energy absorbed by Jupiter. The atmosphere of Jupiter is therefore something of a cross between a normal planetary atmosphere (like Earth’s), which obtains most of its energy from the Sun, and the atmosphere of a star, which is entirely heated by an internal energy source. Most of the internal energy of Jupiter is primordial heat, left over from the formation of the planet 4.5 billion years ago.

Saturn has an internal energy source about half as large as that of Jupiter, which means (since its mass is only about one quarter as great) that it is producing twice as much energy per kilogram of material as does Jupiter. Since Saturn is expected to have much less primordial heat, there must be another source at work generating most of this 2×10^{17} watts of power. This source is the separation of helium from hydrogen in Saturn’s interior. In the liquid hydrogen mantle, the heavier helium forms droplets that sink toward the core, releasing gravitational energy. In effect, Saturn is still differentiating—letting lighter material rise and heavier material fall.

Uranus and Neptune are different. Neptune has a small internal energy source, while Uranus does not emit a measurable amount of internal heat. As a result, these two planets have almost the same atmospheric temperature, in spite of Neptune’s greater

distance from the Sun. No one knows why these two planets differ in their internal heat, but all this shows how nature can contrive to make each world a little bit different from its neighbors.

Magnetic Fields

Each of the giant planets has a strong magnetic field, generated by electric currents in its rapidly spinning interior. Associated with the magnetic fields are the planets' *magnetospheres*, which are regions around the planet within which the planet's own magnetic field dominates over the general interplanetary magnetic field. The magnetospheres of these planets are their largest features, extending millions of kilometers into space.

In the late 1950s, astronomers discovered that Jupiter was a source of radio waves that got more intense at longer rather than at shorter wavelengths—just the reverse of what is expected from thermal radiation (radiation caused by the normal vibrations of particles within all matter). Such behavior is typical, however, of the radiation emitted when high-speed electrons are accelerated by a magnetic field. We call this synchrotron radiation because it was first observed on Earth in particle accelerators, called synchrotrons. This was our first hint that Jupiter must have a strong magnetic field.

Later observations showed that the radio waves are coming from a region surrounding Jupiter with a diameter several times that of the planet itself (Figure 11.8). The evidence suggested that a vast number of charged atomic particles must be circulating around Jupiter, spiraling around the lines of force of a magnetic field associated with the planet. This is just what we observe happening, but on a smaller scale, in the Van Allen belt around Earth. The magnetic fields of Saturn, Uranus, and Neptune, discovered by the spacecraft that first passed close to these planets, work in a similar way, but are not as strong.

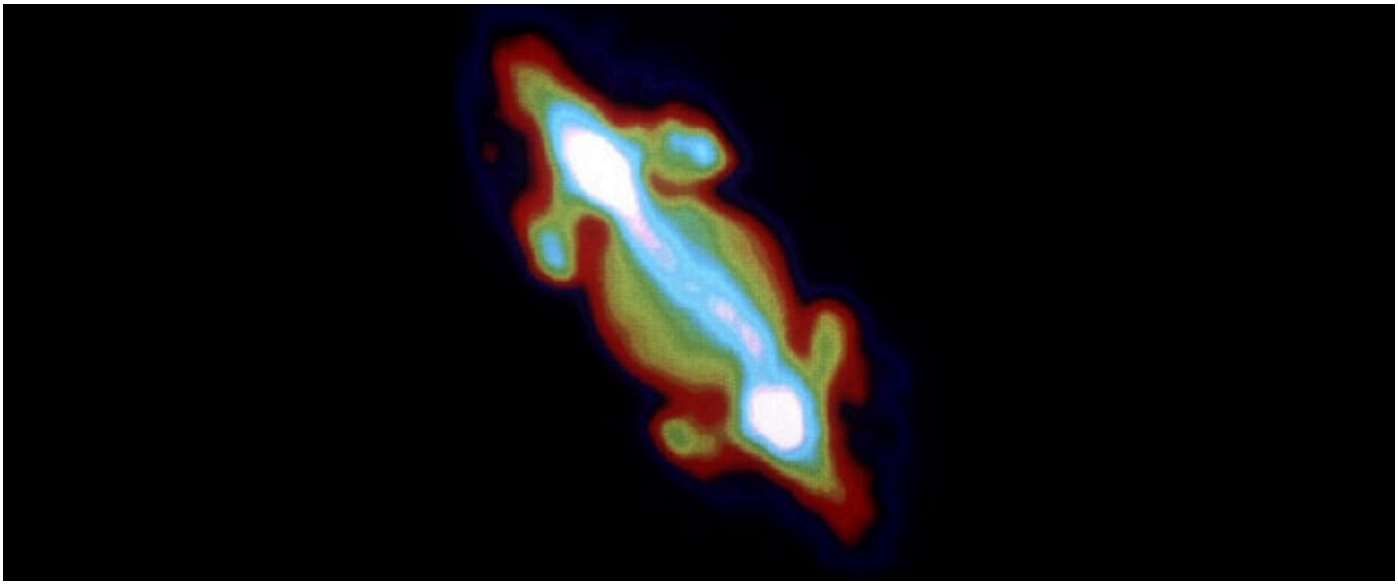


Figure 11.8 : Jupiter in Radio Waves. This false-color image of Jupiter was made with the Very Large Array (of radio telescopes) in New Mexico. We see part of the magnetosphere, brightest in the middle because the largest number of charged particles are in the equatorial zone of Jupiter. The planet itself is slightly smaller than the green oval in the center. Different colors are used to indicate different intensities of synchrotron radiation. (credit: modification of work by I. de Pater (UC Berkeley) NRAO, AUI, NSF)

Link to Learning

Learn more about the magnetosphere of Jupiter and why we continue to be interested in it from this brief NASA video.

Inside each magnetosphere, charged particles spiral around in the magnetic field; as a result, they can be accelerated to high energies. These charged particles can come from the Sun or from the neighborhood of the planet itself. In Jupiter's case, Io, one of its moons, turns out to have volcanic eruptions that blast charged particles into space and right into the jovian magnetosphere.

The axis of Jupiter's magnetic field (the line that connects the magnetic north pole with the magnetic south pole) is not aligned exactly with the axis of rotation of the planet; rather, it is tipped by about 10°. Uranus and Neptune have even greater magnetic tilts, of 60° and 55°, respectively. Saturn's field, on the other hand, is perfectly aligned with its rotation axis. Why different planets

have such different magnetic tilts is not well understood.

The physical processes around the jovian planets turn out to be milder versions of what astronomers find in many distant objects, from the remnants of dead stars to the puzzling distant powerhouses we call quasars. One reason to study the magnetospheres of the giant planets and Earth is that they provide nearby accessible analogues of more energetic and challenging cosmic processes.

This page titled 9.2: The Giant Planets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 11.3: The Giant Planets by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

9.3: The Gailean Moons of Jupiter

Learning Objectives

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

- Describe the major features we can observe about Callisto and what we can deduce from them
- Explain the evidence for tectonic and volcanic activity on Ganymede
- Explain what may be responsible for the unusual features on the icy surface of Europa
- Describe the major distinguishing characteristic of Io
- Explain how tidal forces generate the geological activity we see on Europa and Io

From 1996 to 1999, the Galileo spacecraft careered through the jovian system on a complex but carefully planned trajectory that provided repeated close encounters with the large Galilean moons. More recently, the Juno spacecraft, also a Jupiter orbiter, has provided a few close looks at Ganymede and Europa. (Beginning in 2004, we received an even greater bonanza of information about Titan, obtained from the Cassini spacecraft and its Huygens probe, which landed on its surface. We include Titan, Saturn's one big moon, here for comparison.) Table 12.1 summarizes some basic facts about these large moons (plus our own Moon for comparison).

Table 12.1: The Largest Moons

Name	Diameter (km)	Mass (Earth's Moon = 1)	Density (g/cm^3)	Reflectivity (%)
Moon	3476	1.0	3.3	12
Callisto	4820	1.5	1.8	20
Ganymede	5270	2.0	1.9	40
Europa	3130	0.7	3.0	70
Io	3640	1.2	3.5	60
Titan	5150	1.9	1.9	20

Callisto: An Ancient, Primitive World

We begin our discussion of the Galilean moons with the outermost one, Callisto, not because it is remarkable but because it is not. This makes it a convenient object with which other, more active, worlds can be compared. Its distance from Jupiter is about 2 million kilometers, and it orbits the planet in 17 days. Like our own Moon, Callisto rotates in the same period as it revolves, so it always keeps the same face toward Jupiter. Callisto's day thus equals its month: 17 days. Its noontime surface temperature is only 130 K (about 140 °C below freezing), so that water ice is stable (it never evaporates) on its surface year round.

Callisto has a diameter of 4820 kilometers, almost the same as the planet Mercury (Figure 12.3). Yet its mass is only one-third as great, which means its density (the mass divided by the volume) must be only one-third as great as well. This tells us that Callisto has far less of the rocky and metallic materials found in the inner planets and must instead be an icy body through much of its interior. Callisto can show us how the geology of an icy object compares with those made primarily of rock.

Unlike the worlds we have studied so far, Callisto has not fully *differentiated* (separated into layers of different density materials). We can tell that it lacks a dense core from the details of its gravitational pull on the Galileo spacecraft. This surprised scientists, who expected that all the big icy moons would be differentiated. It should be easier for an icy body to differentiate than for a rocky one because the melting temperature of ice is so low. Only a little heating will soften the ice and get the process started, allowing the rock and metal to sink to the center while the slushy ice floats to the surface. Yet Callisto seems to have frozen solid before the process of differentiation was complete.

The surface of Callisto is covered with impact craters, like the lunar highlands. The survival of these craters tells us that an icy

object can retain impact craters on its surface. Callisto is unique among the planet-sized objects of the solar system in the apparent absence of interior forces to drive geological change. You might say that this moon was stillborn, and it has remained geologically dead for more than 4 billion years (Figure 12.3).

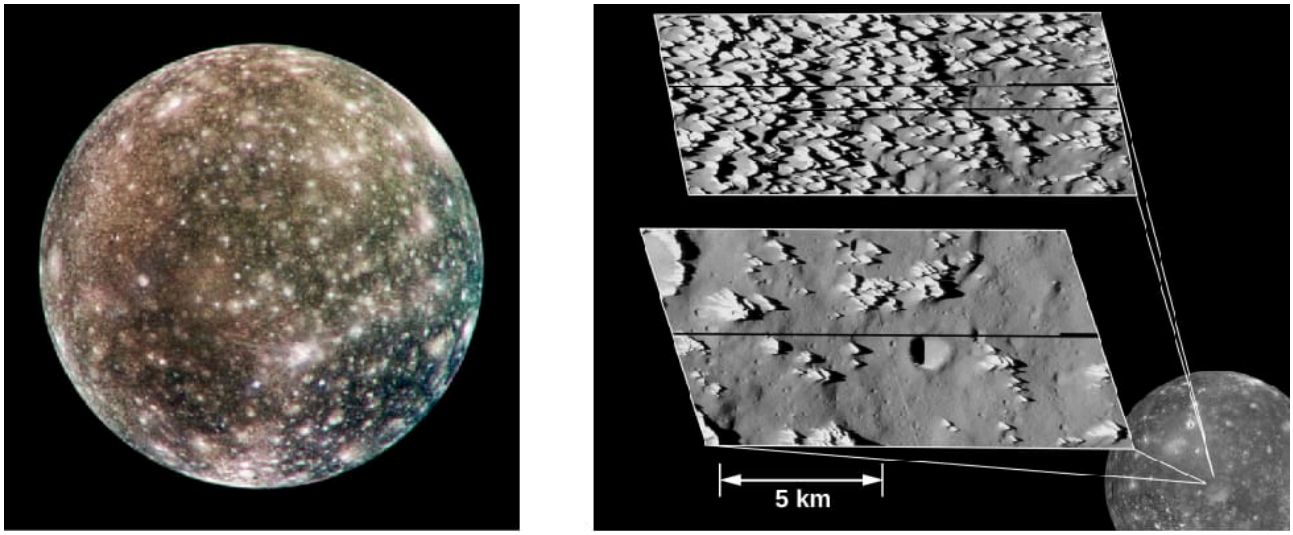


Figure 12.3 : Callisto. (a) Jupiter's outermost large moon shows a heavily cratered surface. Astronomers believe that the bright areas are mostly ice, while the darker areas are more eroded, ice-poor material. (b) These high-resolution images, taken by NASA's Galileo spacecraft in May 2001, show the icy spires (top) on Callisto's surface, with darker dust that has slid down as the ice erodes, collecting in the low-lying areas. The spires are about 80 to 100 meters tall. As the surface erodes even further, the icy spires eventually disappear, leaving impact craters exposed, as shown in the lower image. (credit a: modification of work by NASA/JPL/DLR; credit b: modification of work by NASA/JPL/Arizona State University, Academic Research Lab)

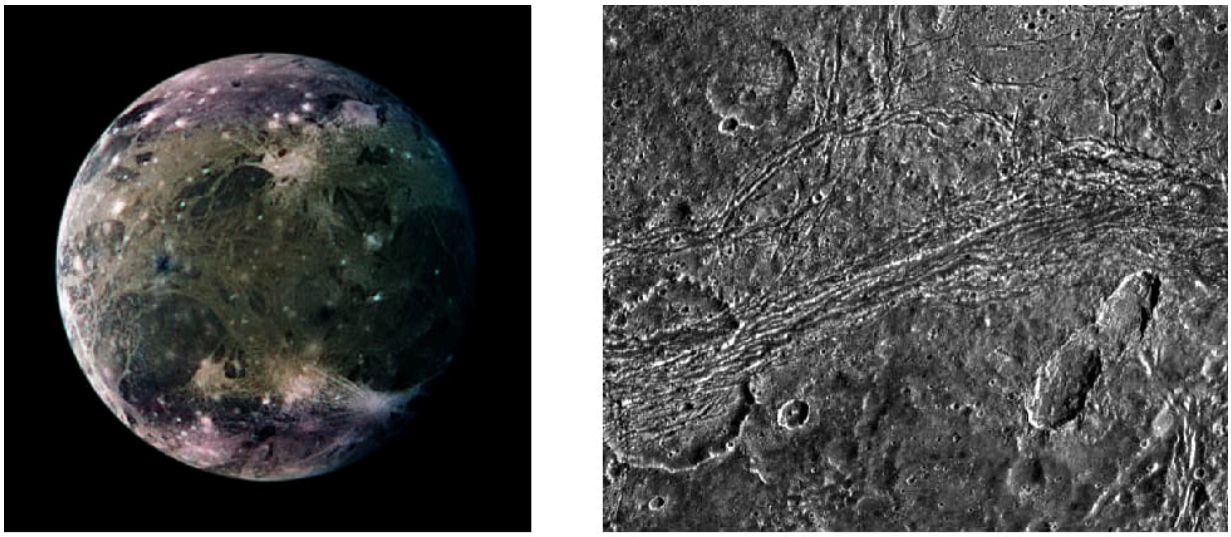
In thinking about ice so far from the Sun, we must take care not to judge its behavior from the much warmer ice we know and love on Earth. At the temperatures of the outer solar system, ice on the surface is nearly as hard as rock, and it behaves similarly. Ice on Callisto does not deform or flow like ice in glaciers on Earth.

Ganymede, the Largest Moon

Ganymede, the largest moon in the solar system, also shows a great deal of cratering (Figure 12.4). Recall from Other Worlds: An Introduction to the Solar System that we can use crater counts on solid worlds to estimate the age of the surface. The more craters, the longer the surface has been exposed to battering from space, and the older it must therefore be. About one-quarter of Ganymede's surface seems to be as old and heavily cratered as that of Callisto; the rest formed more recently, as we can tell by the sparse covering of impact craters as well as the relative freshness of those craters. If we judge from crater counts, this fresher terrain on Ganymede is somewhat younger than the lunar maria or the martian volcanic plains, perhaps 2 to 3 billion years old.

The differences between Ganymede and Callisto are more than skin deep. Ganymede is a differentiated world, like the terrestrial planets. Measurements of its gravity field tell us that the rock sank to form a core about the size of our Moon, with a mantle and crust of ice "floating" above it. In addition, the Galileo spacecraft discovered that Ganymede has a magnetic field, the sure signature of a partially molten interior. There is very likely liquid water trapped within the interior. Thus, Ganymede is not a dead world but rather a place of intermittent geological activity powered by an internal heat source. Some surface features could be as young as the surface of Venus (a few hundred million years).

The younger terrain was formed by tectonic and volcanic forces (Figure 12.4). In some places, the crust apparently cracked, flooding many of the craters with water from the interior. Extensive mountain ranges were formed from compression of the crust, forming long ridges with parallel valleys spaced a few kilometers apart. In some areas, older impact craters were split and pulled apart. There are even indications of large-scale crustal movements that are similar to the plate tectonics of Earth.



(a) (b)

Figure 12.4 : Ganymede. (a) This global view of Ganymede, the largest moon in the solar system, was taken by Voyager 2. The colors are enhanced to make spotting differences easier. Darker places are older, more heavily cratered regions; the lighter areas are younger (the reverse of our Moon). The brightest spots are sites of geologically recent impacts. (b) This close-up of Nicholson Regio on Ganymede shows an old impact crater (on the lower left-hand side) that has been split and pulled apart by tectonic forces. Against Ganymede's dark terrain, a line of grooves and ridges appears to cut through the crater, deforming its circular shape. (credit a: modification of work by NASA/JPL/DLR; credit b: modification of work by NASA/JPL/Brown University)

Why is Ganymede so different from Callisto? Possibly the small difference in size and internal heating between the two led to this divergence in their evolution. But more likely the gravity of Jupiter is to blame for Ganymede's continuing geological activity. Ganymede is close enough to Jupiter that *tidal forces* from the giant planet may have episodically heated its interior and triggered major convulsions on its crust.

A tidal force results from the unequal gravitational pull on two sides of a body. In a complex kind of modern dance, the large moons of Jupiter are caught in the varying gravity grip of both the giant planet and each other. This leads to gravitational flexing or kneading in their centers, which can heat them—an effect called tidal heating. (A fuller explanation is given in the section on Io.) We will see as we move inward to Europa and Io that the role of jovian tides becomes more important for moons close to the planet.

Europa, a Moon with an Ocean

Europa and Io, the inner two Galilean moons, are not icy worlds like most of the moons of the outer planets. With densities and sizes similar to our Moon, they appear to be predominantly rocky objects. How did they fail to acquire a majority share of the ice that must have been plentiful in the outer solar system at the time of their formation?

The most probable cause is Jupiter itself, which was hot enough to radiate a great deal of infrared energy during the first few million years after its formation. This infrared radiation would have heated the disk of material near the planet that would eventually coalesce into the closer moons. Thus, any ice near Jupiter was vaporized, leaving Europa and Io with compositions similar to planets in the inner solar system.

Despite its mainly rocky composition, Europa has an ice-covered surface, as astronomers have long known from examining spectra of sunlight reflected from it. In this it resembles Earth, which has a layer of water on its surface, but in Europa's case the water is capped by a thick crust of ice. There are very few impact craters in this ice, indicating that the surface of Europa is in a continual state of geological self-renewal. Judging from crater counts, the surface must be no more than a few million years old, and perhaps substantially less. In terms of its ability to erase impact craters, Europa is more geologically active than Earth.

When we look at close-up photos of Europa, we see a strange, complicated surface (Figure 12.5). For the most part, the icy crust is extremely smooth, but it is crisscrossed with cracks and low ridges that often stretch for thousands of kilometers. Some of these long lines are single, but most are double or multiple, looking rather like the remnants of a colossal freeway system.



Figure 12.5 : Evidence for an Ocean on Europa. (a) A close-up of an area called Conamara Chaos is shown here with enhanced color. This view is 70 kilometers wide in its long dimension. It appears that Conamara is a region where Europa's icy crust is (or recently was) relatively thin and there is easier access to the possible liquid or slushy ocean beneath. Not anchored to solid crust underneath, many of the ice blocks here seem to have slid or rotated from their original positions. In fact, the formations seen here look similar to views of floating sea-ice and icebergs in Earth's Arctic Ocean. (b) In this high-resolution view, the ice is *wrinkled* and crisscrossed by long ridges. Where these ridges intersect, we can see which ones are older and which younger; the younger ones cross over the older ones. While superficially this system of ridges resembles a giant freeway system on Europa, the ridges are much wider than our freeways and are a natural result of the flexing of the moon. (credit a: modification of work by NASA/JPL/University of Arizona; credit b: modification of work by NASA/JPL)

It is very difficult to make straight lines on a planetary surface. In discussing Mars, we explained that when Percival Lowell saw what appeared to him to be straight lines (the so-called martian "canals"), he attributed them to the engineering efforts of intelligent beings. We now know the lines on Mars were optical illusions, but the lines on Europa are real. These long cracks can form in the icy crust if it is floating without much friction on an ocean of liquid water (Figure 12.6).

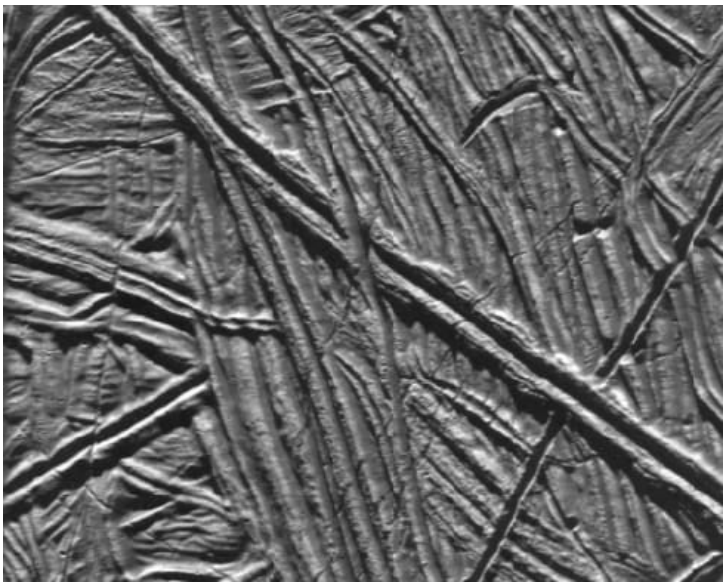


Figure 12.6 : Very High-Resolution Galileo Image of One Young Double Ridge on Europa. The area in this picture is only 15 kilometers across. It appears to have formed when viscous icy material was forced up through a long, straight crack in the crust. Note how the young ridge going from top left toward bottom right lies on top of older features, which are themselves on top of even older ones. (credit: modification of work by NASA/JPL)

The close-up Galileo images appear to confirm the existence of a global ocean. In many places, the surface of Europa looks just as we would expect for a thick layer of ice that was broken up into giant icebergs and ice floes and then refrozen in place. When the ice breaks, water or slush from below may be able to seep up through the cracks and make the ridges and multiple-

line features we observe. Many episodes of ice cracking, shifting, rotating, and refreezing are required to explain the complexity we see. The icy crust might vary in thickness from a kilometer or so up to 20 kilometers. Further confirmation that a liquid ocean exists below the ice comes from measurements of the small magnetic field induced by Europa's interactions with the magnetosphere of Jupiter. The "magnetic signature" of Europa is that of a liquid water ocean, not one of ice or rock.

If Europa really has a large ocean of liquid water under its ice, then it may be the only place in the solar system, other than Earth, with really large amounts of liquid water.[1] To remain liquid, this ocean must be warmed by heat escaping from the interior of Europa. Hot (or at least warm) springs might be active there, analogous to those we have discovered in the deep oceans of Earth. The necessary internal heat is generated by tidal heating (see the discussion later in this chapter).

Link to Learning

A short film with planetary scientist Kevin Hand explains why Europa is so interesting for future exploration. Or listen to this more in-depth talk on Europa.

What makes the idea of an ocean with warm springs exciting is the discovery in Earth's oceans of large ecosystems clustered around deep ocean hot springs. Such life derives all its energy from the mineral-laden water and thrives independent of the sunlight shining on Earth's surface. Is it possible that similar ecosystems could exist today under the ice of Europa?

Many scientists now think that Europa is the most likely place beyond Earth to find life in the solar system. In response, NASA is designing a Europa mission to characterize its liquid ocean and its ice crust, and to identify locations where material from inside has risen to the surface. Such interior material might reveal direct evidence for microbial life. This "Europa Clipper" mission will not orbit Europa, where the intense energetic particles from Jupiter's magnetosphere would quickly destroy its electronics, but instead will swoop in for brief close flybys. It is planned for launch in 2024.

Io, a Volcanic Moon

Io, the innermost of Jupiter's Galilean moons, is in many ways a close twin of our Moon, with nearly the same size and density. We might therefore expect it to have experienced a similar history. Its appearance, as photographed from space, tells us another story, however (Figure 12.7). Instead of being a dead cratered world, Io turns out to have the highest level of volcanism in the solar system, greatly exceeding that of Earth.

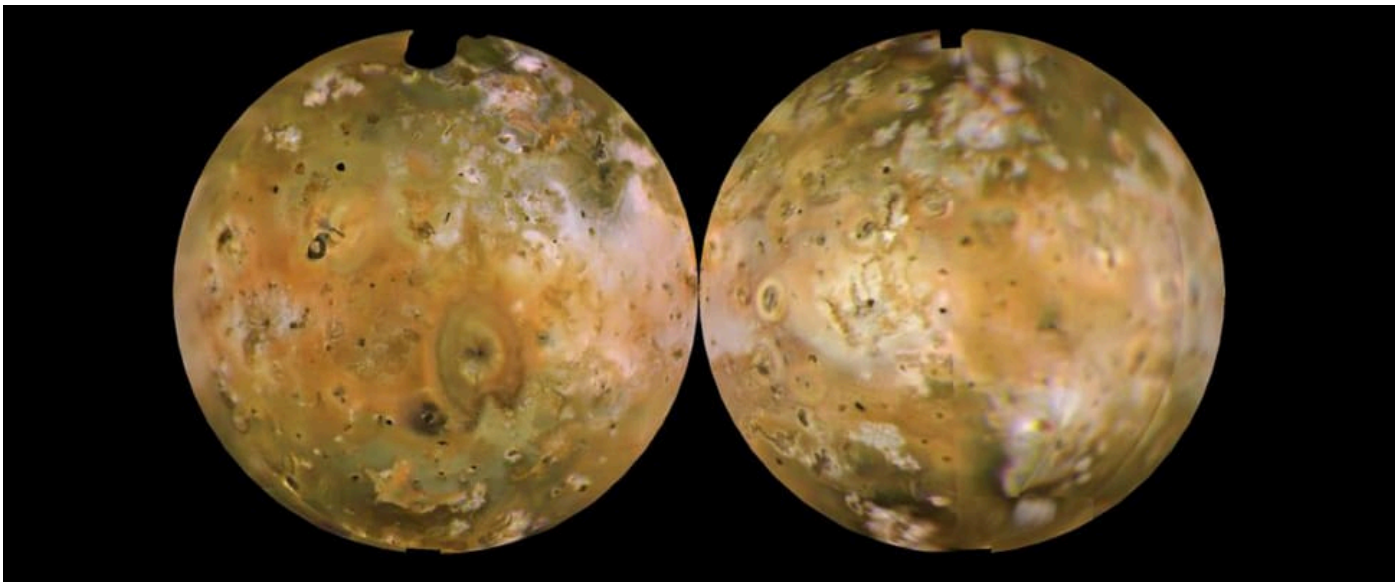


Figure 12.7 : Two Sides of Io. This composite image shows both sides of the volcanically active moon Io. The orange deposits are sulfur snow; the white is sulfur dioxide. (Carl Sagan once quipped that Io looks as if it desperately needs a shot of penicillin.) (credit: modification of work by NASA/JPL/USGS)

Io's active volcanism was discovered by the Voyager spacecraft. Eight volcanoes were seen erupting when Voyager 1 passed in

March 1979, and six of these were still active four months later when Voyager 2 passed. With the improved instruments carried by the Galileo spacecraft, more than 50 eruptions were found during 1997 alone. Many of the eruptions produce graceful plumes that extend hundreds of kilometers out into space (Figure 12.8).

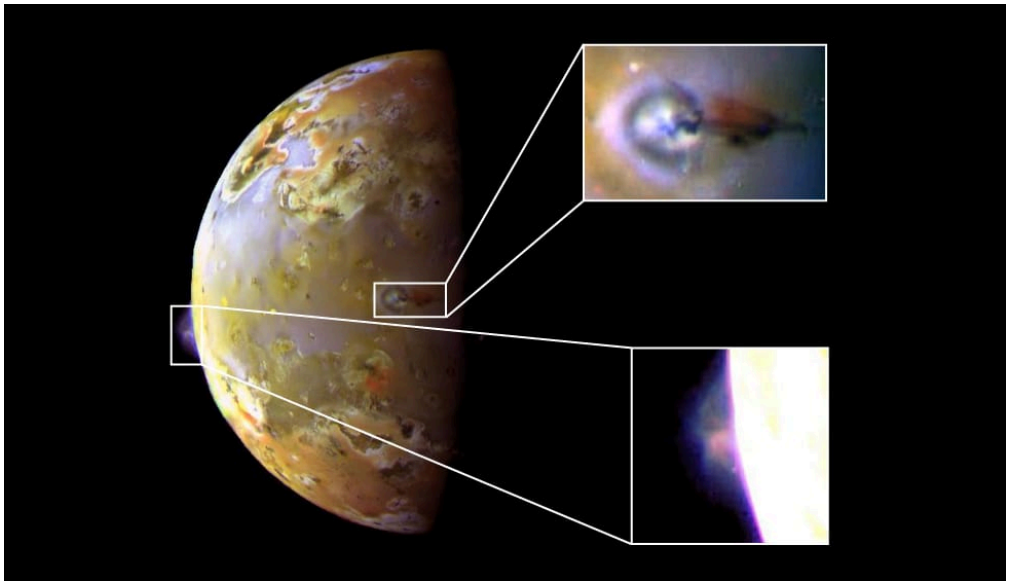


Figure 12.8 : Volcanic Eruptions on Io. This composite image from NASA's Galileo spacecraft shows close-ups (the two inset photos) of two separate volcanic eruptions on Jupiter's volcanic moon, Io. In the upper inset image, you can see a close up of a bluish plume rising about 140 kilometers above the surface of the volcano. In the lower inset image is the Prometheus plume, rising about 75 kilometers from Io's surface. The Prometheus plume is named for the Greek god of fire. (credit: modification of work by NASA/JPL)

Link to Learning

Watch a brief movie made from Voyager and Galileo data, showing a rotating Io with its dramatic surface features.

The Galileo data show that most of the volcanism on Io consists of hot silicate lava, like the volcanoes on Earth. Sometimes the hot lava encounters frozen deposits of sulfur and sulfur dioxide. When these icy deposits are suddenly heated, the result is great eruptive plumes far larger than any ejected from terrestrial volcanoes. As the rising plumes cool, the sulfur and sulfur dioxide recondense as solid particles that fall back to the surface in colorful "snowfalls" that extend as much as a thousand kilometers from the vent. Major new surface features were even seen to appear between Galileo orbits, as shown in Figure 12.9.



Figure 12.9 : Volcanic Changes on Io. These three images were taken of the same 1700-kilometer-square region of Io in April 1997, September 1997, and July 1999. The dark volcanic center called Pillan Patera experienced a huge eruption, producing a dark deposit some 400 kilometers across (seen as the grey area in the upper center of the middle image). In the right image, however, some of the new dark deposit is already being covered by reddish material from the volcano Pele. Also, a small unnamed volcano to the right of Pillan has erupted since 1997, and some of its dark deposit and a yellow ring around it are visible on the right image (to the right of the grey spot). The color range is exaggerated in these images. (credit: modification of work by NASA/JPL/University of Arizona)

As the Galileo mission drew to a close, controllers were willing to take risks in getting close to Io. Approaching this moon is a dangerous maneuver because the belts of atomic particles trapped in Jupiter's magnetic environment are at their most intense near Io's orbit. Indeed, in its very first pass by Io, the spacecraft absorbed damaging radiation beyond its design levels. To keep the system working at all, controllers had to modify or disable various fault-protection software routines in the onboard computers. In spite of these difficulties, the spacecraft achieved four successful Io flybys, obtaining photos and spectra of the surface with unprecedented resolution.

Maps of Io reveal more than 100 recently active volcanoes. Huge flows spread out from many of these vents, covering about 25% of the moon's total surface with still-warm lava. From these measurements, it seems clear that the bright surface colors that first attracted attention to Io are the result of a thin veneer of sulfur compounds. The underlying volcanism is driven by eruptions of molten silicates, just like on Earth (Figure 12.10).



Figure 12.10 : Lava Fountains on Io. Galileo captured a number of eruptions along the chain of huge volcanic calderas (or pits) on Io called Tvashtar Catena in this false-color image combining infrared and visible light. The bright orange-yellow areas at left are places where fresh, hot lava is erupting from below ground. (credit: modification of work by NASA/JPL)

Tidal Heating

How can Io remain volcanically active in spite of its small size? The answer, as we hinted earlier, lies in the effect of gravity, through tidal heating. Io is about the same distance from Jupiter as our Moon is from Earth. Yet Jupiter is more than 300 times more massive than Earth, causing forces that pull Io into an elongated shape, with a several-kilometer-high bulge extending toward Jupiter.

If Io always kept exactly the same face turned toward Jupiter, this bulge would not generate heat. However, Io's orbit is not exactly circular due to gravitational perturbations (tugs) from Europa and Ganymede. In its slightly eccentric orbit, Io twists back and forth with respect to Jupiter, at the same time moving nearer and farther from the planet on each revolution. The twisting and flexing heat Io, much as repeated flexing of a wire coat hanger heats the wire.

After billions of years, this constant flexing and heating have taken their toll on Io, driving away water and carbon dioxide and other gases, so that now sulfur and sulfur compounds are the most volatile materials remaining. Its interior is entirely melted, and the crust itself is constantly recycled by volcanic activity.

In moving inward toward Jupiter from Callisto to Io, we have encountered more and more evidence of geological activity and internal heating, culminating in the violent volcanism on Io. Three of these surfaces are compared in Figure 12.11. Just as the character of the planets in our solar system depends in large measure on their distance from the Sun (and on the amount of heat they receive), so it appears that distance from a giant planet like Jupiter can play a large role in the composition and evolution of its moons (at least partly due to differences in internal heating of each moon by Jupiter's unrelenting tidal forces).

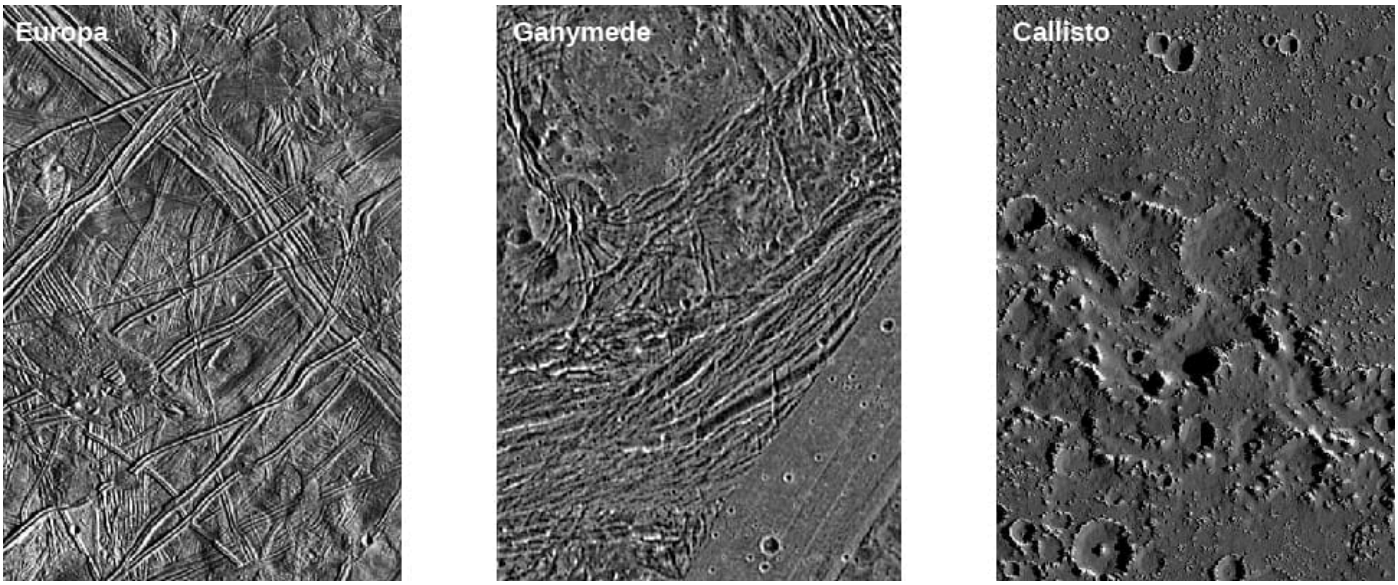


Figure 12.11 : Three Icy Moons. These Galileo images compare the surfaces of Europa, Ganymede, and Callisto at the same resolution. Note that the number of craters (and thus the age of the surface we see) increases as we go from Europa to Ganymede to Callisto. The Europa image is one of those where the system of cracks and ridges resembles a freeway system. (credit: modification of work by NASA/JPL/DLR)

Footnotes

- 1Ganymede and Saturn's moon Enceladus may have smaller amounts of liquid water under their surfaces.

This page titled 9.3: The Gailean Moons of Jupiter is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.3: The Gailean Moons of Jupiter by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

9.4: Titan and Triton

Learning Objectives

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

- Explain how the thick atmosphere of Titan makes bodies of liquid on its surface possible
- Describe what we learned from the landing on Titan with the Huygens probe
- Discuss the features we observed on the surface of Triton when Voyager 2 flew by

We shift our attention now to small worlds in the more distant parts of the solar system. Saturn's large moon Titan turns out to be a weird cousin of Earth, with many similarities in spite of frigid temperatures. The Cassini observations of Titan have provided some of the most exciting recent discoveries in planetary science. Neptune's moon Triton also has unusual characteristics and resembles Pluto, which we will discuss in the following section.

Titan, a Moon with Atmosphere and Hydrocarbon Lakes

Titan, first seen in 1655 by the Dutch astronomer Christiaan Huygens, was the first moon discovered after Galileo saw the four large moons of Jupiter. Titan has roughly the same diameter, mass, and density as Callisto or Ganymede. Presumably it also has a similar composition—about half ice and half rock. However, Titan is unique among moons, with a thick atmosphere and lakes and rivers and falling rain (although these are not composed of water but of hydrocarbons such as ethane and methane, which can stay liquid at the frigid temperatures on Titan). Titan is the only moon or planet other than Earth where we have found evidence of bodies of surface liquids.

The 1980 Voyager flyby of Titan determined that the surface density of its atmosphere is four times greater than that on Earth. The atmospheric pressure on this moon is 1.6 bars, higher than that on any other moon and, remarkably, even higher than that of the terrestrial planets Mars and Earth. The atmospheric composition is primarily nitrogen, an important way in which Titan's atmosphere resembles Earth's.

Also detected in Titan's atmosphere were carbon monoxide (CO), hydrocarbons (compounds of hydrogen and carbon) such as methane (CH_4), ethane (C_2H_6), and propane (C_3H_8), and nitrogen compounds such as hydrogen cyanide (HCN), cyanogen (C_2N_2), and cyanoacetylene (HC_3N). Their presence indicates an active chemistry in which sunlight interacts with atmospheric nitrogen and methane to create a rich mix of organic molecules. There are also multiple layers of hydrocarbon haze and clouds in the atmosphere, as illustrated in Figure 12.12.

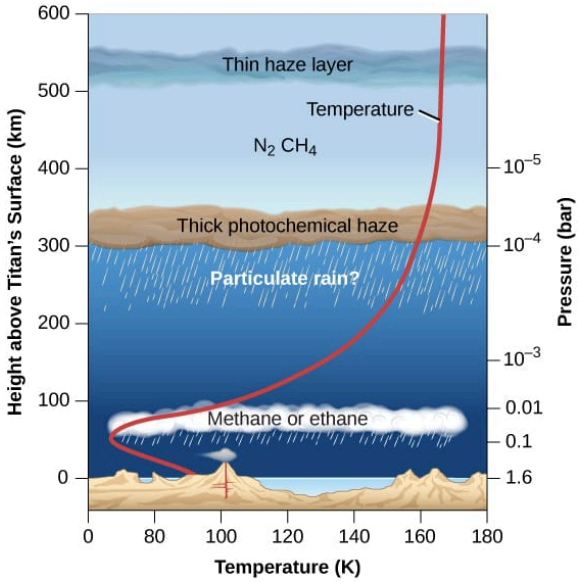


Figure 12.12 : Structure of Titan's Atmosphere. Some characteristics of Titan's atmosphere resemble those of Earth's atmosphere, although it is much colder than our planet. The red line indicates the temperature of Titan's atmosphere at different altitudes.

These Voyager discoveries motivated a much more ambitious exploration program using the NASA Cassini Saturn orbiter and a probe to land on Titan called Huygens, built by the European Space Agency. The orbiter, which included several cameras, spectrometers, and a radar imaging system, made dozens of close flybys of Titan between 2004 and 2015, each yielding radar and infrared images of portions of the surface (see Exploring the Outer Planets). The Huygens probe successfully descended by parachute through the atmosphere, photographing the surface from below the clouds, and landing on January 14, 2005. This was the first (and so far the only) spacecraft landing on a moon in the outer solar system.

At the end of its parachute descent, the 319-kilogram Huygens probe safely touched down, slid a short distance, and began sending data back to Earth, including photos and analyses of the atmosphere. It appeared to have landed on a flat, boulder-strewn plain, but both the surface and the boulders were composed of water ice, which is as hard as rock at the temperature of Titan (see Figure 12.13).

The photos taken during descent showed a variety of features, including drainage channels, suggesting that Huygens had landed on the shore of an ancient hydrocarbon lake. The sky was deep orange, and the brightness of the Sun was a thousand times less than sunlight on Earth (but still more than a hundred times brighter than under the full moon on Earth). Titan's surface temperature was 94 K (–179 °C). The warmer spacecraft heated enough of the ice where it landed for its instruments to measure released hydrocarbon gas. Measurements on the surface continued for more than an hour before the probe succumbed to the frigid temperature.

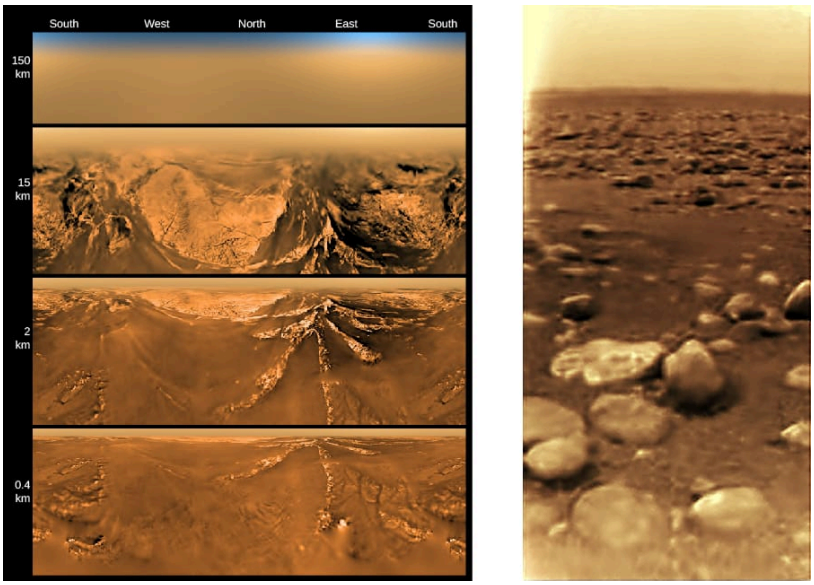


Figure 12.13 : Views of the Surface of Titan. The left image shows the views of Titan from the descent camera, in a flattened projection, at different altitudes. The right image, taken after landing, shows a boulder-strewn surface illuminated by faint reddish sunlight. The boulders are composed of water ice. (credit left: modification of work by ESA/NASA/JPL/University of Arizona; credit right: modification of work by ESA/NASA/JPL/University of Arizona; processed by Andrey Pivovarov)

Radar and infrared imaging of Titan from the Cassini orbiter gradually built up a picture of a remarkably active surface on this moon, complex and geologically young (Figure 12.14). There are large methane lakes near the polar regions that interact with the methane in the atmosphere, much as Earth's water oceans interact with the water vapor in our atmosphere. The presence of many erosional features indicates that atmospheric methane can condense and fall as rain, then flow down valleys to the big lakes. Thus, Titan has a low-temperature equivalent of the water cycle on Earth, with liquid on the surface that evaporates, forms clouds, and then condenses to fall as rain—but on Titan the liquid is a combination of methane, ethane, and a trace of other hydrocarbons. It is a weirdly familiar and yet utterly alien landscape.

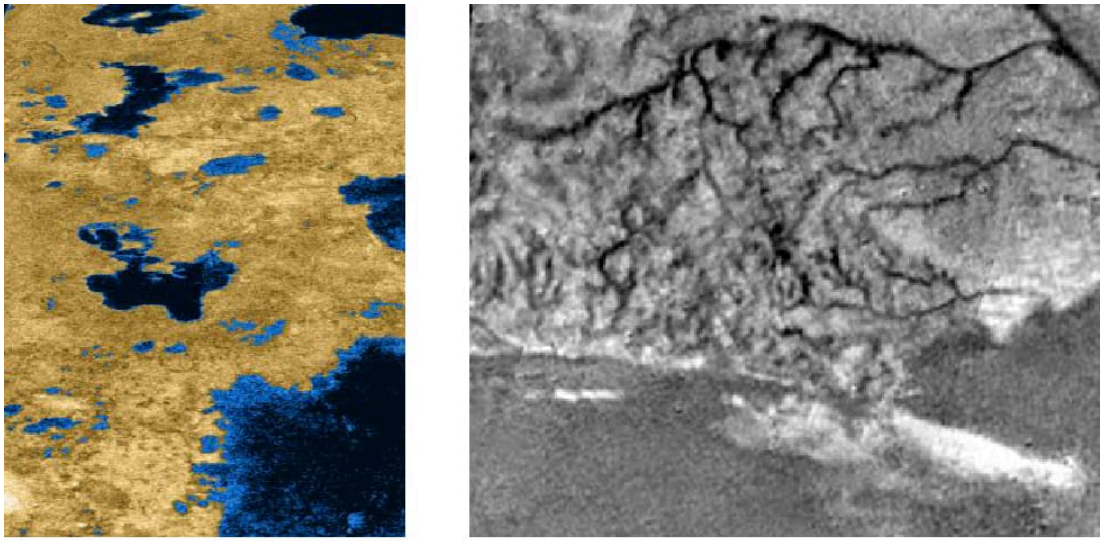


Figure 12.14 : Titan's Lakes. (a) This Cassini image from a September 2006 flyby shows the liquid lakes on Titan. Their composition is most likely a combination of methane and ethane. (Since this is a radar image, the colors are artificially added. The dark blue areas are the smooth surfaces of the liquid lakes, and yellow is the rougher solid terrain around them.) (b) This mosaic of Titan's surface from the Cassini-Huygens mission shows in detail a high ridge area and many narrow, sinuous erosion channels that appear to be part of a widespread network of "rivers" carved by flowing hydrocarbons. (credit a: modification of work by NASA/JPL-Caltech/USGS; credit b; modification of work by NASA/JPL/ESA/University of Arizona)

These discoveries raise the question of whether there could be life on Titan. Hydrocarbons are fundamental for the formation of

the large carbon molecules that are essential to life on our planet. However, the temperature on Titan is far too low for liquid water or for many of the chemical processes that are essential to life as we know it. There remains, though, an intriguing possibility that Titan might have developed a different form of low-temperature carbon-based life that could operate with liquid hydrocarbons playing the role of water. The discovery of such "life as we don't know it" could be even more exciting than finding life like ours on Mars. If such a truly alien life is present on Titan, its existence would greatly expand our understanding of the nature of life and of habitable environments.

NASA has selected a new mission to Titan for launch in 2027. Called *Dragonfly*, this mission is a drone that will fly in Titan's atmosphere, with emphasis on study of pre-biotic chemistry. Other future proposed missions include a balloon operating in the atmosphere and a "boat" floating in one of the Titan lakes.

Link to Learning

The Cassini mission scientists and the visual presentation specialists at NASA's Jet Propulsion Laboratory have put together some nice films from the images taken by Cassini and Huygens. See, for example, the Titan approach and the flyover of the Northern lakes district.

Triton and Its Volcanoes

Neptune's largest moon Triton (don't get its name confused with Titan) has a diameter of 2720 kilometers and a density of 2.1 g/cm^3, indicating that it's probably composed of about 75% rock mixed with 25% water ice. Measurements indicate that Triton's surface has the coldest temperature of any of the worlds our robot representatives have visited. Because its reflectivity is so high (about 80%), Triton reflects most of the solar energy that falls on it, resulting in a surface temperature between 35 and 40 K.

The surface material of Triton is made of frozen water, nitrogen, methane, and carbon monoxide. Methane and nitrogen exist as gas in most of the solar system, but they are frozen at Triton's temperatures. Only a small quantity of nitrogen vapor persists to form an atmosphere. Although the surface pressure of this atmosphere is only 16 millionths of a bar, this is sufficient to support thin haze or cloud layers.

Triton's surface, like that of many other moons in the outer solar system, reveals a long history of geological evolution (Figure 12.15). Although some impact craters are found, many regions have been flooded fairly recently by the local version of "lava" (perhaps water or water-ammonia mixtures). There are also mysterious regions of jumbled or mountainous terrain.



Figure 12.15 : Neptune's Moon Triton. This mosaic of Voyager 2 images of Triton shows a wide range of surface features. The pinkish area at the bottom is Triton's large southern polar cap. The south pole of Triton faces the Sun here, and the slight heating effect is driving some of the material northward, where it is colder. (credit: modification of work by NASA/JPL/USGS)

The Voyager flyby of Triton took place at a time when the moon's southern pole was tipped toward the Sun, allowing this part of the surface to enjoy a period of relative warmth. (Remember that "warm" on Triton is still outrageously colder than anything we experience on Earth.) A polar cap covers much of Triton's southern hemisphere, apparently evaporating along the northern edge. This polar cap may consist of frozen nitrogen that was deposited during the previous winter.

Remarkably, the Voyager images showed that the evaporation of Triton's polar cap generates geysers or volcanic plumes of nitrogen gas (see Figure 12.16). (Fountains of such gas rose about 10 kilometers high, visible in the thin atmosphere because dust from the surface rose with them and colored them dark.) These plumes differ from the volcanic plumes of Io in their composition and also in that they derive their energy from sunlight warming the surface rather than from internal heat.

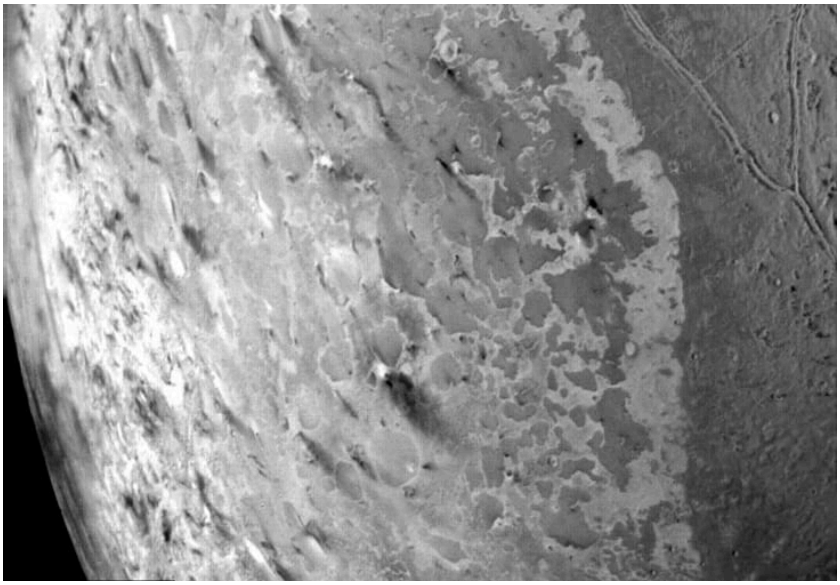


Figure 12.16 : Triton's Geysers. This close-up view shows some of the geysers on Neptune's moon Triton, with the long trains of dust pointing to the lower right in this picture. (credit: modification of work by NASA/JPL)

This page titled 9.4: Titan and Triton is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.4: Titan and Triton by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10: ASTEROIDS, COMETS, AND DWARF PLANETS

10.1: Asteroids
10.2: Meteors
10.3: Meteorites - Stones from Heaven
10.4: The "Long-Haired" Comets
10.5: The Origin and Fate of Comets and Related Objects
10.6: Pluto and Charon
10.7: Cosmic Influences on the Evolution of Earth

This page titled 10: Asteroids, Comets, and Dwarf Planets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

10.1: Asteroids

Learning Objectives

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

- Outline the story of the discovery of asteroids and describe their typical orbits
- Describe the composition and classification of the various types of asteroids
- Discuss what was learned from spacecraft missions to several asteroids

The asteroids are mostly found in the broad space between Mars and Jupiter, a region of the solar system called the *asteroid belt*. Asteroids are too small to be seen without a telescope; the first of them was not discovered until the beginning of the nineteenth century.

Discovery and Orbits of the Asteroids

In the late 1700s, many astronomers were hunting for an additional planet they thought should exist in the gap between the orbits of Mars and Jupiter. The Sicilian astronomer Giovanni Piazzi thought he had found this missing planet in 1801, when he discovered the first asteroid (or as it was later called, "minor planet") orbiting at 2.8 AU from the Sun. His discovery, which he named Ceres, was quickly followed by the detection of three other little planets in similar orbits.

Clearly, there was not a single missing planet between Mars and Jupiter but rather a whole group of objects, each much smaller than our Moon. (An analogous discovery history has played out in slow motion in the outer solar system. Pluto was discovered beyond Neptune in 1930 and was initially called a planet, but early in the twenty-first century, several other similar objects were found. We now call all of them dwarf planets.)

By 1890, more than 300 of these minor planets or asteroids had been discovered by sharp-eyed observers. In that year, Max Wolf at Heidelberg introduced astronomical photography to the search for asteroids, greatly accelerating the discovery of these dim objects. In the twenty-first century, searchers use computer-driven electronic cameras, another leap in technology. Currently many asteroid searches are made with cameras in space. More than a million asteroids now have well-determined orbits.

Asteroids are given a number (corresponding to the order of discovery) and sometimes also a name. Originally, the names of asteroids were chosen from goddesses in Greek and Roman mythology. After exhausting these and other female names (including, later, those of spouses, friends, flowers, cities, and others), astronomers turned to the names of colleagues (and other people of distinction) whom they wished to honor. For example, asteroids 2410, 4859, and 68448 are named Morrison, Fraknoi, and Sidneywolff, for the three senior authors of this textbook.

The largest asteroid is Ceres (numbered 1), with a diameter just less than 1000 kilometers. As we saw, Ceres was considered a planet when it was discovered but later was called an asteroid (the first of many.) Now, it has again been reclassified and is considered one of the dwarf planets, like Pluto (see the chapter on Moons, Rings and Pluto). We still find it convenient, however, to discuss Ceres as the largest of the asteroids. Two other asteroids, Pallas and Vesta, have diameters of about 500 kilometers, and about 15 more are larger than 250 kilometers (see Table 13.1). The number of asteroids increases rapidly with decreasing size; there are about 100 times more objects 10 kilometers across than there are 100 kilometers across.

Link to Learning

The Minor Planet Center is a worldwide repository of data on asteroids. Visit it online to find out about the latest discoveries related to the small bodies in our solar system. (Note that some of the material on this site is technical; it's best to click on the menu tab for the "public" for information more at the level of this textbook.) Watch this video about asteroid size comparison for a dramatic visualization of the sizes of some better-known asteroids.

Table 13.1: The Largest Asteroids

#	Name	Year of Discovery	Orbit's Semimajor Axis (AU)	Diameter (km)	Compositional Class
1	Ceres	1801	2.77	940	C (carbonaceous)
2	Pallas	1802	2.77	540	C (carbonaceous)
3	Juno	1804	2.67	265	S (stony)
4	Vesta	1807	2.36	510	basaltic
10	Hygiea	1849	3.14	410	C (carbonaceous)
16	Psyche	1852	2.92	265	M (metallic)
31	Euphrosyne	1854	3.15	250	C (carbonaceous)
52	Europa	1858	3.10	280	C (carbonaceous)
65	Cybele	1861	3.43	280	C (carbonaceous)
87	Sylvia	1866	3.48	275	C (carbonaceous)
451	Patientia	1899	3.06	260	C (carbonaceous)
511	Davida	1903	3.16	310	C (carbonaceous)
704	Interamnia	1910	3.06	310	C (carbonaceous)

The asteroids all revolve about the Sun in the same direction as the planets, and most of their orbits lie near the plane in which Earth and other planets circle. The majority of asteroids are in the asteroid belt, the region between Mars and Jupiter that contains all asteroids with orbital periods between 3.3 to 6 years (Figure 13.2). Although more than 75% of the known asteroids are in the belt, they are not closely spaced (as they are sometimes depicted in science fiction movies). The volume of the belt is actually very large, and the typical spacing between objects (down to 1 kilometer in size) is several million kilometers. (This was fortunate for spacecraft like Galileo, Cassini, *Rosetta*, and New Horizons, which needed to travel through the asteroid belt without a collision.)

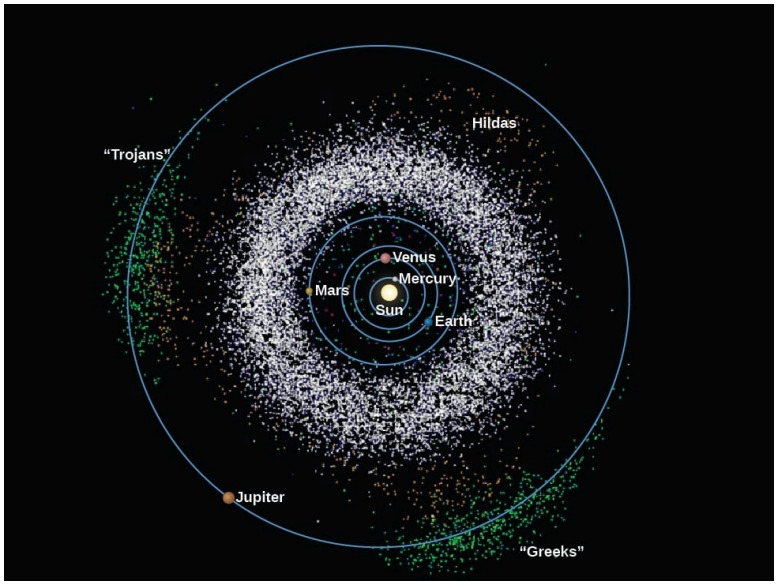


Figure 13.2 : Asteroids in the Solar System. This computer-generated diagram shows the positions of the asteroids known in 2006. If the asteroid sizes were drawn to scale, none of the dots representing an asteroid would be visible. Here, the asteroid dots are too big and give a false impression of how crowded the asteroid belt would look if you were in it. Note that in addition to those in the asteroid belt, there are also asteroids in the inner solar system and some along Jupiter's orbit (such as the Trojans and Greeks groups), controlled by the giant planet's gravity.

Still, over the long history of our solar system, there have been a good number of collisions among the asteroids themselves. In 1918, the Japanese astronomer Kiyotsugu Hirayama found that some asteroids fall into *families*, groups with similar orbital characteristics. He hypothesized that each family may have resulted from the breakup of a larger body or, more likely, from the

collision of two asteroids. Slight differences in the speeds with which the various fragments left the collision scene account for the small spread in orbits now observed for the different asteroids in a given family. Several dozen such families exist, and observations have shown that individual members of most families have similar compositions, as we would expect if they were fragments of a common parent.

Link to Learning

You can see a dramatic animated video showing the orbits of 100,000 asteroids found by one sky survey. As the 3-minute video goes on, you get to see the orbits of the planets and how the asteroids are distributed in the solar system. But note that all such videos are misleading in one sense. The asteroids themselves are really small compared to the distances covered, so they have to be depicted as larger points to be visible. If you were in the asteroid belt, there would be far more empty space than asteroids.

Composition and Classification

Asteroids are as different as black and white. The majority are very dark, with reflectivity of only 3 to 4%, like a lump of coal. However, another large group has a typical reflectivity of 15%. To understand more about these differences and how they are related to chemical composition, astronomers study the spectrum of the light reflected from asteroids for clues about their composition.

The dark asteroids are revealed from spectral studies to be *primitive* bodies (those that have changed little chemically since the beginning of the solar system) composed of silicates mixed with dark, organic carbon compounds. These are known as C-type asteroids (“C” for carbonaceous). Two of the largest asteroids, Ceres and Pallas, are primitive, as are almost all of the asteroids in the outer part of the belt. In 2020 and 2021, two of these C-type asteroids, Bennu and Ryugu, were visited by spacecraft that briefly touched their surfaces.

The second most populous group is the S-type asteroids, where “S” stands for a stony or silicate composition. Here, the dark carbon compounds are missing, resulting in higher reflectivity and clearer spectral signatures of silicate minerals. The S-type asteroids are also chemically primitive, but their different composition indicates that they were probably formed in a different location in the solar system from the C-type asteroids.

Asteroids of a third class, much less numerous than those of the first two, are composed primarily of metal and are called M-type asteroids (“M” for metallic). Spectroscopically, the identification of metal is difficult, but for at least the largest M-type asteroid, Psyche, this identification has been confirmed by radar. Since a metal asteroid, like an airplane or ship, is a much better reflector of radar than is a stony object, Psyche appears bright when we aim a radar beam at it. In 2022, a NASA mission, conveniently named Psyche, will be launched toward this asteroid.

How did such metal asteroids come to be? We suspect that each came from a parent body large enough for its molten interior to settle out or differentiate, and the heavier metals sank to the center. When this parent body shattered in a later collision, the fragments from the core were rich in metals. There is enough metal in even a 1-kilometer M-type asteroid to supply the world with iron and many other industrial metals for the foreseeable future, if we could bring one safely to Earth.

In addition to the M-type asteroids, a few other asteroids show signs of early heating and differentiation. These have basaltic surfaces like the volcanic plains of the Moon and Mars; the large asteroid Vesta (discussed in a moment) is in this last category.

The different classes of asteroids are found at different distances from the Sun (Figure 13.3). By tracing how asteroid compositions vary with distance from the Sun, we can reconstruct some of the properties of the solar nebula from which they originally formed.

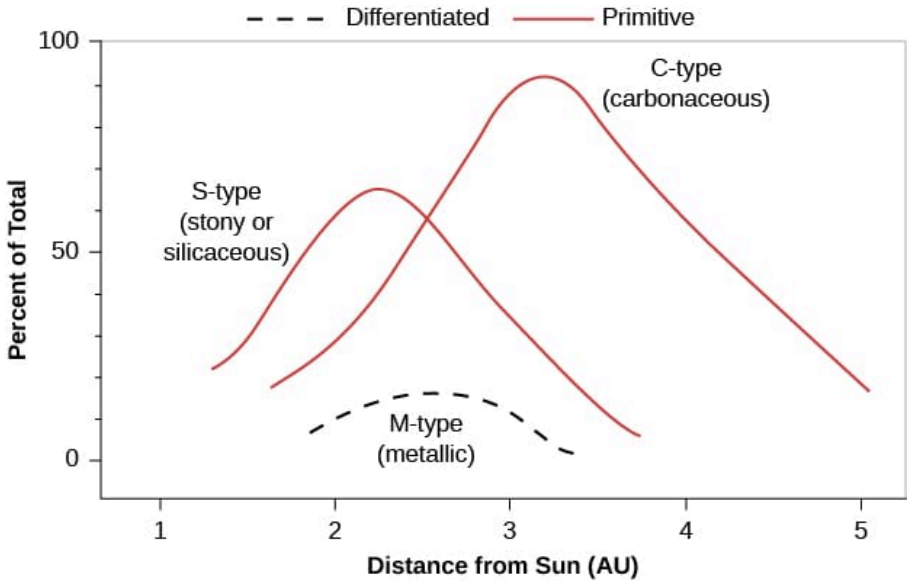


Figure 13.3 : Where Different Types of Asteroids Are Found. Asteroids of different composition are distributed at different distances from the Sun. The S-type and C-type are both primitive; the M-type consists of cores of differentiated parent bodies.

Vesta: A Differentiated Asteroid

Vesta is one of the most interesting of the asteroids. It orbits the Sun with a semi-major axis of 2.4 AU in the inner part of the asteroid belt. Its relatively high reflectivity of almost 30% makes it the brightest asteroid, so bright that it is actually visible to the unaided eye if you know just where to look. But its real claim to fame is that its surface is covered with basalt, indicating that Vesta is a differentiated object that must once have been volcanically active, in spite of its small size (about 500 kilometers in diameter).

Meteorites from Vesta's surface (Figure 13.4), identified by comparing their spectra with that of Vesta itself, have landed on Earth and are available for direct study in the laboratory. We thus know a great deal about this asteroid. The age of the lava flows from which these meteorites derived has been measured at 4.4 to 4.5 billion years, very soon after the formation of the solar system. This age is consistent with what we might expect for volcanoes on Vesta; whatever process heated such a small object was probably intense and short-lived. In 2016, a meteorite fell in Turkey that could be identified with a particular lava flow as revealed by the orbiting *Dawn* spacecraft.

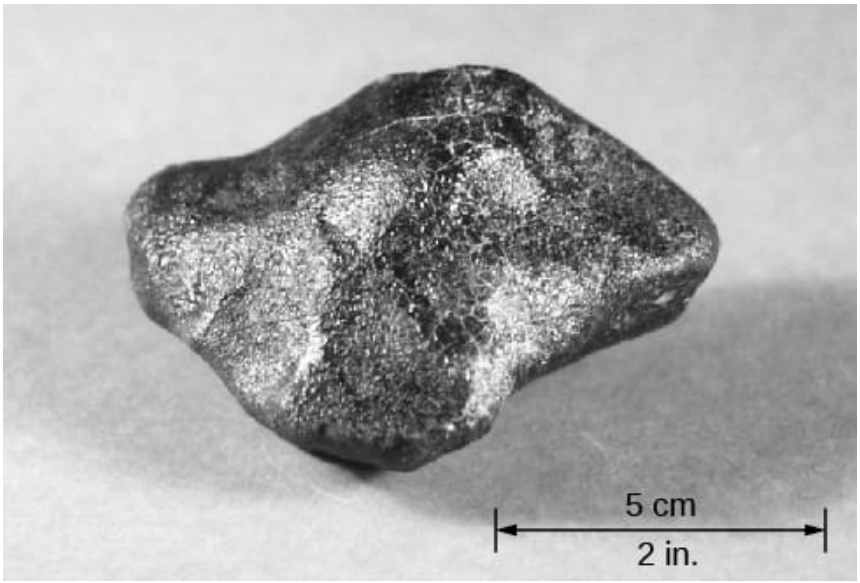


Figure 13.4 : Piece of Vesta. This meteorite (rock that fell from space) has been identified as a volcanic fragment from the crust of asteroid Vesta. (credit: modification of work by R. Kempton (New England Meteoritical Services))

Asteroids Up Close

On the way to its 1995 encounter with Jupiter, the Galileo spacecraft was targeted to fly close to two main-belt S-type asteroids called Gaspra and Ida. The Galileo camera revealed both as long and highly irregular (resembling a battered potato), as befits fragments from a catastrophic collision (Figure 13.5).

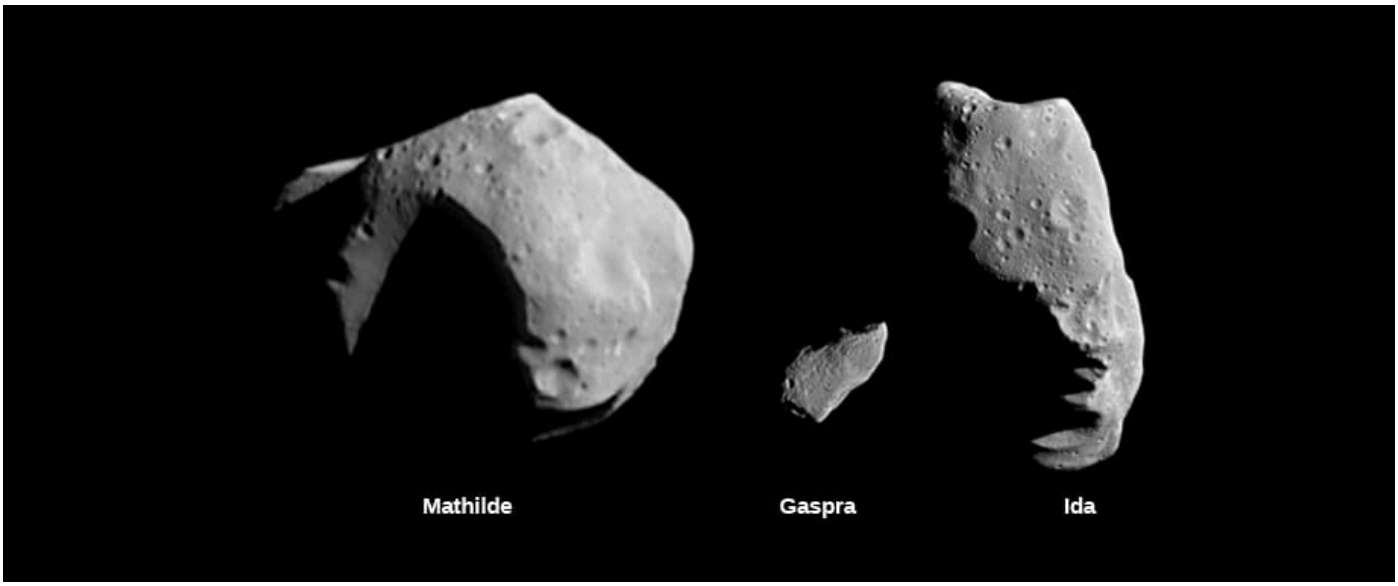


Figure 13.5 : Mathilde, Gaspra, and Ida. The first three asteroids photographed from spacecraft flybys, printed to the same scale. Gaspra and Ida are S-type and were investigated by the Galileo spacecraft; Mathilde is C-type and was a flyby target for the NEAR-Shoemaker spacecraft. (credit: modification of work by NEAR Project, Galileo Project, NASA)

The detailed images allowed us to count the craters on Gaspra and Ida, and to estimate the length of time their surfaces have been exposed to collisions. The Galileo scientists concluded that these asteroids are only about 200 million years old (that is, the collisions that formed them took place about 200 million years ago). Calculations suggest that an asteroid the size of Gaspra or Ida can expect another catastrophic collision sometime in the next billion years, at which time it will be disrupted to form another generation of still-smaller fragments.

The greatest surprise of the Galileo flyby of Ida was the discovery of a moon (which was then named Dactyl), in orbit about the asteroid (Figure 13.6). Although only 1.5 kilometers in diameter, smaller than many college campuses, Dactyl provides scientists with something otherwise beyond their reach—a measurement of the mass and density of Ida using Kepler's laws. The moon's distance of about 100 kilometers and its orbital period of about 24 hours indicate that Ida has a density of approximately 2.5 g/cm^3, which matches the density of primitive rocks. Subsequently, both large visible-light telescopes and high-powered planetary radar have discovered many other asteroid moons, so that we are now able to accumulate valuable data on asteroid masses and densities.



Figure 13.6 : Ida and Dactyl. The asteroid Ida and its tiny moon Dactyl (the small body off to its right), were photographed by the Galileo spacecraft in 1993. Irregularly shaped Ida is 56 kilometers in its longest dimension, while Dactyl is about 1.5 kilometers across. The colors have been intensified in this image; to the eye, all asteroids look basically gray. (credit: modification of work by NASA/JPL)

By the way, Phobos and Deimos, the two small moons of Mars, are probably captured asteroids (Figure 13.7). They were first studied at close range by the Viking orbiters in 1977 and later by *Mars Global Surveyor*. Both are irregular, somewhat elongated, and heavily created, resembling other smaller asteroids. Their largest dimensions are about 26 kilometers and 16 kilometers, respectively. The small outer moons of Jupiter and Saturn were probably also captured from passing asteroids, perhaps early in the history of the solar system.

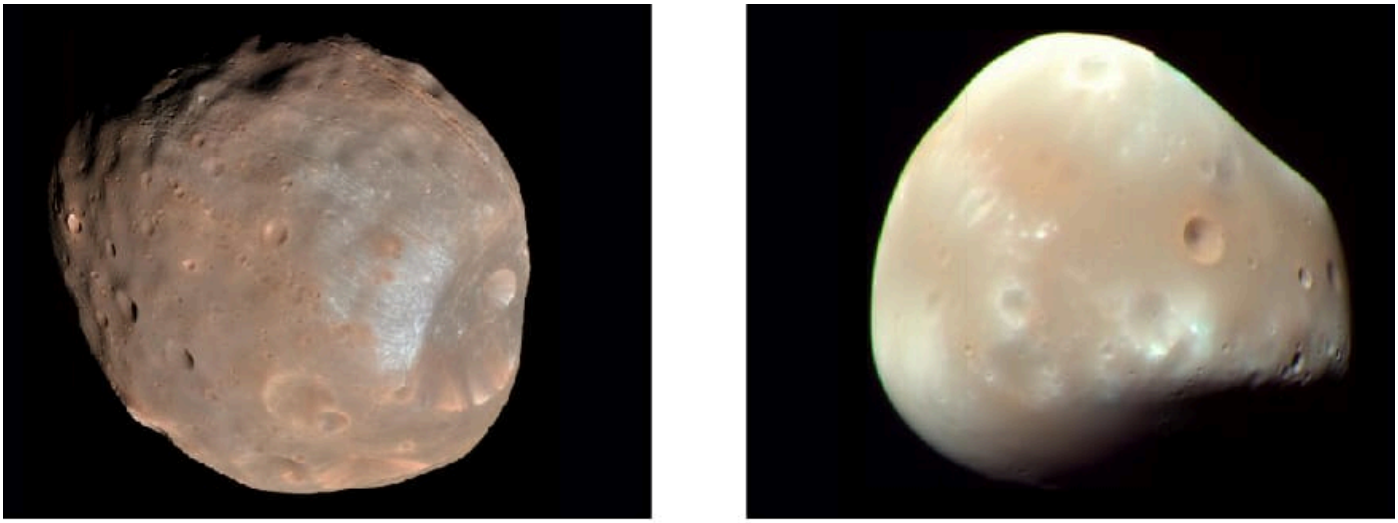


Figure 13.7 : Moons of Mars. The two small moons of Mars, (a) Phobos and (b) Deimos, were discovered in 1877 by American astronomer Asaph Hall. Their surface materials are similar to many of the asteroids in the outer asteroid belt, leading astronomers to believe that the two moons may be captured asteroids. (credit a: modification of work by NASA; credit b: modification of work by NASA/JPL-Caltech/University of Arizona)

Beginning in the 1990s, spacecraft have provided close looks at several more asteroids. The Near Earth Asteroid Rendezvous (NEAR) spacecraft went into orbit around the S-type asteroid Eros, becoming a temporary moon of this asteroid. On its way to Eros, the *NEAR* spacecraft was renamed after planetary geologist Eugene Shoemaker, a pioneer in our understanding of craters and impacts.

For a year, the NEAR-Shoemaker spacecraft orbited the little asteroid at various altitudes, measuring its surface and interior composition as well as mapping Eros from all sides (Figure 13.8). The data showed that Eros is made of some of the most

chemically primitive materials in the solar system. Several other asteroids have been revealed as made of loosely bound rubble throughout, but not Eros. Its uniform density (about the same as that of Earth's crust) and extensive global-scale grooves and ridges show that it is a cracked but solid rock.

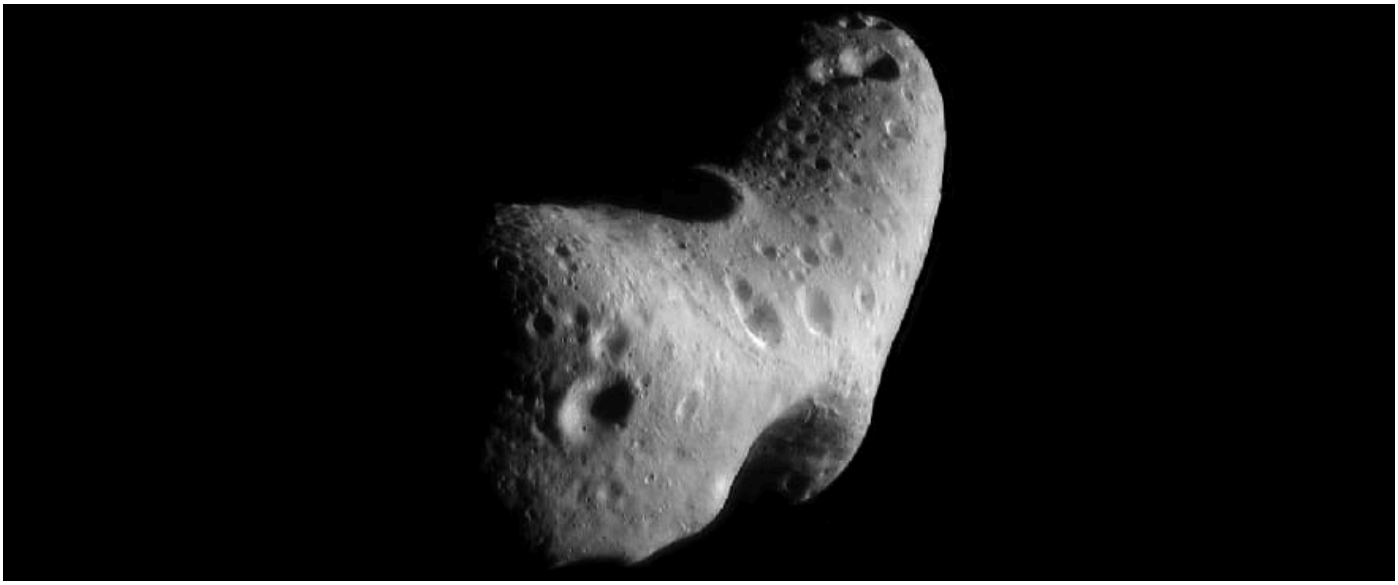


Figure 13.8 : Looking Down on the North Pole of Eros. This view was constructed from six images of the asteroid taken from an altitude of 200 kilometers. The large crater at the top has been named Psyche (after the maiden who was Eros' lover in classical mythology) and is about 5.3 kilometers wide. A saddle-shaped region can be seen directly below it. Craters of many different sizes are visible. (credit: modification of work by NASA/JHUPL)

Eros has a good deal of loose surface material that appears to have slid down toward lower elevations. In some places, the surface rubble layer is 100 meters deep. The top of loose soil is dotted with scattered, half-buried boulders. There are so many of these boulders that they are more numerous than the craters. Of course, with the gravity so low on this small world, a visiting astronaut would find loose boulders rolling toward her pretty slowly and could easily leap high enough to avoid being hit by one. Although the NEAR-Shoemaker spacecraft was not constructed as a lander, at the end of its orbital mission in 2000, it was allowed to fall gently to the surface, where it continued its chemical analysis for another week.

In 2003, Japan's Hayabusa 1 mission not only visited a small asteroid but also brought back samples to study in laboratories on Earth. The target S-type asteroid, Itokawa (shown in Figure 13.9), is much smaller than Eros, only about 500 meters long. This asteroid is elongated and appears to be the result of the collision of two separate asteroids long ago. There are almost no impact craters, but an abundance of boulders (like a pile of rubble) on the surface.

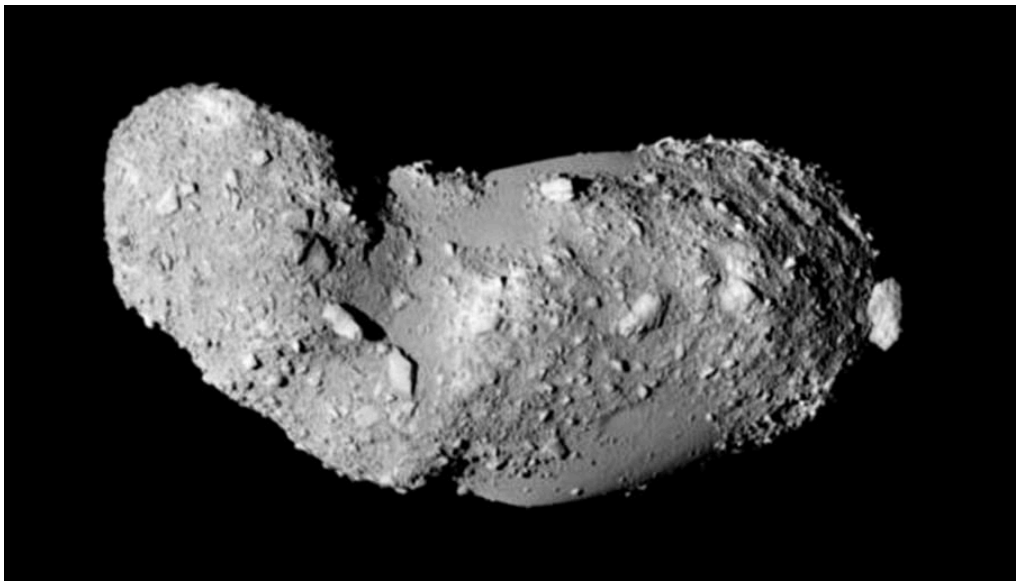


Figure 13.9 : Asteroid Itokawa. The surface of asteroid Itokawa appears to have no craters. Astronomers have hypothesized that its surface consists of rocks and ice chunks held together by a small amount of gravity, and its interior is probably also a similar rubble pile. (credit: JAXA)

The *Hayabusa* spacecraft was designed not to land, but to touch the surface just long enough to collect a small sample. This tricky maneuver failed on its first try, with the spacecraft briefly toppling over on its side. Eventually, the controllers were successful in picking up a few grains of surface material and transferring them into the return capsule (Figure 13.10). We estimate that the dust grains *Hayabusa* picked up had been exposed on the surface of the asteroid for about 8 million years.

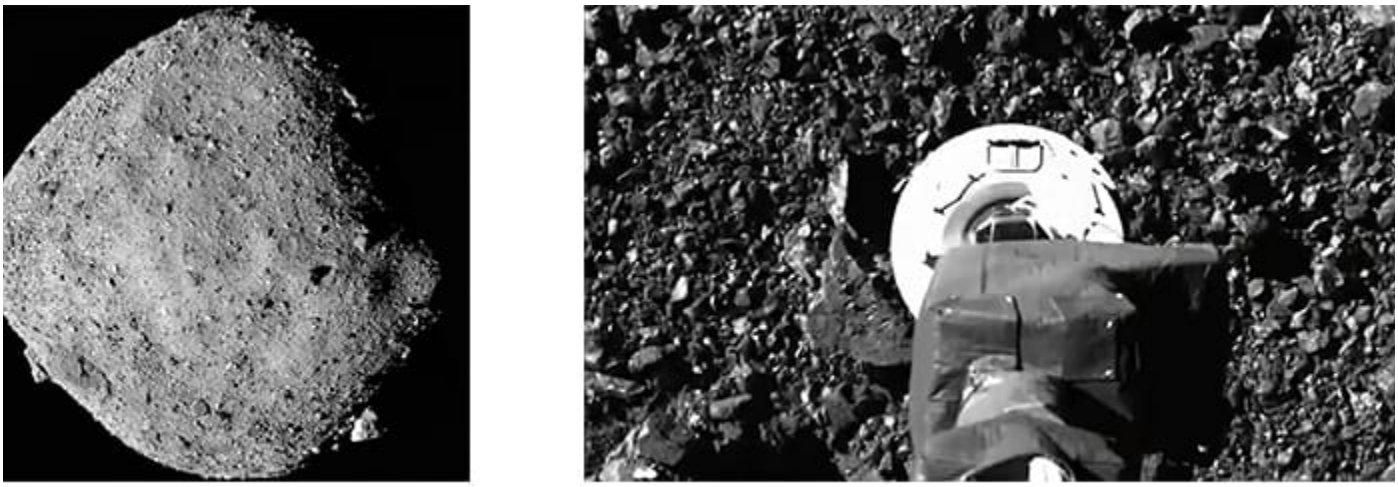


Figure 13.10 : Examining an Asteroid Up Close. (a) Near-Earth Asteroid Bennu, imaged by the Osiris-Rex Spacecraft in 2020. Bennu is a small (half-km diameter), rapidly spinning dark asteroid. Its low density suggests it is a loosely-bound rubble pile. The surface is very rough, with protruding rocks and boulders. (b) Preparing to collect a sample, the collection arm is shown just before touch-down on Bennu. For scale, the diameter of the sampling device is 30 cm. The sample of about 100 grams of surface rocks and dust will be returned to Earth in 2023. a: modification of work by NASA Goddard/University of Arizona; credit b: modification of work by NASA Goddard/University of Arizona

Following on the success of Hayabusa, both NASA and the Japanese space agency, JAXA, sent more advanced asteroid sample-return missions. Both targeted small Earth-approaching asteroids, only about 1 km in diameter. Each target is a dark object, apparently a rubble pile formed relatively recently after being broken apart in a collision. Hayabusa-2 visited the asteroid Ryugu, collected a small sample, and successfully returned it to Earth in December 2020. The U.S. mission called OSIRIS-REx visited Bennu, an asteroid that actually has a small chance of colliding with Earth in the next century. Its sample will be returned to Earth in September 2023.

The most ambitious asteroid space mission (called Dawn) has visited the two largest main belt asteroids, Ceres and Vesta,

orbiting each for about a year (Figure 13.11). Their large sizes (diameters of about 1000 and 500 kilometers, respectively) make them appropriate for comparison with the planets and large moons. Both turned out to be heavily cratered, implying their surfaces are old. On Vesta, we have now actually located the large impact craters that ejected the basaltic meteorites previously identified as coming from this asteroid. These craters are so large that they sample several layers of Vesta's crustal material.

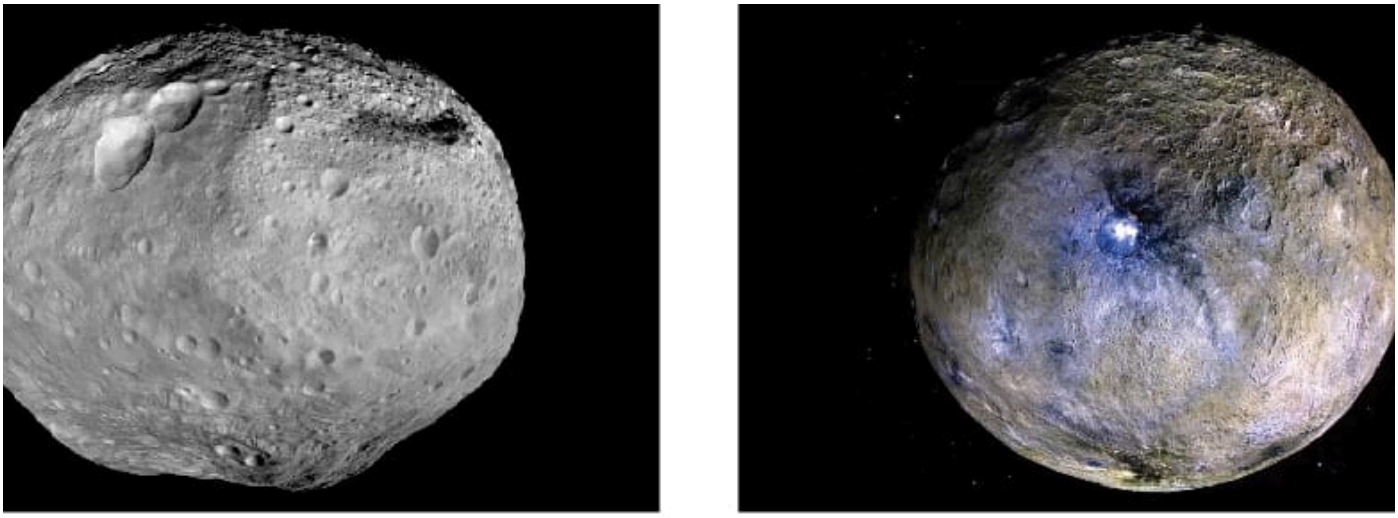


Figure 13.11 : Vesta and Ceres. The NASA *Dawn* spacecraft took these images of the large asteroids (a) Vesta and (b) Ceres. (a) Note that Vesta is not round, as Ceres (which is considered a dwarf planet) is. A mountain twice the height of Mt. Everest on Earth is visible at the very bottom of the Vesta image. (b) The image of Ceres has its colors exaggerated to bring out differences in composition. You can see a white feature in Occator crater near the center of the image. (credit a, b: modification of work by NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

Ceres has not had a comparable history of giant impacts, so its surface is covered with craters that look more like those from the lunar highlands. One big surprise at Ceres is the presence of very bright white spots, associated primarily with the central peaks of large craters (Figure 13.12). The light-colored mineral is primarily salt, released from the interior. After repeated close flybys, data from the NASA Dawn spacecraft indicated that Ceres has (or has had) a subsurface ocean of water, with occasional eruptions on the surface. The most dramatic is the 4 kilometer tall ice volcano called Ahuna Mons (see Figure 13.12).

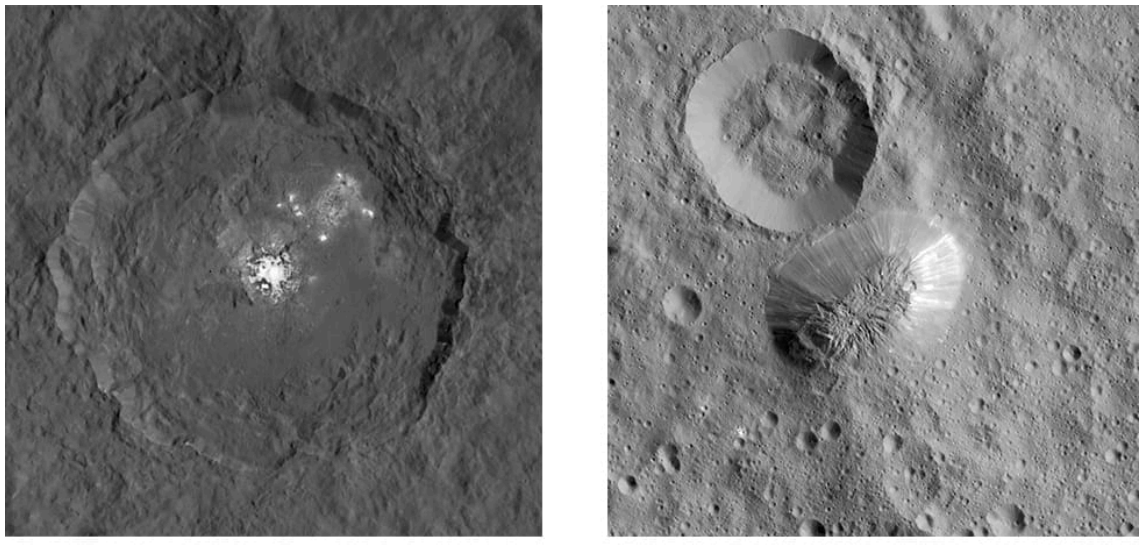


Figure 13.12 : White Spots in a Larger Crater on Ceres. White Spots in a Larger Crater on Ceres. (a) These bright features appear to be salt deposits in a Ceres crater called Occator, which is 92 kilometers across. (b) Ahuna Mons is an isolated mountain on Ceres, 4 kilometers high. It is thought to be an intrusion of ice from the interior. (credit a: modification of work by NASA/JPL-Caltech/UCLA/MPS/DLR/IDA; credit b: modification of work by NASA/JPL-Caltech/UCLA/MPS/DLR/IDA/PSI)

In late 2017, something entirely new was discovered: an interstellar asteroid. This visitor was found at a distance of 33 million kilometers with a survey telescope on Haleakala, Hawaii. As astronomers followed up on the discovery, it quickly became apparent that this asteroid was travelling far too fast to be part of the Sun's family. Its orbit is a hyperbola, and when discovered it was already rapidly leaving the inner solar system. Although it was too distant for imaging by even large telescopes, its size and shape could be estimated from its brightness and rapid light fluctuations. It is highly elongated, with an approximately cylindrical shape. The nominal dimensions are about 200 meters in length and only 35 meters across, the most extreme of any natural object. Large objects, like planets and moons, are pulled by their own gravity into roughly spherical shapes, and even small asteroids and comets (often described as "potato-shaped") rarely have irregularities of more than a factor of two.

This asteroid was named 'Oumuamua, a Hawaiian word meaning "scout" or "first to reach out." In a way, the discovery of an interstellar asteroid or comet was not unexpected. Early in solar system history, before the planet orbits sorted themselves into stable, non-intersecting paths all in the same plane, we estimate that quite a lot of mass was ejected, either whole planets or more numerous smaller fragments. Even today, an occasional comet coming in from the outer edges of the solar system can have its orbit changed by gravitational interaction with Jupiter and the Sun, and some of these escape on hyperbolic trajectories. In 2019, astronomers discovered an interstellar comet entering our solar system—more on this in the section on comets. As we have recently learned that planetary systems are common, the question became: where are similar debris objects ejected from other planetary systems? Now we have found two, and improved surveys will soon add others to this category.

Link to Learning

View an artist's rendering of the asteroid 'Oumuamua (www.openstax.org/l/30oumuamua) by the ESO. Although it was not close enough to Earth to be imaged, its long slender shape was indicated by its rapid variation in brightness as it rotated.

Link to Learning

The space agencies involved with the Dawn mission have produced nice animated "flyover" videos of Vesta and Ceres available online.

This page titled 10.1: Asteroids is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 13.2: Asteroids by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10.2: Meteors

Learning Objectives

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

- Explain what a meteor is and why it is visible in the night sky
- Describe the origins of meteor showers

As we saw in Comets and Asteroids: Debris of the Solar System, the ices in comets evaporate when they get close to the Sun, together spraying millions of tons of rock and dust into the inner solar system. There is also dust from asteroids that have collided and broken up. Earth is surrounded by this material. As each of the larger dust or rock particles enters Earth's atmosphere, it creates a brief fiery trail; this is often called a *shooting star*, but it is properly known as a meteor.

Observing Meteors

Meteors are created as tiny solid particles that enter Earth's atmosphere from interplanetary space. Since the particles move at speeds of many kilometers per second, friction with the air vaporizes them at altitudes between 80 and 130 kilometers. The resulting flashes of light fade out within a few seconds. These "shooting stars" got their name because at night their luminous vapors look like stars moving rapidly across the sky. To be visible, a meteor must be within about 200 kilometers of the observer. On a typical dark, moonless night, an alert observer can see half a dozen meteors per hour. These *sporadic meteors*—those not associated with a meteor shower (explained in the next section)—are random occurrences. Over the entire Earth, the total number of meteors bright enough to be visible totals about 25 million per day.

The typical meteor is produced by a particle with a mass of less than 1 gram—no larger than a pea. How can we see such a small particle? The light you see comes from the much larger region of heated, glowing gas surrounding this little grain of interplanetary material. Because of its high speed, the energy in a pea-sized meteor is as great as that of an artillery shell fired on Earth, but this energy is dispersed high in Earth's atmosphere. (When these tiny projectiles hit an airless body like the Moon, they do make small craters and generally pulverize the surface.)

If a particle the size of a golf ball strikes our atmosphere, it produces a much brighter trail called a fireball (Figure 14.2). A piece as large as a bowling ball has a fair chance of surviving its fiery entry if its approach speed is not too high. The total mass of meteoric material entering Earth's atmosphere is estimated to be about 100 tons per day (which seems like a lot if you imagine it all falling in one place, but remember it is spread out all over our planet's surface).



Figure 14.2 : Fireball. When a larger piece of cosmic material strikes Earth's atmosphere, it can make a bright fireball. This time-lapse meteor image was captured in April 2014 at the Atacama Large Millimeter/Submillimeter Array (ALMA). The visible trail results from the burning gas around the particle. (credit: modification of work by ESO/C Malin)

Link to Learning

While it is difficult to capture images of fireballs and other meteors with still photography, it's easy to capture the movement of these objects on video. The American Meteor Society maintains a website on which their members can share such videos.

Meteor Showers

Many—perhaps most—of the meteors that strike Earth are associated with specific comets. Some of these periodic comets still return to our view; others have long ago fallen apart, leaving only a trail of dust behind them. The dust particles from a given comet retain approximately the orbit of their parent, continuing to move together through space but spreading out over the orbit with time. When Earth, in its travels around the Sun, crosses such a dust stream, we see a sudden burst of meteor activity that usually lasts several hours; such an event is called a meteor shower.

The dust particles and pebbles that produce meteor showers are moving together in space before they encounter Earth. Thus, as we look up at the atmosphere, their parallel paths seem to come toward us from a place in the sky called the *radiant*. This is the direction in space from which the meteor stream seems to be diverging, just as long railroad tracks seem to diverge from a single spot on the horizon (Figure 14.3). Meteor showers are often designated by the constellation in which this radiant is located: for example, the Perseid meteor shower has its radiant in the constellation of Perseus. But you are likely to see shower meteors anywhere in the sky, not just in the constellation of the radiant. The characteristics of some of the more famous meteor showers are summarized in Table 14.1.

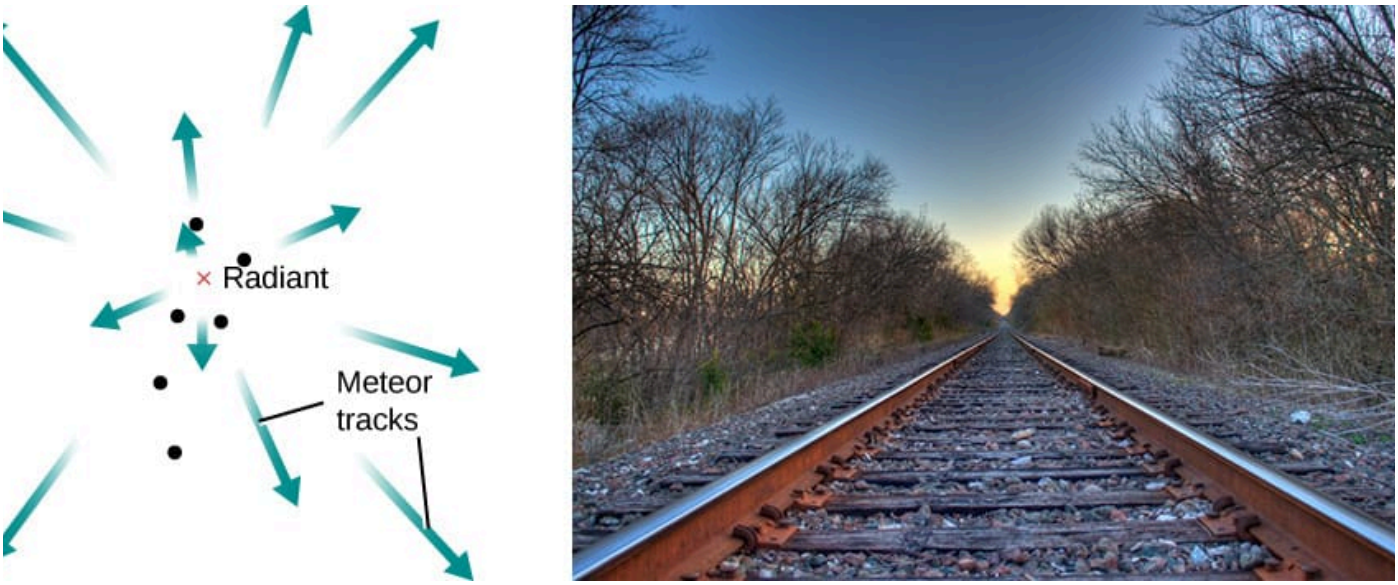


Figure 14.3 : Radiant of a Meteor Shower. The tracks of the meteors diverge from a point in the distance, just as long, parallel railroad tracks appear to do. (credit “tracks”: Nathan Vaughn)

Table 14.1: Major Annual Meteor Showers

Shower Name	Date of Maximum	Associated Parent Object	Comet’s Period (years)
Quadrantid	January 3–4	2003EH (asteroid)	—
Lyrid	April 22	Comet Thatcher	415
Eta Aquarid	May 4–5	Comet Halley	76
Delta Aquarid	July 29–30	Comet Machholz	—
Perseid	August 11–12	Comet Swift-Tuttle	133
Orionid	October 20–21	Comet Halley	76
Southern Taurid	October 31	Comet Encke	3
Leonid	November 16–17	Comet Tempel-Tuttle	33
Geminid	December 13	Phaethon (asteroid)	1.4

The meteoric dust is not always evenly distributed along the orbit of the comet, so during some years more meteors are seen when Earth intersects the dust stream, and in other years fewer. For example, a very clumpy distribution is associated with the Leonid meteors, which in 1833 and again in 1866 (after an interval of 33 years—the period of the comet) yielded the most spectacular showers (sometimes called *meteor storms*) ever recorded (Figure 14.4). During the Leonid storm on November 17, 1866, up to a hundred meteors were observed per second in some locations. The Leonid shower of 2001 was not this intense, but it peaked at nearly a thousand meteors per hour—one every few seconds—observable from any dark viewing site.



Figure 14.4 : Leonid Meteor Storm. A painting depicts the great meteor shower or storm of 1833, shown with a bit of artistic license.

The most dependable annual meteor display is the Perseid shower, which appears each year for about three nights near August 11. In the absence of bright moonlight, you can see one meteor every few minutes during a typical Perseid shower. Astronomers estimate that the total combined mass of the particles in the Perseid swarm is nearly a billion tons; the comet that gave rise to the particles in that swarm, called Swift-Tuttle, must originally have had at least that much mass. However, if its initial mass were comparable to the mass measured for Comet Halley, then Swift-Tuttle would have contained several hundred billion tons, suggesting that only a very small fraction of the original cometary material survives in the meteor stream.

Link to Learning

The California Academy of Sciences has a short animated guide on "How to Observe a Meteor Shower."

No shower meteor has ever survived its flight through the atmosphere and been recovered for laboratory analysis. However, there are other ways to investigate the nature of these particles and thereby gain additional insight into the comets from which they are derived. Analysis of the flight paths of meteors shows that most of them are very light or porous, with densities typically less than 1.0 g/cm^3. If you placed a fist-sized lump of meteor material on a table in Earth's gravity, it might well fall apart under its own weight.

Such light particles break up very easily in the atmosphere, accounting for the failure of even relatively large shower meteors to reach the ground. Comet dust is apparently fluffy, rather inconsequential stuff. NASA's Stardust mission used a special substance, called aerogel, to collect these particles. We can also infer this from the tiny comet particles recovered in Earth's atmosphere with high-flying aircraft (see Figure 13.19). This fluff, by its very nature, cannot reach Earth's surface intact. However, more substantial fragments from asteroids do make it into our laboratories, as we will see in the next section.

Seeing for Yourself: Showering with the Stars

Observing a meteor shower is one of the easiest and most enjoyable astronomy activities for beginners (Figure 14.5). The best thing about it is that you don't need a telescope or binoculars—in fact, they would positively get in your way. What you do need is a site far from city lights, with an unobstructed view of as much sky as possible. While the short bright lines in the sky made by individual meteors could, in theory, be traced back to a radiant point (as shown in Figure 14.3), the quick blips of light that represent the end of the meteor could happen anywhere above you.



Figure 14.5 : Perseid Meteor Shower. This twenty-second exposure shows a meteor during the 2015 Perseid meteor shower. (credit: NASA/Bill Ingalls)

The key to observing meteor showers is not to restrict your field of view, but to lie back and scan the sky alertly. Try to select a good shower (see the list in Table 14.1) and a night when the Moon will not be bright at the time you are observing. The Moon, street lights, vehicle headlights, bright flashlights, and cell phone and tablet screens will all get in the way of your seeing the faint meteor streaks.

You will see more meteors after midnight, when you are on the hemisphere of Earth that faces forward—in the direction of Earth's revolution around the Sun. Before midnight, you are observing from the "back side" of Earth, and the only meteors you see will be those that traveled fast enough to catch up with Earth's orbital motion.

When you've gotten away from all the lights, give your eyes about 15 minutes to get "dark adapted"—that is, for the pupils of your eyes to open up as much as possible. (This adaptation is the same thing that happens in a dark movie theater. When you first enter, you can't see a thing, but eventually, as your pupils open wider, you can see pretty clearly by the faint light of the screen—and notice all that spilled popcorn on the floor.)

Seasoned meteor observers find a hill or open field and make sure to bring warm clothing, a blanket, and a thermos of hot coffee or chocolate with them. (It's also nice to take along someone with whom you enjoy sitting in the dark.) Don't expect to see fireworks or a laser show: meteor showers are subtle phenomena, best approached with a patience that reflects the fact that some of the dust you are watching burn up may first have been gathered into its parent comet more than 4.5 billion years ago, as the solar system was just forming.

This page titled 10.2: Meteors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 14.2: Meteors by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10.3: Meteorites - Stones from Heaven

Learning Objectives

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

- Explain the origin of meteorites and the difference between a meteor and a meteorite
- Describe how most meteorites have been found
- Explain how primitive stone meteorites are significantly different from other types
- Explain how the study of meteorites informs our understanding of the age of the solar system.

Any fragment of interplanetary debris that survives its fiery plunge through Earth's atmosphere is called a meteorite. Meteorites fall only very rarely in any one locality, but over the entire Earth thousands fall each year. Some meteorites are loners, but many are fragments from the breakup in the atmosphere of a single larger object. These rocks from the sky carry a remarkable record of the formation and early history of the solar system.

Extraterrestrial Origin of Meteorites

Occasional meteorites have been found throughout history, but their extraterrestrial origin was not accepted by scientists until the beginning of the nineteenth century. Before that, these strange stones were either ignored or considered to have a supernatural origin.

The falls of the earliest recovered meteorites are lost in the fog of mythology. A number of religious texts speak of stones from heaven, which sometimes arrived at opportune moments to smite the enemies of the authors of those texts. At least one sacred meteorite has apparently survived in the form of the Ka'aba, the holy black stone in Mecca that is revered by Islam as a relic from the time of the Patriarchs—although understandably, no chip from this sacred stone has been subject to detailed chemical analysis.

The modern scientific history of the meteorites begins in the late eighteenth century, when a few scientists suggested that some strange-looking stones had such peculiar composition and structure that they were probably not of terrestrial origin. The idea that indeed "stones fall from the sky" was generally accepted only after a scientific team led by French physicist Jean-Baptiste Biot investigated a well-observed fall in 1803.

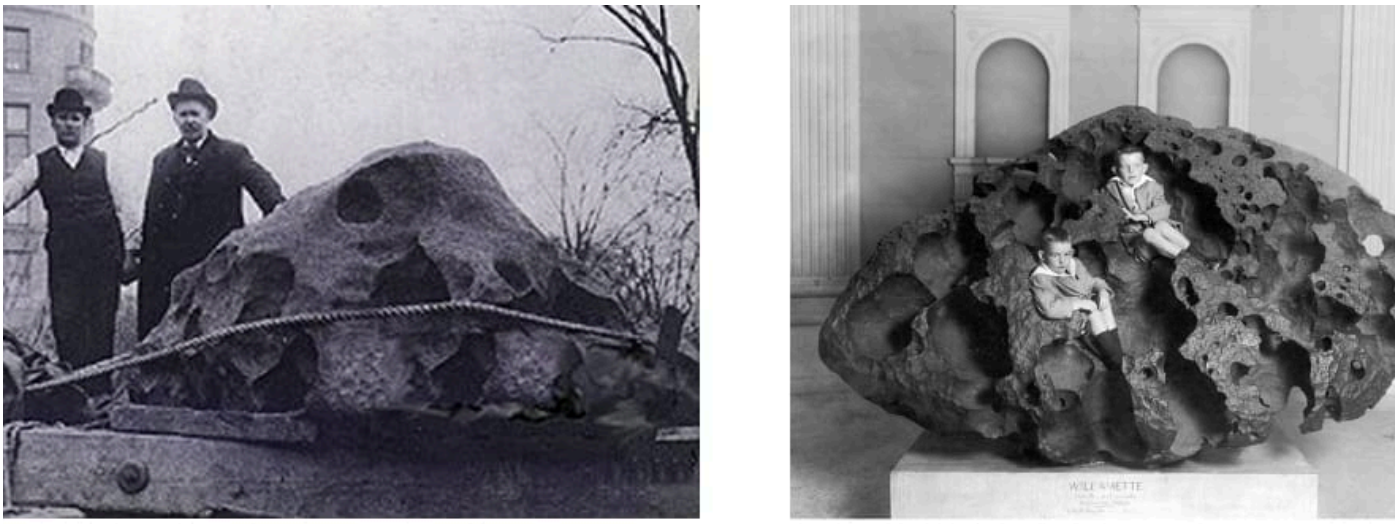
Meteorites sometimes fall in groups or showers. Such a fall occurs when a single larger object breaks up during its violent passage through the atmosphere. It is important to remember that such a *shower of meteorites* has nothing to do with a *meteor shower*. No meteorites have ever been recovered in association with meteor showers. Whatever the ultimate source of the meteorites, they do not appear to come from the comets or their associated particle streams.

Meteorite Falls and Finds

Meteorites are found in two ways. First, sometimes bright meteors (fireballs) are observed to penetrate the atmosphere to low altitudes. If we search the area beneath the point where the fireball burned out, we may find one or more remnants that reached the ground. Observed *meteorite falls*, in other words, may lead to the recovery of fallen meteorites. (A few meteorites have even hit buildings or, very rarely, people; see Making Connections: Some Striking Meteorites). The 2013 Chelyabinsk fireball, which we discussed in the chapter on Comets and Asteroids: Debris of the Solar System, produced tens of thousands of small meteorites, many of them easy to find because these dark stones fell on snow.

There are, however, many false alarms about meteorite falls. Most observers of a bright fireball conclude that part of it hit the ground, but that is rarely the case. Every few months news outlets report that a meteorite has been implicated in the start of a fire. Such stories have always proved to be wrong. The meteorite is ice-cold in space, and most of its interior remains cold even after its brief fiery plunge through the atmosphere. A freshly fallen meteorite is more likely to acquire a coating of frost than to start a fire.

People sometimes discover unusual-looking rocks that turn out to be meteoritic; these rocks are termed *meteorite finds*. Now that the public has become meteorite-conscious, many unusual fragments, not all of which turn out to be from space, are sent to experts each year. Some scientists divide these objects into two categories: “meteorites” and “meteorwrongs.” Outside Antarctica (see the next paragraph), genuine meteorites turn up at an average rate of 25 or so per year. Most of these end up in natural history museums or specialized meteoritical laboratories throughout the world, although there is a growing group of private collectors interested in them as well. (Figure 14.6)



(a) (b)

Figure 14.6 : Meteorite Find. (a) This early twentieth century photo shows a 15-ton iron meteorite found in the Willamette Valley in Oregon. Although known to Native Americans in the area, it was “discovered” by an enterprising local farmer in 1902, who proceeded to steal it and put it on display. (b) It was eventually purchased for the American Museum of Natural History and is now on display in the museum’s Rose Center in New York City as the largest iron meteorite in the United States. In this 1911 photo, two young boys are perched in the meteor’s crevices.

Since the 1980s, sources in the Antarctic have dramatically increased our knowledge of meteorites. More than ten thousand meteorites have been recovered from the Antarctic as a result of the motion of the ice in some parts of that continent (Figure 14.7). Meteorites that fall in regions where ice accumulates are buried and then carried slowly to other areas where the ice is gradually worn away. After thousands of years, the rock again finds itself on the surface, along with other meteorites carried to these same locations. The ice thus concentrates the meteorites that have fallen both over a large area and over a long period of time. Once on the surface, the rocks stand out in contrast to the ice and are thus easier to spot than in other places on our rocky planet.



(a) (b)

Figure 14.7 : Antarctic Meteorite. (a) The US Antarctic Search for Meteorites (ANSMET) team recovers a meteorite from the Antarctic ice during a 2001–2002 mission. (b) The team is shown with some of the equipment used in the search. (credit a, b: modification of work by NASA)

Making Connections: Some Striking Meteorites

Although meteorites fall regularly onto Earth's surface, few of them have much of an impact on human civilization. There is so much water and uninhabited land on our planet that rocks from space typically fall where no one even sees them come down. But given the number of meteorites that land each year, you may not be surprised that a few have struck buildings, cars, and even people. In September 1938, for example, a meteorite plunged through the roof of Edward McCain's garage, where it became embedded in the seat of his Pontiac Coupe (Figure 14.8).

In November 1982, Robert and Wanda Donahue of Wethersfield, Connecticut, were watching *M*A*S*H** on television when a 6-pound meteorite came thundering through their roof, making a hole in the living room ceiling. After bouncing, it finally came to rest under their dining room table.

Eighteen-year-old Michelle Knapp of Peekskill, New York, got quite a surprise one morning in October 1992. She had just purchased her very first car, her grandmother's 1980 Chevy Malibu. But she awoke to find its rear end mangled and a crater in the family driveway—thanks to a 3-pound meteorite. Michelle was not sure whether to be devastated by the loss of her car or thrilled by all the media attention.

In June 1994, Jose Martin and his wife were driving from Madrid, Spain, to a golfing vacation when a fist-sized meteorite crashed through the windshield of their car, bounced off the dashboard, broke Jose's little finger, and then landed in the back seat. Before Martin, the most recent person known to have been struck by a meteorite was Annie Hodges of Sylacauga, Alabama. In November 1954, she was napping on a couch when a meteorite came through the roof, bounced off a large radio set, and hit her first on the arm and then on the leg.

The fireball that exploded at an altitude of about 20 kilometers near the Russian city of Chelyabinsk on February 15, 2013, produced a very large meteorite shower, and quite a few of the small rocks hit buildings. None is known to have hit people, however, and the individual meteorites were so small that they did not do much damage—much less than the shockwave from the exploding fireball, which broke the glass in thousands of windows.



Figure 14.8 : Benld Meteorite. A meteorite (inset) left a hole in the seat cushion of Edward McCain's car. (credit: "Shsilver"/Wikimedia Commons)

Meteorite Classification

The meteorites in our collections have a wide range of compositions and histories, but traditionally they have been placed into three broad classes. First are the irons, composed of nearly pure metallic nickel-iron. Second are the stones, the term used for any silicate or rocky meteorite. Third are the rarer stony-irons, made (as the name implies) of mixtures of stone and metallic iron (Figure 14.9).

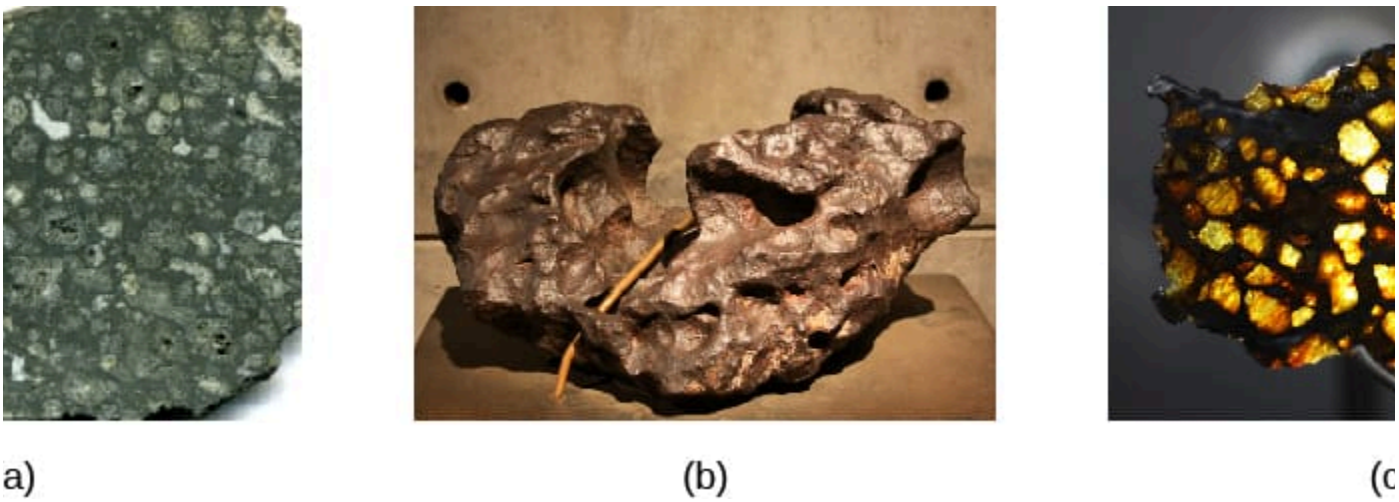


Figure 14.9 : Meteorite Types. (a) This piece of the Allende carbonaceous meteorite has white inclusions that may date back to before the formation of the solar nebula. (b) This fragment is from the iron meteorite responsible for the formation of Meteor Crater in Arizona. (c) This piece of the Imilac stony-iron meteorite is a beautiful mixture of green olivine crystals and metallic iron. (credit a: modification of work by James St. John; credit b: modification of work by "Taty2007"/Wikimedia Commons; credit c: modification of work by Juan Manuel Fluxà)

Of these three types, the irons and stony-irons are the most obviously extraterrestrial because of their metallic content. Pure iron almost never occurs naturally on Earth; it is generally found here as an oxide (chemically combined with oxygen) or other mineral ore. Therefore, if you ever come across a chunk of metallic iron, it is sure to be either human-made or a meteorite.

The stones are much more common than the irons but more difficult to recognize. Often laboratory analysis is required to demonstrate that a particular sample is really of extraterrestrial origin, especially if it has lain on the ground for some time and been subject to weathering. The most scientifically valuable stones are those collected immediately after they fall, or the Antarctic samples preserved in a nearly pristine state by ice.

Table 14.2 summarizes the frequencies of occurrence of the different classes of meteorites among the fall, find, and Antarctic categories.

Table 14.2: Frequency of Occurrence of Meteorite Classes

Class	Falls (%)	Finds (%)	Antarctic (%)
Primitive stones	88	51	85
Differentiated stones	8	2	12
Irons	3	42	2
Stony-irons	1	5	1

Ages and Compositions of Meteorites

It was not until the ages of meteorites were measured and their compositions analyzed in detail that scientists appreciated their true significance. The meteorites include the oldest and most primitive materials available for direct study in the laboratory. The ages of stony meteorites can be determined from the careful measurement of radioactive isotopes and their decay products. Almost all meteorites have radioactive ages between 4.50 and 4.56 billion years, as old as any ages we have measured in the solar system. The few younger exceptions are igneous rocks that have been ejected from cratering events on the Moon or Mars (and have made their way to Earth).

The average age for the most primitive meteorites, calculated using the most accurate values now available for radioactive half-lives, is 4.56 billion years, with an uncertainty of less than 0.01 billion years. This value (which we round off to 4.5 billion years in this book) is taken to represent the *age of the solar system*—the time since the first solids condensed and began to form into larger bodies.

The traditional classification of meteorites into irons, stones, and stony-irons is easy to use because it is obvious from inspection which category a meteorite falls into (although it may be much more difficult to distinguish a meteoritic stone from a terrestrial rock). More scientifically significant, however, is the distinction between *primitive* and *differentiated* meteorites. The differentiated meteorites are fragments of larger parent bodies that were molten before they broke up, allowing the denser materials (such as metals) to sink to their centers. Like many rocks on Earth, they have been subject to a degree of chemical reshuffling, with the different materials sorted according to density. Differentiated meteorites include the irons, which come from the metal cores of their parent bodies; stony-irons, which probably originate in regions between a metal core and a stony mantle; and some stones that are composed of mantle or crust material from their differentiated parent bodies.

The Most Primitive Meteorites

For information on the *earliest* history of the solar system, we turn to the primitive meteorites—those made of materials that have *not* been subject to great heat or pressure since their formation. We can look at the spectrum of sunlight reflected from asteroids and compare their compositions with those of primitive meteorites. Such analysis indicates that their parent bodies are almost certainly asteroids. Since asteroids are believed to be fragments left over from the formation process of the solar system, it makes sense that they should be the parent bodies of the primitive meteorites.

The great majority of the meteorites that reach Earth are primitive stones. Many of them are composed of light-colored gray silicates with some metallic grains mixed in, but there is also an important group of darker stones called *carbonaceous meteorites*. As their name suggests, these meteorites contain carbon, but we also find various complex organic molecules in them—chemicals based on carbon, which on Earth are the chemical building blocks of life. In addition, some of them contain chemically bound water, and many are depleted in metallic iron. The carbonaceous (or C-type) asteroids are concentrated in the outer part of the asteroid belt.

Among the most useful of these meteorites have been the Allende meteorite that fell in Mexico (see Figure 14.9), the Murchison meteorite that fell in Australia (both in 1969), and the Tagish Lake meteorite that landed in a winter snowdrift on Tagish Lake, Canada, in 2000. (The fragile bits of dark material from the Tagish Lake meteorite were readily visible against the white snow, although at first they were mistaken for wolf droppings.)

The Murchison meteorite (Figure 14.10) is known for the variety of organic chemicals it has yielded. Most of the carbon

compounds in carbonaceous meteorites are complex, tarlike substances that defy exact analysis. Murchison also contains 16 amino acids (the building blocks of proteins), 11 of which are rare on Earth. The most remarkable thing about these organic molecules is that they include equal numbers with right-handed and left-handed molecular symmetry. Amino acids can have either kind of symmetry, but all life on Earth has evolved using only the *left-handed* versions to make proteins. The presence of both kinds of amino acids clearly demonstrates that the ones in the meteorites had an extraterrestrial origin.



Figure 14.10 : Murchison Meteorite. A fragment of the meteorite that fell near the small town of Murchison, Australia, is shown next to a small sample of its material in a test tube, used for analysis of its chemical makeup.

These naturally occurring amino acids and other complex organic molecules in Murchison—formed without the benefit of the sheltering environment of planet Earth—show that a great deal of interesting chemistry must have taken place when the solar system was forming. If so, then perhaps some of the molecular building blocks of life on Earth were first delivered by primitive meteorites and comets. This is an interesting idea because our planet was probably much too hot for any organic materials to survive its earliest history. But after Earth's surface cooled, the asteroid and comet fragments that pelted it could have refreshed its supply of organic materials.

This page titled 10.3: Meteorites - Stones from Heaven is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 14.3: Meteorites - Stones from Heaven by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10.4: The "Long-Haired" Comets

Learning Objectives

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

- Characterize the general physical appearance of comets
- Explain the range of cometary orbits
- Describe the size and composition of a typical comet's nucleus
- Discuss the atmospheres of comets
- Summarize the discoveries of the Rosetta mission

Comets differ from asteroids primarily in their icy composition, a difference that causes them to brighten dramatically as they approach the Sun, forming a temporary atmosphere. In some early cultures, these so-called "hairy stars" were considered omens of disaster. Today, we no longer fear comets, but eagerly anticipate those that come close enough to us to put on a good sky show.

Appearance of Comets

A comet is a relatively small chunk of icy material (typically a few kilometers across) that develops an atmosphere as it approaches the Sun. Later, there may be a very faint, nebulous tail, extending several million kilometers away from the main body of the comet. Comets have been observed from the earliest times: accounts of comets are found in the histories of virtually all ancient civilizations. The typical comet, however, is not spectacular in our skies, instead having the appearance of a rather faint, diffuse spot of light somewhat smaller than the Moon and many times less brilliant. (Comets seemed more spectacular to people before the invention of artificial lighting, which compromises our view of the night sky.)

Like the Moon and planets, comets appear to wander among the stars, slowly shifting their positions in the sky from night to night. Unlike the planets, however, most comets appear at unpredictable times, which perhaps explain why they frequently inspired fear and superstition in earlier times. Comets typically remain visible for periods that vary from a couple of weeks to several months. We'll say more about what they are made of and how they become visible after we discuss their motions.

Note that still images of comets give the impression that they are moving rapidly across the sky, like a bright meteor or shooting star. Looking only at such images, it is easy to confuse comets and meteors. But seen in the real sky, they are very different: the meteor burns up in our atmosphere and is gone in a few seconds, whereas the comet may be visible for weeks in nearly the same part of the sky.

Comet Orbits

The study of comets as members of the solar system dates from the time of Isaac Newton, who first suggested that they orbited the Sun on extremely elongated ellipses. Newton's colleague Edmund Halley (see the Edmund Halley: Astronomy's Renaissance Man feature box) developed these ideas, and in 1705, he published calculations of 24 comet orbits. In particular, he noted that the orbits of the bright comets that had appeared in the years 1531, 1607, and 1682 were so similar that the three could well be the same comet, returning to perihelion (closest approach to the Sun) at average intervals of 76 years. If so, he predicted that the object should next return about 1758. Although Halley had died by the time the comet appeared as he predicted, it was given the name Comet Halley (rhymes with "valley") in honor of the astronomer who first recognized it as a permanent member of our solar system, orbiting around the Sun. Its aphelion (furthest point from the Sun) is beyond the orbit of Neptune.

We now know from historical records that Comet Halley has actually been observed and recorded on every passage near the Sun since 239 BCE at intervals ranging from 74 to 79 years. The period of its return varies somewhat because of orbital changes produced by the pull of the giant planets. In 1910, Earth was brushed by the comet's tail, causing much needless public concern. Comet Halley last appeared in our skies in 1986 (Figure 13.16), when it was met by several spacecraft that gave us a wealth of information about its makeup; it will return in 2061.



Figure 13.16 : Comet Halley. This composite of three images (one in red, one in green, one in blue) shows Comet Halley as seen with a large telescope in Chile in 1986. During the time the three images were taken in sequence, the comet moved among the stars. The telescope was moved to keep the image of the comet steady, causing the stars to appear in triplicate (once in each color) in the background. (credit: modification of work by ESO)

Voyagers in Astronomy: Edmund Halley: Astronomy's Renaissance Man

Edmund Halley (Figure 13.17), a brilliant astronomer who made contributions in many fields of science and statistics, was by all accounts a generous, warm, and outgoing person. In this, he was quite the opposite of his good friend Isaac Newton, whose great work, the *Principia* (see Orbits and Gravity), Halley encouraged, edited, and helped pay to publish. Halley himself published his first scientific paper at age 20, while still in college. As a result, he was given a royal commission to go to Saint Helena (a remote island off the coast of Africa where Napoleon would later be exiled) to make the first telescopic survey of the southern sky. After returning, he received the equivalent of a master's degree and was elected to the prestigious Royal Society in England, all at the age of 22.

In addition to his work on comets, Halley was the first astronomer to recognize that the so-called "fixed" stars move relative to each other, by noting that several bright stars had changed their positions since Ptolemy's publication of the ancient Greek catalogs. He wrote a paper on the possibility of an infinite universe, proposed that some stars may be variable, and discussed the nature and size of *nebulae* (glowing cloudlike structures visible in telescopes). While in Saint Helena, Halley observed the planet Mercury going across the face of the Sun and developed the mathematics of how such transits could be used to establish the size of the solar system.

In other fields, Halley published the first table of human life expectancies (the precursor of life-insurance statistics); wrote papers on monsoons, trade winds, and tides (charting the tides in the English Channel for the first time); laid the foundations for the systematic study of Earth's magnetic field; studied evaporation and how inland waters become salty; and even designed an underwater diving bell. He served as a British diplomat, advising the emperor of Austria and squiring the future czar of Russia around England (avidly discussing, we are told, both the importance of science and the quality of local brandy).

In 1703, Halley became a professor of geometry at Oxford, and in 1720, he was appointed Astronomer Royal of England. He continued observing Earth and the sky and publishing his ideas for another 20 years, until death claimed him at age 85.

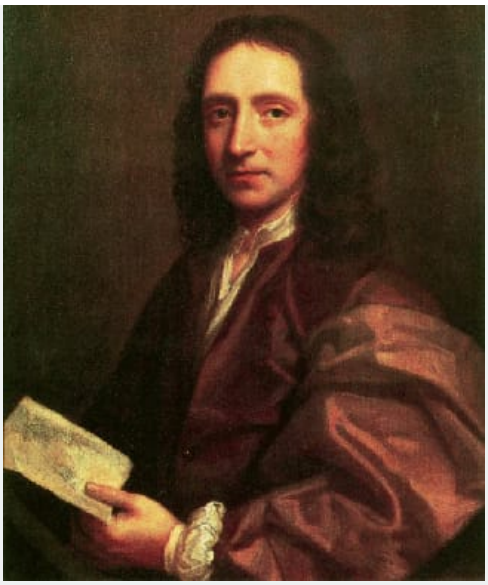


Figure 13.17 : Edmund Halley (1656–1742). Halley was a prolific contributor to the sciences. His study of comets at the turn of the eighteenth century helped predict the orbit of the comet that now bears his name.

Only a few comets return in a time measureable in human terms (shorter than a century or two), like Comet Halley does; these are called *short-period* comets. Many short-period comets have had their orbits changed by coming too close to one of the giant planets—most often Jupiter (and they are thus sometimes called Jupiter-family comets). Most comets have long periods and will take thousands of years to return, if they return at all. As we will see later in this chapter, most Jupiter-family comets come from a different source than the *long-period* comets (those with orbital periods longer than about a century).

Observational records exist for thousands of comets. We were visited by two bright comets in recent decades. First, in March 1996, came Comet Hyakutake, with a very long tail. A year later, Comet Hale-Bopp appeared; it was as bright as the brightest stars and remained visible for several weeks, even in urban areas (see the image that opens this chapter, Figure 13.1) Since then, there have been few comets visible to the naked eye, and astronomers (both professional and amateur) eagerly await such a visitor.

Table 13.2 lists some well-known comets whose history or appearance is of special interest.

Table 13.2: Some Interesting Comets

Name	Period	Significance
Great Comet of 1577	Long	Tycho Brahe showed it was beyond the Moon (a big step in our understanding)
Great Comet of 1843	Long	Brightest recorded comet; visible in daytime
Daylight Comet of 1910	Long	Brightest comet of the twentieth century
West	Long	Nucleus broke into pieces (1976)
Hyakutake	Long	Passed within 15 million km of Earth (1996)
Hale–Bopp	Long	Brightest recent comet (1997)
Swift-Tuttle	133 years	Parent comet of Perseid meteor shower
Halley	76 years	First comet found to be periodic; explored by spacecraft in 1986
Borrelly	6.8 years	Flyby by Deep Space 1 spacecraft (2000)
Biela	6.7 years	Broke up in 1846 and not seen again
Churyumov-Gerasimenko	6.5 years	Target of Rosetta mission (2014–16)
Wild 2	6.4 years	Target of Stardust sample return mission (2004)
Tempel 1	5.7 years	Target of Deep Impact mission (2005)

Name	Period	Significance
Encke	3.3 years	Shortest known period

The Comet's Nucleus

When we look at an active comet, all we normally see is its temporary atmosphere of gas and dust illuminated by sunlight. This atmosphere is called the comet's head or *coma*. Since the gravity of such small bodies is very weak, the atmosphere is rapidly escaping all the time; it must be replenished by new material, which has to come from somewhere. The source is the small, solid nucleus inside, just a few kilometers across, usually hidden by the glow from the much-larger atmosphere surrounding it. The nucleus is the real comet, the fragment of ancient icy material responsible for the atmosphere and the tail (Figure 13.18).

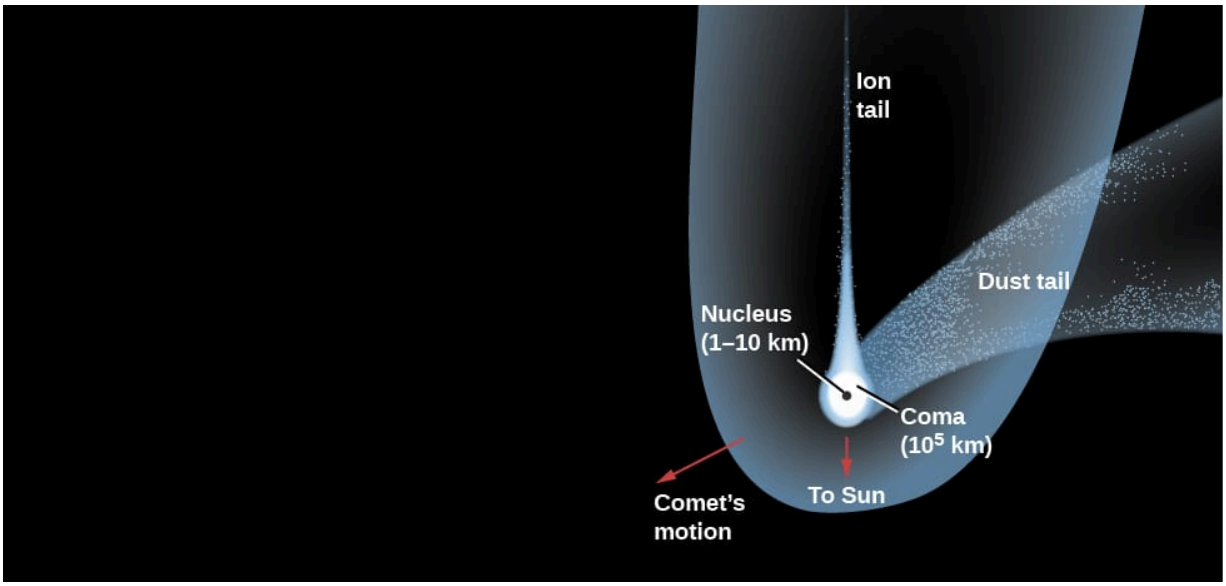


Figure 13.18 : Parts of a Comet. This schematic illustration shows the main parts of a comet. Note that the different structures are not to scale.

The modern theory of the physical and chemical nature of comets was first proposed by Harvard astronomer Fred Whipple in 1950. Before Whipple's work, many astronomers thought that a comet's nucleus might be a loose aggregation of solids, sort of an orbiting "gravel bank," Whipple proposed instead that the nucleus is a solid object a few kilometers across, composed in substantial part of water ice (but with other ices as well) mixed with silicate grains and dust. This proposal became known as the "dirty snowball" model.

The water vapor and other volatiles that escape from the nucleus when it is heated can be detected in the comet's head and tail, and therefore, we can use spectra to analyze what atoms and molecules the nucleus ice consists of. However, we are somewhat less certain of the non-icy component. We have never identified a fragment of solid matter from a comet that has survived passage through Earth's atmosphere. However, spacecraft that have approached comets have carried dust detectors, and some comet dust has even been returned to Earth (see Figure 13.19). It seems that much of the "dirt" in the dirty snowball is dark, primitive hydrocarbons and silicates, rather like the material thought to be present on the dark, primitive asteroids.

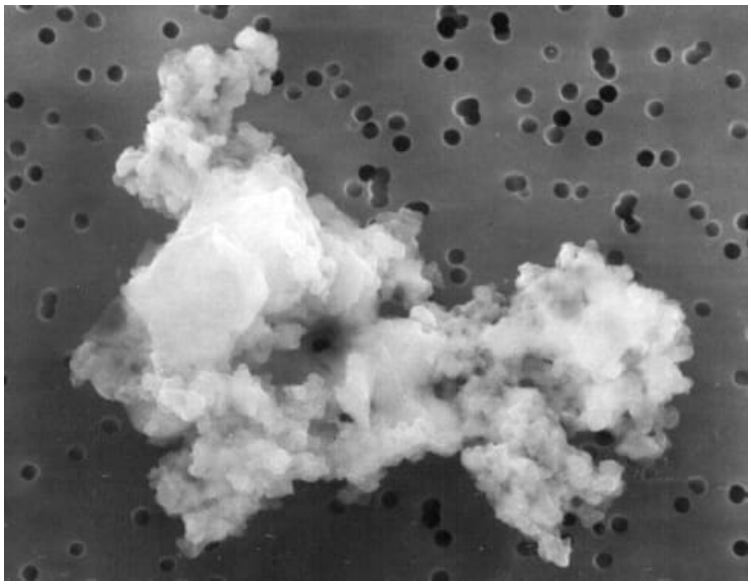


Figure 13.19 : Captured Comet Dust. This particle (seen through a microscope) is believed to be a tiny fragment of cometary dust, collected in the upper atmosphere of Earth. It measures about 10 microns, or 1/100 of a millimeter, across. (credit: NASA/JPL)

Since the nuclei of comets are small and dark, they are difficult to study from Earth. Spacecraft did obtain direct measurements of a comet nucleus, however, in 1986, when three spacecraft swept past Comet Halley at close range (see Figure 13.20). Subsequently, other spacecraft have flown close to other comets. In 2005, the NASA *Deep Impact* spacecraft even carried a probe for a high-speed impact with the nucleus of Comet Tempel 1. But by far, the most productive study of a comet has been by the 2015 Rosetta mission, which we will discuss shortly.



Figure 13.20 : Close-up of Comet Halley. This historic photograph of the black, irregularly shaped nucleus of Comet Halley was obtained by the ESA *Giotto* spacecraft from a distance of about 1000 kilometers. The bright areas are jets of material escaping from the surface. The length of the nucleus is 10 kilometers, and details as small as 1 kilometer can be made out. (credit: modification of work by ESA)

The Comet's Atmosphere

The spectacular activity that allows us to see comets is caused by the evaporation of cometary ices heated by sunlight. Beyond the asteroid belt, where comets spend most of their time, these ices are solidly frozen. But as a comet approaches the Sun, it begins to warm up. If water (H_2O) is the dominant ice, significant quantities vaporize as sunlight heats the surface above 200 K. This happens for the typical comet somewhat beyond the orbit of Mars. The evaporating H_2O in turn releases the dust that was mixed with the ice. Since the comet's nucleus is so small, its gravity cannot hold back either the gas or the dust, both of which flow away into space at speeds of about 1 kilometer per second.

The comet continues to absorb energy as it approaches the Sun. A great deal of this energy goes into the evaporation of its ice, as well as into heating the surface. However, recent observations of many comets indicate that the evaporation is not uniform and that most of the gas is released in sudden spurts, perhaps confined to a few areas of the surface. Expanding into space at a speed of about 1 kilometer per second, the comet's atmosphere can reach an enormous size. The diameter of a comet's head is often as large as Jupiter, and it can sometimes approach a diameter of a million kilometers (Figure 13.21).



Figure 13.21 : Head of Comet Halley. Here we see the cloud of gas and dust that make up the head, or coma, of Comet Halley in 1986. On this scale, the nucleus (hidden inside the cloud) would be a dot too small to see. (credit: modification of work by NASA/W. Liller)

Most comets also develop tails as they approach the Sun. A comet's tail is an extension of its atmosphere, consisting of the same gas and dust that make up its head. As early as the sixteenth century, observers realized that comet tails always point away from the Sun (Figure 13.22), not back along the comet's orbit. Newton proposed that comet tails are formed by a repulsive force of sunlight driving particles away from the head—an idea close to our modern view.

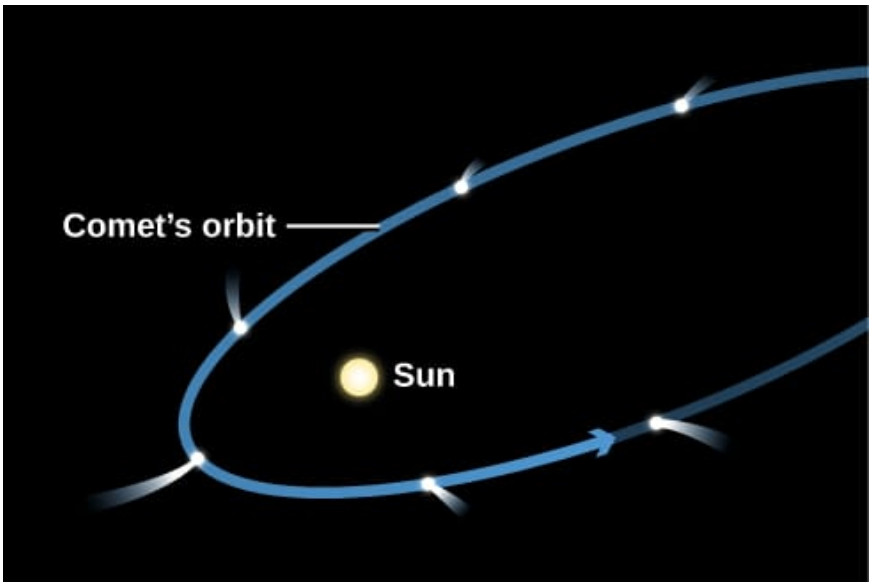


Figure 13.22 : Comet Orbit and Tail. The orientation of a typical comet tail changes as the comet passes perihelion. Approaching the Sun, the tail is behind the incoming comet head, but on the way out, the tail precedes the head.

The two different components that make up the tail (the dust and gas) act somewhat differently. The brightest part of the tail is called the *dust tail*, to differentiate it from a fainter, straight tail made of ionized gas, called the ion tail. The ion tail is carried outward by streams of ions (charged particles) emitted by the Sun. As you can see in Figure 13.23, the smoother dust tail curves a bit, as individual dust particles spread out along the comet's orbit, whereas the straight ion tail is pushed more directly outward

from the Sun by our star's wind of charged particles

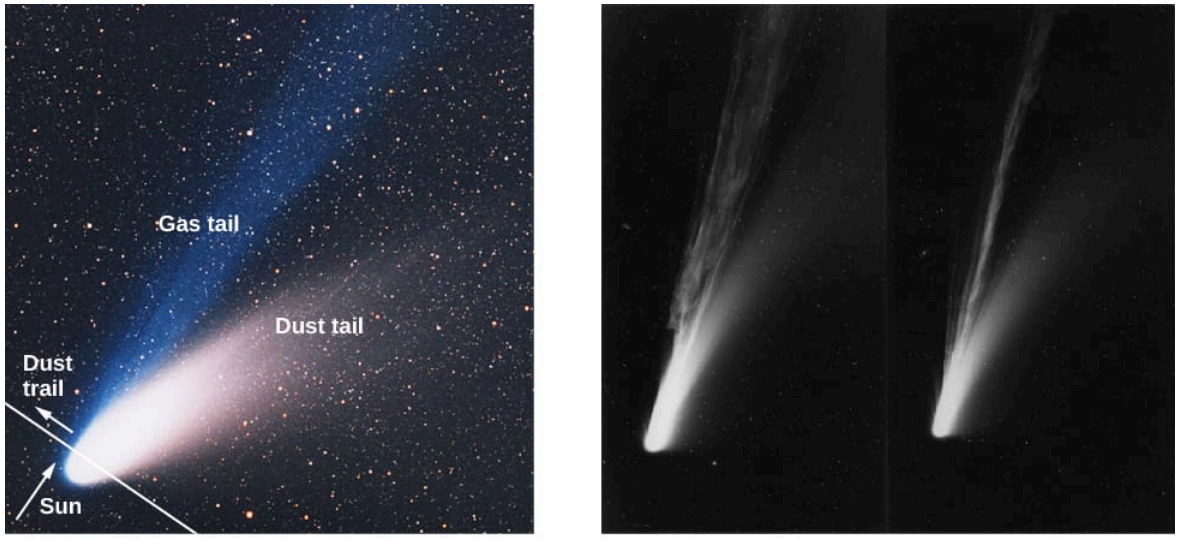


Figure 13.23 : Comet Tails. (a) As a comet nears the Sun, its features become more visible. In this illustration from NASA showing Comet Hale-Bopp, you can see a comet's two tails: the more easily visible dust tail, which can be up to 10 million kilometers long, and the fainter gas tail (or ion tail), which is up to hundreds of millions of kilometers long. The grains that make up the dust tail are the size of smoke particles. (b) Comet Mrkos was photographed in 1957 with a wide-field telescope at Palomar Observatory and also shows a clear distinction between the straight gas tail and the curving dust tail. (credit a: modification of work by ESO/E. Slawik; credit b: modification of work by Charles Kearns, George O. Abell, and Byron Hill)

Link to Learning

These days, comets close to the Sun can be found with spacecraft designed to observe our star. For example, in early July, 2011, astronomers at the ESA/NASA's Solar and Heliospheric Observatory (SOHO) witnessed a comet streaking toward the Sun, one of almost 3000 such sightings. You can also watch a brief video by NASA entitled "Why Are We Seeing So Many Sungrazing Comets?"

The Rosetta Comet Mission

The 1976 Giotto flyby of Comet Halley, which gave us the first picture of an active comet nucleus, was a watershed in our understanding of comets. In the 1990s, European scientists decided to design a much more ambitious mission that would match orbits with an incoming comet and follow it as it approached the Sun. They also proposed that a smaller spacecraft would actually try to land on the comet. The 2-ton main spacecraft was named *Rosetta*, carrying a dozen scientific instruments, and its 100-kilogram lander with nine more instruments was named *Philae*.

The Rosetta mission was launched in 2004. Delays with the launch rocket caused it to miss its original target comet, so an alternate destination was picked, Comet Churyumov-Gerasimenko (named after the two discoverers, but generally denoted 67P). This comet's period of revolution is 6.45 years, making it a Jupiter-family comet.

Since the European Space Agency did not have access to the plutonium-fueled nuclear power sources used by NASA for deep space missions, *Rosetta* had to be solar powered, requiring especially large solar panels. Even these were not enough to keep the craft operating as it matched orbits with 67P near the comet's aphelion. The only solution was to turn off all the spacecraft systems and let it coast for several years toward the Sun, out of contact with controllers on Earth until solar energy was stronger. The success of the mission depended on an automatic timer to turn the power back on as it neared the Sun. Fortunately, this strategy worked.

In August 2014, *Rosetta* began a gradual approach to the comet nucleus, which is a strangely misshapen object about 5 kilometers across, quite different from the smooth appearance of Halley's nucleus (but equally dark). Its rotation period is 12 hours. On November 12, 2014, the *Philae* lander was dropped, descending slowly for 7 hours before gently hitting the surface.

It bounced and rolled, coming to rest under an overhang where there was not enough sunlight to keep its batteries charged. After operating for a few hours and sending data back to the orbiter, *Philae* went silent. The main *Rosetta* spacecraft continued operations, however, as the level of comet activity increased, with steamers of gas jetting from the surface. As the comet approached perihelion in September 2015, the spacecraft backed off to ensure its safety.

The extent of the *Rosetta* images (and data from other instruments) far exceeds anything astronomers had seen before from a comet. The best imaging resolution was nearly a factor of 100 greater than in the best Halley images. At this scale, the comet appears surprisingly rough, with sharp angles, deep pits, and overhangs (Figure 13.24).

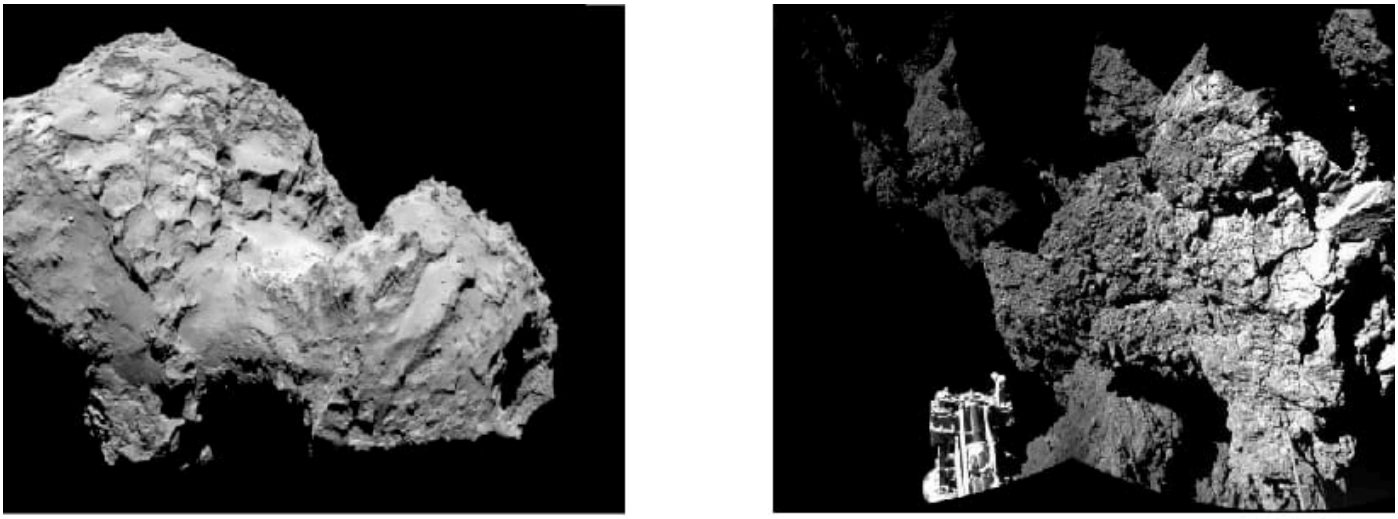


Figure 13.24 : Comet 67P's Strange Shape and Surface Features. (a) This image from the *Rosetta* camera was taken from a distance of 285 kilometers. The resolution is 5 meters. You can see that the comet consists of two sections with a connecting "neck" between them. (b) This close-up view of Comet Churyumov-Gerasimenko is from the *Philae* lander. One of the lander's three feet is visible in the foreground. The lander itself is mostly in shadow. (credit a: modification of work by ESA/Rosetta/MPS for OSIRIS Team MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP/IDA; credit b: modification of work by ESA/Rosetta/Philae/CIVA)

The double-lobed shape of 67P's nucleus has been tentatively attributed to the collision and merger of two independent comet nuclei long ago. The spacecraft verified that the comet's dark surface was covered with organic carbon-rich compounds, mixed with sulfides and iron-nickel grains. 67P has an average density of only 0.5 g/cm^3 (recall water in these units has a density of 1 g/cm^3.) This low density indicates that the comet is quite porous, that is, there is a large amount of empty space among its materials.

We already knew that the evaporation of comet ices was sporadic and limited to small jets, but in comet 67P, this was carried to an extreme. At any one time, more than 99% of the surface is inactive. The active vents are only a few meters across, with the material confined to narrow jets that persist for just a few minutes (Figure 13.25). The level of activity is strongly dependent on solar heating, and between July and August 2015, it increased by a factor of 10. Isotopic analysis of deuterium in the water ejected by the comet shows that it is different from the water found on Earth. Thus, apparently comets like 67P did not contribute to the origin of our oceans or the water in our bodies, as some scientists had thought.



Figure 13.25 : Gas Jets on Comet 67P. "(a and b) Jet activity was photographed by the *Rosetta* spacecraft near perihelion. You can see a jet suddenly appearing; it was active for only a few minutes. (c) This spectacular photo, taken near perihelion, shows the active comet surrounded by multiple jets of gas and dust. (credit a, b: modification of work by ESA/Rosetta/MPS; credit c: modification of work by ESA/Rosetta/NAVCAM)

Link to Learning

The European Space Agency is continuing to make interesting short videos illustrating the challenges and results of the Rosetta and Philae missions. For example, watch "*Rosetta*'s Moment in the Sun" to see some of the images of the comet generating plumes of gas and dust and hear about some of the dangers an active comet poses for the spacecraft.

This page titled 10.4: The "Long-Haired" Comets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 13.4: The "Long-Haired" Comets by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10.5: The Origin and Fate of Comets and Related Objects

Learning Objectives

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

- Describe the traits of the centaur objects
- Chronicle the discovery and describe the composition of the Oort cloud
- Describe trans-Neptunian and Kuiper-belt objects
- Explain the proposed fate of comets that enter the inner solar system

The comets we notice when they come near Earth (especially the ones coming for the first time) are probably the most primitive objects we can study, preserved unchanged for billions of years in the deep freeze of the outer solar system. However, astronomers have discovered many other objects that orbit the Sun beyond the planets.

Centaurs and TNOs

In the outer solar system, where most objects contain large amounts of water ice, the distinction between asteroids and comets breaks down. Astronomers initially still used the name “asteroids” for new objects discovered going around the Sun with orbits that carry them far beyond Jupiter. The first of these objects is Chiron, found in 1977 on a path that carries it from just inside the orbit of Saturn at its closest approach to the Sun out to almost the distance of Uranus (Figure 13.26). The diameter of Chiron is estimated to be about 200 kilometers, much larger than any known comet.

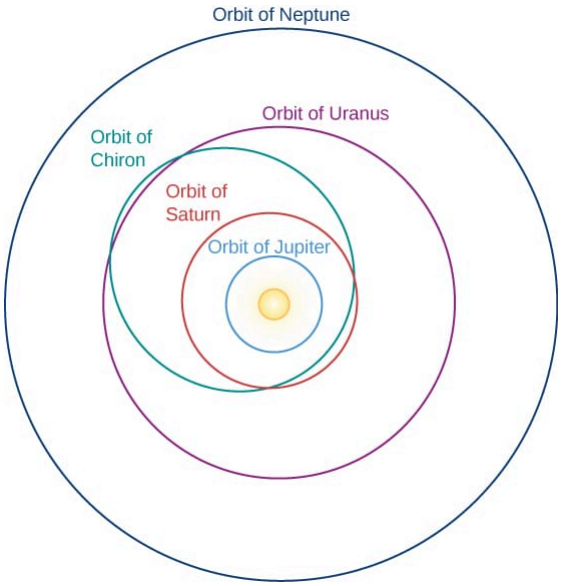


Figure 13.26 : Chiron’s Orbit. Chiron orbits the Sun every 50 years, with its closest approach being inside the orbit of Saturn and its farthest approach out to the orbit of Uranus.

In 1992, a still-more-distant object named Pholus was discovered with an orbit that takes it 33 AU from the Sun, beyond the orbit of Neptune. Pholus has the reddest surface of any object in the solar system, indicating a strange (and still unknown) surface composition. As more objects are discovered in these distant reaches, astronomers decided that they will be given the names of *centaurs* from classical mythology; this is because the centaurs were half human, half horse, and these new objects display some of the properties of both asteroids and comets.

Beyond the orbit of Neptune lies a cold, dark realm populated by objects called simply trans-Neptunian objects (TNOs). The first discovered, and best known, of these TNOs is the dwarf planet Pluto. We discussed Pluto and the New Horizons spacecraft encounter with it in Rings, Moons, and Pluto. The second TNO was discovered in 1992, and now more than a thousand are known,

most of them smaller than Pluto. In late 2018, a team of astronomers announced the discovery of the most distant TNO found so far, orbiting at 120 Astronomical Units, or 3 ½ times the distance of Pluto; they nicknamed it “farout.”

The largest ones after Pluto—named Eris, Makemake, and Haumea—are also classed as dwarf planets. Except for their small size, dwarf planets have many properties in common with the larger planets. Pluto has five moons, and two moons have been discovered orbiting Haumea and one each circling Eris and Makemake.

The Kuiper Belt and the Oort Cloud

TNOs are a part of what is called the Kuiper belt, a large area of space beyond Neptune that is also the source of many comets. Astronomers study the Kuiper belt in two ways. New, more powerful telescopes allow us to discover many of the larger members of the Kuiper belt directly. We can also measure the composition of short-period comets that are thought to originate in the Kuiper belt, where small gravitational perturbations from Neptune can gradually shift their orbits until they can penetrate the inner solar system. More than a thousand Kuiper belt objects have been discovered, and astronomers estimate that there are more than 100,000 with diameters larger than 100 kilometers, in a disk extending out to about 50 AU from the Sun.

Following its successful flyby of Pluto, the New Horizons spacecraft explored the Kuiper Belt. The team eventually identified a member of the belt, 2014 MU16 (subsequently named Arrokoth), that could be reached by a slight change in the spacecraft trajectory. New Horizons flew past Arrokoth on January 1, 2019, at a distance of only 3,500 km, obtaining images and other data. The target turned out to be a contact binary, something new and unexpected (see photo).

By following their orbits backward, we can calculate that the *aphelia* (points farthest from the Sun) of newly discovered comets typically have values near 50,000 AU (more than a thousand times farther than Pluto). This clustering of aphelion distances was first noted by Dutch astronomer Jan Oort, who, in 1950, proposed an idea for the origin of those comets that is still accepted today (Figure 13.27).



(a) (b)

Figure 13.27 : Jan Oort (1900–1992). (a) Jan Oort first suggested that there might be a reservoir of frozen chunks, potential comet nuclei, at the edge of the region of the Sun’s gravitational influence. (b) The first color image of the Kuiper Belt Object named Arrokoth, taken at a distance of 137,000 kilometers from the New Horizons spacecraft on January 1, 2019. (credit (a): This image is copyright by the Leiden Observatory; credit (b): NASA/JHU/SWRI)

It is possible to calculate that a star’s gravitational *sphere of influence*—the distance within which it can exert sufficient gravitation to hold onto orbiting objects—is about one third of its distance to the nearest other stars. Stars in the vicinity of the Sun are spaced in such a way that the Sun’s sphere of influence extends a little beyond 50,000 AU, or about 1 light-year. At such great distances, however, objects in orbit about the Sun can be perturbed by the gravity of passing stars. Some of the perturbed objects can then take on orbits that bring them much closer to the Sun (while others might be lost to the solar system forever).

Oort suggested, therefore, that the new comets we were seeing were examples of objects orbiting the Sun near the edge of its sphere of influence, whose orbits had been disturbed by nearby stars, eventually bringing them close to the Sun where we can

see them.[1] The reservoir of ancient icy objects from which such comets are derived is now called the Oort cloud.

Astronomers estimate that there are about a trillion (10^{12}) comets in the Oort cloud. In addition, we estimate that about 10 times this number of icy objects could be orbiting the Sun in the volume of space between the Kuiper belt (which is gravitationally linked to Neptune) and the Oort cloud. These objects remain undiscovered because they are too faint to be seen directly and their orbits are too stable to permit any of them to be deflected inward close to the Sun. The total number of icy or cometary objects in the outer reaches of our solar system could thus be on the order of 10 trillion (10^{13}), a very large number indeed.

What is the mass represented by 10^{13} comets? We can make an estimate if we assume something about comet sizes and masses. Let us suppose that the nucleus of Comet Halley is typical. Its observed volume is about 600 km^3. If the primary constituent is water ice with a density of about 1 g/cm^3, then the total mass of Halley's nucleus must be about 6×10^{14} kilograms. This is about one ten billionth (10^{-10}) of the mass of Earth.

If our estimate is reasonable and there are 10^{13} comets with this mass out there, their total mass would be equal to about 1000 Earths—comparable to the mass of all the planets put together. Therefore, icy, cometary material could be the most important constituent of the solar system after the Sun itself.

Example 13.1: Mass of the Oort Cloud Comets

Suppose the Oort cloud contains 10^{12} comets with an average diameter of 10 km each. Let's estimate the mass of the total Oort cloud.

Solution

We can start by assuming that typical comets are about the size of Comets Halley and Borrelly, with a diameter of 10 km and a density appropriate to water ice, which is about 1 g/cm^3 or 1000 kg/m^3. We know that $\text{density} = \frac{\text{mass}}{\text{volume}}$, the volume of a sphere, $V = \frac{4}{3}\pi R^3$, and the radius, $R = \frac{1}{2}D$. Therefore, for each comet,

$$\begin{aligned} \text{mass} &= \text{density} \times \text{volume} \\ &= \text{density} \times \tfrac{4}{3}\pi\left(\tfrac{1}{2}D\right)^3 \end{aligned}$$

Given that $10\text{km} = 10^4\text{m}$, each comet's mass is

$$\begin{aligned} \text{mass} &= 1000\,\text{kg/m}^3 \times \tfrac{4}{3} \times 3.14 \times \tfrac{1}{8} \times \left(10^4\right)^3\,\text{m}^3 \\ &\approx 10^{15}\,\text{kg} \\ &= 10^{12}\,\text{tons} \end{aligned}$$

To calculate the total mass of the cloud, we multiply this typical mass for one comet by the number of comets:

$$\begin{aligned} \text{total mass} &= 10^{15}\,\text{kg/comet} \times 10^{12}\,\text{comets} \\ &= 10^{27}\,\text{kg} \end{aligned}$$

Exercise 10.5.1

How does the total mass we calculated above compare to the mass of Jupiter? To the mass of the Sun? (Give a numerical answer.)

Answer available at the end of this section.

Early Evolution of the Planetary System

Comets from the Oort cloud help us sample material that formed very far from the Sun, whereas the short-period comets from the Kuiper belt sample materials that were planetesimals in the solar nebula disk but did not form planets. Studies of the Kuiper

belt also are influencing our understanding of the early evolution of our planetary system.

The objects in the Oort cloud and the Kuiper belt have different histories, and they may therefore have different compositions. Astronomers are therefore very interested in comparing detailed measurements of the comets derived from these two source regions. Most of the bright comets that have been studied in the past (such as Hyakutake and Hale-Bopp) are Oort cloud comets, but P67 and several other comets targeted for spacecraft measurements in the next decade are Jupiter-family comets from the Kuiper belt (see Table 13.2).

The Kuiper belt is made up of ice-and-rock planetesimals, a remnant of the building blocks of the planets. Since it is gravitationally linked to Neptune, it can help us understand the formation and history of the solar system. As the giant planets formed, their gravity profoundly influenced the orbits of Kuiper belt objects. Computer simulations of the early evolution of the planetary system suggest that the gravitational interactions between the giant planets and the remaining planetesimals caused the orbit of Jupiter to drift inward, whereas the orbits of Saturn, Uranus, and Neptune all expanded, carrying the Kuiper belt with them.

Another hypotheses involves a fifth giant planet that was expelled from the solar system entirely as the planetary orbits shifted. Neptune's retrograde (backward-orbiting) moon Triton (which is nearly as large as Pluto) may have been a Kuiper belt object captured by Neptune during the period of shifting orbits. It clearly seems that the Kuiper belt may carry important clues to the way our solar system reached its present planetary configuration.

Making Connections: Comet Hunting as a Hobby

When amateur astronomer David Levy (Figure 13.28), the co-discoverer of Comet Shoemaker-Levy 9, found his first comet, he had already spent 928 fruitless hours searching through the dark night sky. But the discovery of the first comet only whetted his appetite. Since then, he has found 8 others on his own and 13 more working with others. Despite this impressive record, he ranks only third in the record books for number of comet discoveries. But David hopes to break the record someday.

All around the world, dedicated amateur observers spend countless nights scanning the sky for new comets. Astronomy is one of the very few fields of science where amateurs can still make a meaningful contribution, and the discovery of a comet is one of the most exciting ways they can establish their place in astronomical history. Don Machholz, a California amateur (and comet hunter) who has been making a study of comet discoveries, reported that between 1975 and 1995, 38% of all comets discovered were found by amateurs. Those 20 years yielded 67 comets for amateurs, or almost 4 per year. That might sound pretty encouraging to new comet hunters, until they learn that the average number of hours the typical amateur spent searching for a comet before finding one was about 420. Clearly, this is not an activity for impatient personalities.

What do comet hunters do if they think they have found a new comet? First, they must check the object's location in an atlas of the sky to make sure it really is a comet. Since the first sighting of a comet usually occurs when it is still far from the Sun and before it sports a significant tail, it will look like only a small, fuzzy patch. And through most amateur telescopes, so will nebulae (clouds of cosmic gas and dust) and galaxies (distant groupings of stars). Next, they must check that they have not come across a comet that is already known, in which case, they will only get a pat on the back instead of fame and glory. Then they must re-observe or re-image it sometime later to see whether its motion in the sky is appropriate for comets.

Often, comet hunters who think they have made a discovery get another comet hunter elsewhere in the country to confirm it. If everything checks out, the place they contact is the Central Bureau for Astronomical Telegrams at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts (http://www.cbat.eps.harvard.edu/). If the discovery is confirmed, the bureau will send the news out to astronomers and observatories around the world. One of the unique rewards of comet hunting is that the discoverer's name becomes associated with the new comet—a bit of cosmic fame that few hobbies can match.



Figure 13.28 : David Levy. Amateur astronomer David Levy ranks third in the world for comet discoveries. (credit: Andrew Fraknoi)

The Fate of Comets

Any comet we see today will have spent nearly its entire existence in the Oort cloud or the Kuiper belt at a temperature near absolute zero. But once a comet enters the inner solar system, its previously uneventful life history begins to accelerate. It may, of course, survive its initial passage near the Sun and return to the cold reaches of space where it spent the previous 4.5 billion years. At the other extreme, it may collide with the Sun or come so close that it is destroyed on its first perihelion passage (several such collisions have been observed with space telescopes that monitor the Sun). Sometimes, however, the new comet does not come that close to the Sun but instead interacts with one or more of the planets.

Link to Learning

SOHO (the Solar and Heliospheric Observatory) has an excellent collection of videos of comets that come near the Sun. At this site, comet ISON approaches the Sun and is believed to be destroyed in its passage.

A comet that comes within the gravitational influence of a planet has three possible fates. It can (1) impact the planet, ending the story at once; (2) speed up and be ejected, leaving the solar system forever; or (3) be perturbed into an orbit with a shorter period. In the last case, its fate is sealed. Each time it approaches the Sun, it loses part of its material and also has a significant chance of collision with a planet. Once the comet is in this kind of short-period orbit, its lifetime starts being measured in thousands, not billions, of years.

A few comets end their lives catastrophically by breaking apart (sometimes for no apparent reason) (Figure 13.29). Especially spectacular was the fate of the faint Comet Shoemaker-Levy 9, which broke into about 20 pieces when it passed close to Jupiter in July 1992. The fragments of Shoemaker-Levy were actually captured into a very elongated, two-year orbit around Jupiter, more than doubling the number of known jovian moons. This was only a temporary enrichment of Jupiter's family, however, because in July 1994, all the comet fragments crashed unto Jupiter, releasing energy equivalent to millions of megatons of TNT.

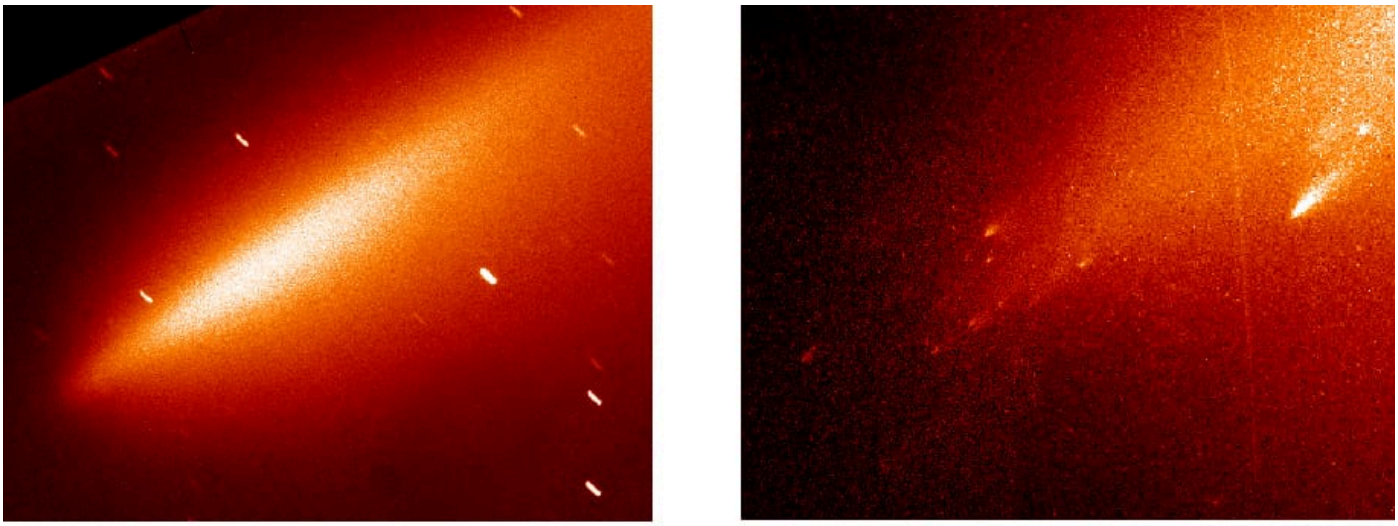


Figure 13.29 : Breakup of Comet LINEAR. (a) A ground-based view with much less detail and (b) a much more detailed photo with the Hubble Space Telescope, showing the multiple fragments of the nucleus of Comet LINEAR. The comet disintegrated in July 2000 for no apparent reason. (Note in the left view, the fragments all blend their light together, and can't be distinguished. The short diagonal white lines are stars that move in the image, which is keeping track of the moving comet.) (credit a: modification of work by the University of Hawaii; credit b: modification of work by NASA, Harold Weaver (the Johns Hopkins University), and the HST Comet LINEAR Investigation Team)

As each cometary fragment streaked into the jovian atmosphere at a speed of 60 kilometers per second, it disintegrated and exploded, producing a hot fireball that carried the comet dust as well as atmospheric gases to high altitudes. These fireballs were clearly visible in profile, with the actual point of impact just beyond the jovian horizon as viewed from Earth (Figure 13.30). As each explosive plume fell back into Jupiter, a region of the upper atmosphere larger than Earth was heated to incandescence and glowed brilliantly for about 15 minutes, a glow we could detect with infrared-sensitive telescopes.

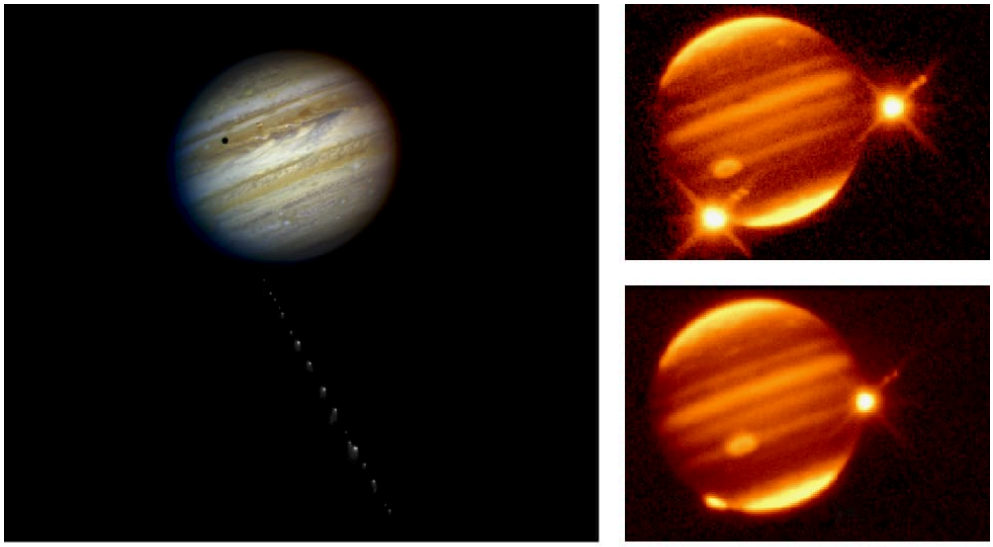


Figure 13.30 : Comet Impact on Jupiter. (a) The "string" of white objects are fragments of Comet Shoemaker-Levy 9 approaching Jupiter. (b) The first fragment of the comet impacts Jupiter, with the point of contact on the bottom left side in this image. On the right is Jupiter's moon, Io. The equally bright spot in the top image is the comet fragment flaring to maximum brightness. The bottom image, taken about 20 minutes later, shows the lingering flare from the impact. The Great Red Spot is visible near the center of Jupiter. These infrared images were taken with a German-Spanish telescope on Calar Alto in southern Spain. (credit a: modification of work by ESA; credit b: modification of work by Tom Herbst, Max-Planck-Institut fuer Astronomie, Heidelberg, Doug Hamilton, Max-Planck-Institut fuer Kernphysik, Heidelberg, Hermann Boehnhardt, Universitaets-Sternewarte, Muenchen, and Jose Luis Ortiz Moreno, Instituto de Astrofisica de Andalucia, Granada)

After this event, dark clouds of debris settled into the stratosphere of Jupiter, producing long-lived "bruises" (each still larger than Earth) that could be easily seen through even small telescopes (Figure 13.31). Millions of people all over the world peered

at Jupiter through telescopes or followed the event via television or online. Another, smaller, impact feature was seen on Jupiter in summer 2009 (and six more since), indicating that the 1994 events were by no means unique. Seeing these large, impact explosions on Jupiter helps us to appreciate the disaster that would happen to our planet if we were hit by a comet or asteroid.

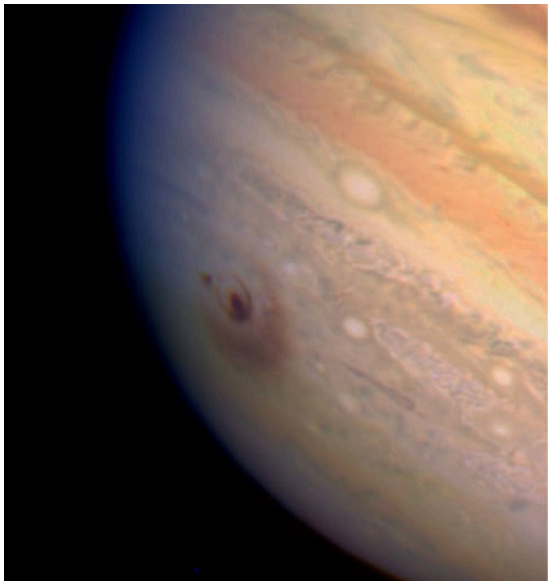


Figure 13.31 : Impact Dust Cloud on Jupiter. These features result from the impact of Comet Shoemaker-Levy 9 with Jupiter, seen with the Hubble Space Telescope 105 minutes after the impact that produced the dark rings (the compact back dot came from another fragment). The inner edge of the diffuse, outer ring is about the same size as Earth. Later, the winds on Jupiter blended these features into a broad spot that remained visible for more than a month. (credit: modification of work by H. Hammel, MIT, and NASA/ESA)

For comets that do not meet so dramatic an end, measurements of the amount of gas and dust in their atmospheres permit us to estimate the total losses during one orbit. Typical loss rates are up to a million tons per day from an active comet near the Sun, adding up to some tens of millions of tons per orbit. At that rate, a typical comet will be gone after a few thousand orbits. This will probably be the fate of Comet Halley in the long run.

Earlier we mentioned that comets can also escape the solar system. In 2019, astronomers discovered comet 2I/Borisov, which came from a planetary system other than ours. It is only the second interstellar object we have found, and the first that is rich in volatile materials such as water and carbon compounds. The presence of an atmosphere allowed its composition to be studied. Borisov contains much more carbon monoxide than normal solar system comets, which tells us that it was probably formed in a cooler location, perhaps around a red dwarf star.

Link to Learning

Watch this 4-minute NASA retrospective on the collision of Comet Shoemaker-Levy 9 with Jupiter History Channel video to learn more.

Footnotes

- 1We now know that not every comet we see originates in the Kuiper Belt or Oort Cloud. In 2017 and 2019, astronomers discovered two “interstellar comets,” (named 1I/Oumuamua and 2I/Borisov) whose orbits indicated that they came from outside the solar system! (Oumoumua means scout or messenger in Hawaiian.)

Answers

Exercise 10.5.1

The mass of Jupiter is about $1.9 \times 10^{27}\text{kg}$. The mass of the Oort cloud calculated above is 10^{27} kg. So the cloud would contain about half a Jupiter of mass. The mass of the Sun is $2 \times 10^{30}\text{kg}$. This means the Oort cloud would be

$$\frac{10^{27}\,\text{kg}}{(2 \times 10^{30}\,\text{kg})} = 0.0005 \times \text{the mass of the Sun}$$

This page titled 10.5: The Origin and Fate of Comets and Related Objects is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 13.5: The Origin and Fate of Comets and Related Objects by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10.6: Pluto and Charon

Learning Objectives

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

- Compare the orbital characteristics of Pluto with those of the planets
- Describe information about Pluto's surface deduced from the New Horizons images
- Note some distinguishing characteristics of Pluto's large moon Charon

Pluto is not a moon, but we discuss it here because its size and composition are similar to many moons in the outer solar system. Our understanding of Pluto (and its large moon Charon) have changed dramatically as a result of the New Horizons flyby in 2015.

Is Pluto a Planet?

Pluto was discovered through a careful, systematic search, unlike Neptune, whose position was calculated from gravitational theory. Nevertheless, the history of the search for Pluto began with indications that Uranus had slight departures from its predicted orbit, departures that could be due to the gravitation of an undiscovered "Planet X." Early in the twentieth century, several astronomers, most notably Percival Lowell, then at the peak of his fame as an advocate of intelligent life on Mars, became interested in searching for this ninth planet.

Lowell and his contemporaries based their calculations primarily on tiny unexplained irregularities in the motion of Uranus. Lowell's computations indicated two possible locations for a perturbing Planet X; the more likely of the two was in the constellation Gemini. He predicted a mass for the planet intermediate between the masses of Earth and Neptune (his calculations gave about 6 Earth masses). Other astronomers, however, obtained other solutions from the tiny orbital irregularities, even including one model that indicated two planets beyond Neptune.

At his Arizona observatory, Lowell searched without success for the unknown planet from 1906 until his death in 1916, and the search was not renewed until 1929. In February 1930, a young observing assistant named Clyde Tombaugh (see the Clyde Tombaugh: From the Farm to Fame feature box), comparing photographs he made on January 23 and 29 of that year, found a faint object whose motion appeared to be about right for a planet far beyond the orbit of Neptune (Figure 12.17). The new planet was named for Pluto, the Roman god of the underworld, who dwelt in remote darkness, just like the new planet. The choice of this name, among hundreds suggested, was helped by the fact that the first two letters were Percival Lowell's initials.

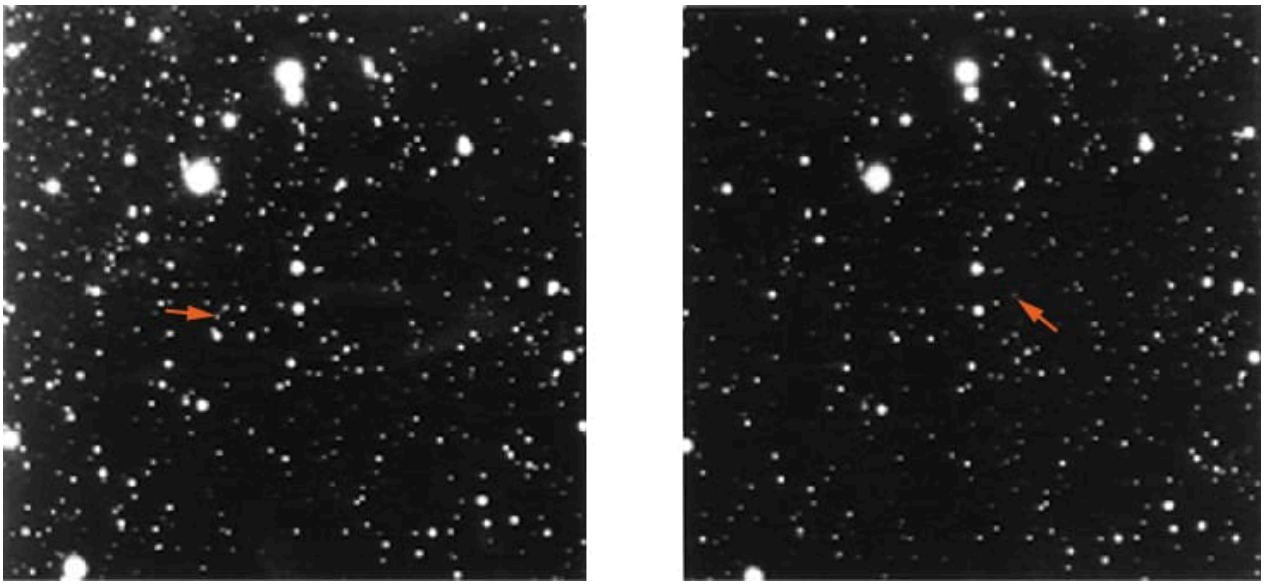


Figure 12.17 : Pluto's Motion. Portions of the two photographs by which Clyde Tombaugh discovered Pluto in 1930. The left one was taken on January 23 and the right on January 29. Note that Pluto, indicated by an arrow, has moved among the stars during those six nights. If we hadn't put an arrow next to it, though, you probably would never have spotted the dot that moved. (credit: modification of work by the Lowell Observatory Archives)

Although the discovery of Pluto appeared initially to be a vindication of gravitational theory similar to the earlier triumph of Adams and Le Verrier in predicting the position of Neptune, we now know that Lowell's calculations were wrong. When its mass and size were finally measured, it was found that Pluto could not possibly have exerted any measurable pull on either Uranus or Neptune. Astronomers are now convinced that the reported small anomalies in the motions of Uranus are not, and never were, real.

From the time of its discovery, it was clear that Pluto was not a giant like the other four outer solar system planets. For a long time, it was thought that the mass of Pluto was similar to that of Earth, so that it was classed as a fifth terrestrial planet, somehow misplaced in the far outer reaches of the solar system. There were other anomalies, however, as Pluto's orbit was more eccentric and inclined to the plane of our solar system than that of any other planet. Only after the discovery of its moon Charon in 1978 could the mass of Pluto be measured, and it turned out to be far less than the mass of Earth.

In addition to Charon, Pluto has four much smaller moons. Subsequent observations of Charon showed that this moon is in a retrograde orbit and has a diameter of about 1200 kilometers, more than half the size of Pluto itself (Figure 12.18). This makes Charon the moon whose size is the largest fraction of its parent planet. We could even think of Pluto and Charon as a double world. Seen from Pluto, Charon would be as large as eight full moons on Earth.

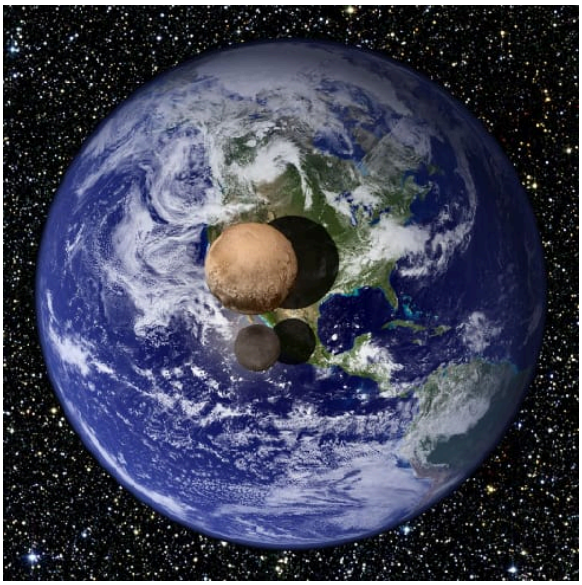


Figure 12.18 : Comparison of the Sizes of Pluto and Its Moon Charon with Earth. This graphic vividly shows how tiny Pluto is relative to a terrestrial planet like Earth. That is the primary justification for putting Pluto in the class of dwarf planets rather than terrestrial planets. (credit: modification of work by NASA)

To many astronomers, Pluto seemed like the odd cousin that everyone hopes will not show up at the next family reunion. Neither its path around the Sun nor its size resembles either the giant planets or the terrestrial planets. In the 1990s, astronomers began to discover additional small objects in the far outer solar system, showing that Pluto was not unique. We will discuss these trans-neptunian objects later with other small bodies, in the chapter on Comets and Asteroids: Debris of the Solar System. One of them (called Eris) is nearly the same size as Pluto, and another (Makemake) is substantially smaller. It became clear to astronomers that Pluto was so different from the other planets that it needed a new classification. Therefore, it was called a *dwarf planet*, meaning a planet much smaller than the terrestrial planets. We now know of many small objects in the vicinity of Pluto and we have classified several as dwarf planets.

A similar history was associated with the discovery of the asteroids. When the first asteroid (Ceres) was discovered at the beginning of the nineteenth century, it was hailed as a new planet. In the following years, however, other objects were found with similar orbits to Ceres. Astronomers decided that these should not all be considered planets, so they invented a new class of objects, called minor planets or asteroids. Today, Ceres is also called a dwarf planet. Both minor planets and dwarf planets are part of a whole belt or zones of similar objects (as we will discuss in Comets and Asteroids: Debris of the Solar System).

So, is Pluto a planet? Our answer is yes, but it is a *dwarf planet*, clearly not in the same league with the eight major planets (four giants and four terrestrials). While some people were upset when Pluto was reclassified, we might point out that a dwarf tree is still a type of tree and (as we shall see) a dwarf galaxy is still a type of galaxy.

Voyagers in Astronomy: Clyde Tombaugh: From the Farm to Fame

Clyde Tombaugh discovered Pluto when he was 24 years old, and his position as staff assistant at the Lowell Observatory was his first paying job. Tombaugh had been born on a farm in Illinois, but when he was 16, his family moved to Kansas. There, with his uncle's encouragement, he observed the sky through a telescope the family had ordered from the Sears catalog. Tombaugh later constructed a larger telescope on his own and devoted his nights (when he wasn't too tired from farm work) to making detailed sketches of the planets (Figure 12.19).

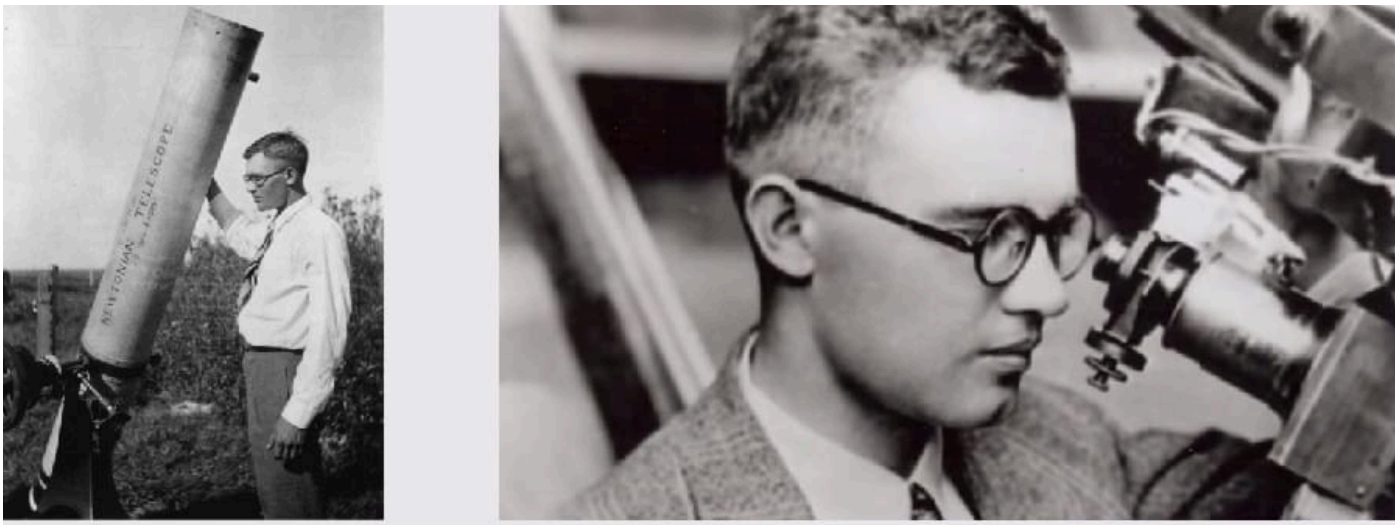


Figure 12.19 : Clyde Tombaugh (1906–1997). (a) Tombaugh is pictured on his family farm in 1928 with a 9-inch telescope he built. (b) Here Tombaugh is looking through an eyepiece at the Lowell Observatory. (credit b: modification of work by NASA)

In 1928, after a hailstorm ruined the crop, Tombaugh decided he needed a job to help support his family. Although he had only a high school education, he thought of becoming a telescope builder. He sent his planet sketches to the Lowell Observatory, seeking advice about whether such a career choice was realistic. By a wonderful twist of fate, his query arrived just when the Lowell astronomers realized that a renewed search for a ninth planet would require a very patient and dedicated observer.

The large photographic plates (pieces of glass with photographic emulsion on them) that Tombaugh was hired to take at night and search during the day contained an average of about 160,000 star images each. How to find Pluto among them? The technique involved taking two photographs about a week apart. During that week, a planet would move a tiny bit, while the stars remained in the same place relative to each other. A new instrument called a “blink comparator” could quickly alternate the two images in an eyepiece. The stars, being in the same position on the two plates, would not appear to change as the two images were “blinked.” But a moving object would appear to wiggle back and forth as the plates were alternated.

After examining more than 2 million stars (and many false alarms), Tombaugh found his planet on February 18, 1930. The astronomers at the observatory checked his results carefully, and the find was announced on March 13, the 149th anniversary of the discovery of Uranus. Congratulations and requests for interviews poured in from around the world. Visitors descended on the observatory in scores, wanting to see the place where the first new planet in almost a century had been discovered, as well as the person who had discovered it.

In 1932, Tombaugh took leave from Lowell, where he had continued to search and blink, to get a college degree. Eventually, he received a master’s degree in astronomy and taught navigation for the Navy during World War II. In 1955, after working to develop a rocket-tracking telescope, he became a professor at New Mexico State University, where he helped found the astronomy department. He died in 1997; some of his ashes were placed inside the New Horizons spacecraft to Pluto.

Link to Learning

Here is a touching video about Tombaugh’s life as described by his children.

The Nature of Pluto

Almost everything we know about Pluto and its moons comes from the New Horizons spacecraft, which flew by in 2015 before traveling on into the outermost parts of the planetary system. Using data from the New Horizons probe, astronomers have measured the diameter of Pluto as 2370 kilometers, only 60 percent as large as our Moon. From the diameter and mass, we find

a density of 1.9 g/cm^3, suggesting that Pluto is a mixture of rocky materials and water ice in about the same proportions as many outer-planet moons.

Parts of Pluto's surface are highly reflective, and its spectrum demonstrates the presence on its surface of frozen methane, carbon monoxide, and nitrogen. The maximum surface temperature ranges from about 50 K when Pluto is farthest from the Sun to 60 K when it is closest. Even this small difference is enough to cause a partial sublimation (going from solid to gas) of the methane and nitrogen ice. This generates an atmosphere when Pluto is close to the Sun, and it freezes out when Pluto is farther away. Observations of distant stars seen through this thin atmosphere indicate that the surface pressure is about a ten-thousandth of Earth's. Because Pluto is a few degrees warmer than Triton, its atmospheric pressure is about ten times greater. This atmosphere contains several distinct haze layers, presumably caused by photochemical reactions, like those in Titan's atmosphere (Figure 12.20).

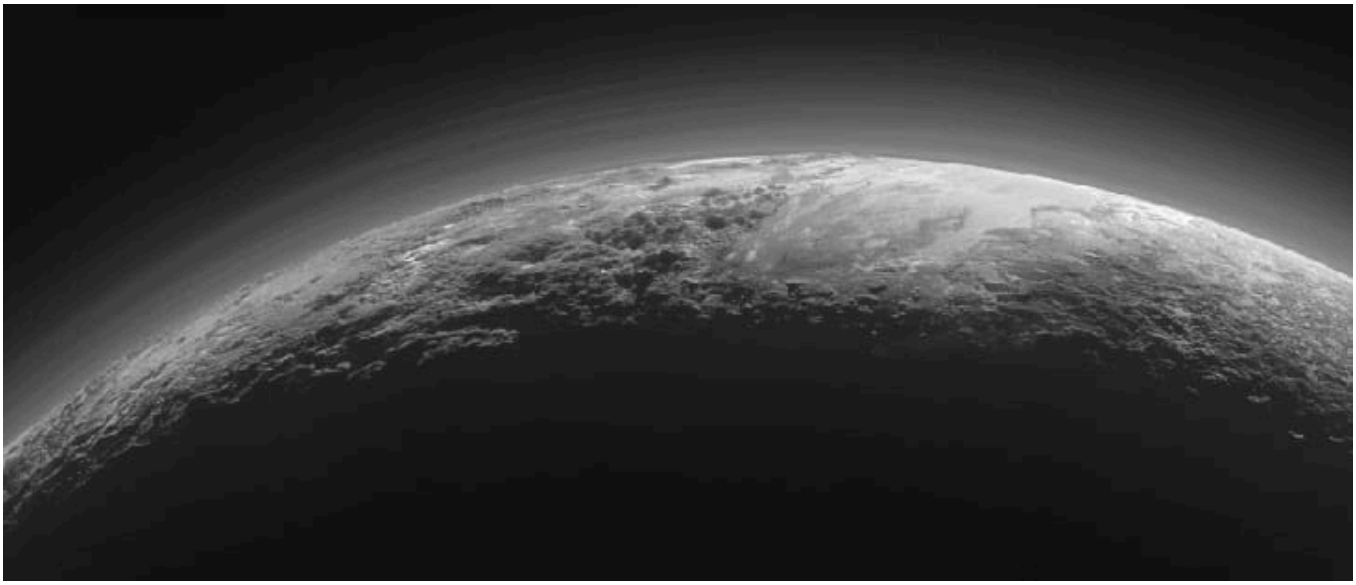


Figure 12.20 : Haze Layers in the Atmosphere of Pluto. This is one of the highest-resolution photos of Pluto, taken by the New Horizons spacecraft 15 minutes after its closest approach. It shows 12 layers of haze. Note also the range of mountains with heights up to 3500 meters. (credit: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

Reaching Pluto with a spacecraft was a major challenge, especially in an era when reduced NASA budgets could not support large, expensive missions like Galileo and Cassini. Yet like Galileo and Cassini, a Pluto mission would require a nuclear electric system that used the heat from plutonium to generate the energy to power the instruments and keep them operating far from the warmth of the Sun. NASA made available one of the last of its nuclear generators for such a mission. Assuming an affordable but highly capable spacecraft could be built, there was still the problem of getting to Pluto, nearly 5 billion kilometers from Earth, without waiting decades. The answer was to use Jupiter's gravity to slingshot the spacecraft toward Pluto.

The 2006 launch of New Horizons started the mission with a high speed, and the Jupiter flyby just a year later gave it the required additional boost. The New Horizons spacecraft arrived at Pluto in July 2015, traveling at a relative speed of 14 kilometers per second (or about 50,000 kilometers per hour). With this high speed, the entire flyby sequence was compressed into just one day. Most of the data recorded near closest approach could not be transmitted to Earth until many months later, but when it finally arrived, astronomers were rewarded with a treasure trove of images and data.

Geology of Pluto

Pluto is not the geologically dead world that many anticipated for such a small object—far from it. The division of the surface into areas with different composition and surface texture is apparent in the global color photo shown in Figure 12.21. The reddish color is enhanced in this image to bring out differences in color more clearly. The darker parts of the surface are cratered, but adjacent to them is a nearly featureless light area in the lower right quadrant of this image. This is a huge ice-filled depression called the Sputnik Plains, named for the first artificial Earth satellite. The absence of impact craters suggests a surface no more than 10 million years old. The dark areas show the colors of photochemical haze or smog similar to that in the atmosphere of Titan. The dark material that is staining these old surfaces could come from Pluto's atmospheric haze or from chemical reactions taking place at the surface due to the action of sunlight.



Figure 12.21 : Global Color Image of Pluto. This New Horizons image clearly shows the variety of terrains on Pluto. The dark area in the lower left is covered with impact craters, while the large light area in the center and lower right is a flat basin devoid of craters. The colors you see are somewhat enhanced to bring out subtle differences. (credit: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

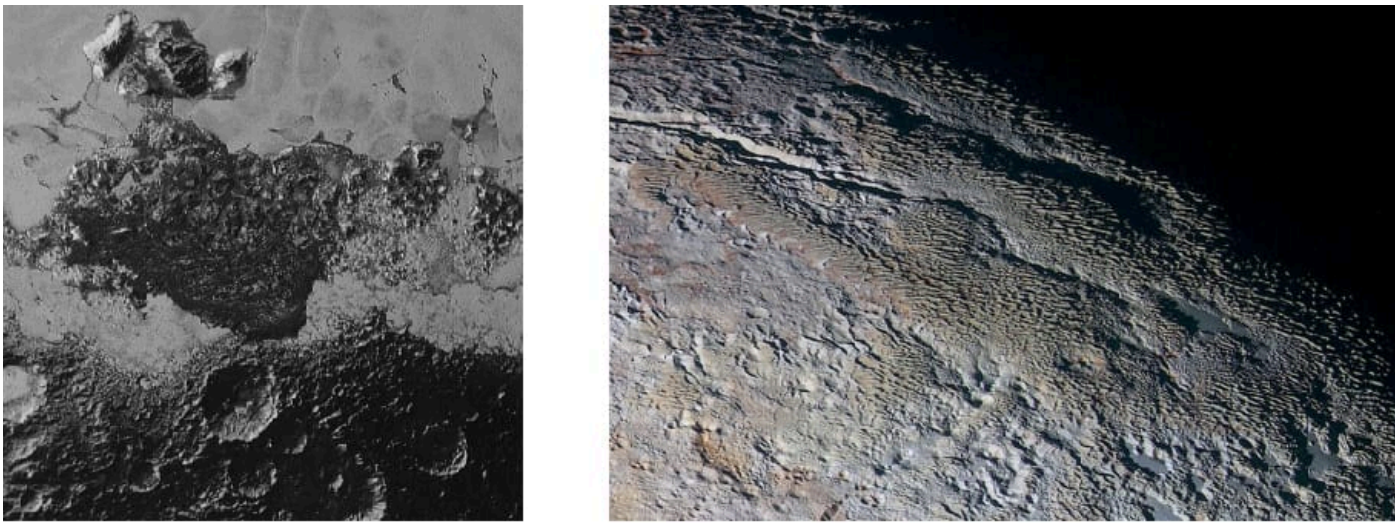
Figure 12.22 shows some of the remarkable variety of surface features New Horizons revealed. we are finding on Pluto. Toward the right on this image, we see the shoreline of the Sputnik Plains, showing evidence of relatively recent geological activity. The nitrogen and carbon monoxide ice that fills the Sputnik depression shows cells or polygons with an average width of more than 30 kilometers, caused by slow convection in the ice.[2]



Figure 12.22 : Diversity of Terrain on Pluto. This enhanced color view of a strip of Pluto's surface about 80 kilometers long shows a variety of different surface features. From left to right, we first cross a region of "badlands" with some craters showing, and then move across a wide range of mountains made of water ice and coated with the redder material we saw in the previous image. Then, at right, we arrive at the "shoreline" of the great sea of frozen nitrogen that the mission scientists have named the "Sputnik Plains." This nitrogen sea is divided into mysterious cells or segments that are many kilometers across. (credit: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

Figure 12.23 shows another view of the boundary between different types of geology. The width of this image is 250 kilometers, and it shows dark, ancient, heavily cratered terrain; dark, uncratered terrain with a hilly surface; smooth, geologically young terrain; and a small cluster of mountains more than 3000 meters high. In the best images, the light areas of nitrogen ice seem to have flowed much like glaciers on Earth, covering some of the older terrain underneath them.

The isolated mountains in the midst of the smooth nitrogen plains are probably also made of water ice, which is very hard at the temperatures on Pluto and can float on frozen nitrogen. Additional mountains, and some hilly terrain that reminded the mission scientists of snakeskin, are visible in part (b) of Figure 12.23.



(a) (b)

Figure 12.23 : Diversity of Terrains on Pluto. (a) In this photo, about 250 kilometers across, we can see many different kinds of terrain. At the bottom are older, cratered highlands; a V-shaped region of hills without cratering points toward the bottom of the image. Surrounding the V-shaped dark region is the smooth, brighter frozen nitrogen plain, acting as glaciers on Earth do. Some isolated mountains, made of frozen water ice, are floating in the nitrogen near the top of the picture. (b) This scene is about 390 kilometers across. The rounded mountains, quite different from those we know on Earth, are named Tartarus Dorsa. The patterns, made of repeating ridges with the more reddish terrain between them, are not yet understood. (credit a, b: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/ Southwest Research Institute)

A Quick Look at Charon

To add to the mysteries of Pluto, we show in Figure 12.24 one of the best New Horizons images of Pluto's large moon Charon. Recall from earlier that Charon is roughly half Pluto's size (its diameter is about the size of Texas). Charon keeps the same side toward Pluto, just as our Moon keeps the same side toward Earth. What is unique about the Pluto-Charon system, however, is that Pluto also keeps its same face toward Charon. Like two dancers embracing, these two constantly face each other as they spin across the celestial dance floor. Astronomers call this a double tidal lock.



(a) (b)

Figure 12.24 : Pluto's Large Moon Charon. (a) In this New Horizons image, the color has been enhanced to bring out the color of the moon's strange red polar cap. Charon has a diameter of 1214 kilometers, and the resolution of this image is 3 kilometers. (b) Here we see the moon from a slightly different angle, in true color. The inset shows an area about 390 kilometers from top to bottom. Near the top left is an intriguing feature—what appears to be a mountain in the middle of a depression or moat. (credit a, b: modification of work by NASA/JHUAPL/SwRI)

What New Horizons showed was another complex world. There are scattered craters in the lower part of the image, but much of the rest of the surface appears smooth. Crossing the center of the image is a belt of rough terrain, including what appear to be tectonic valleys, as if some forces had tried to split Charon apart. Topping off this strange image is a distinctly red polar cap, of unknown composition. Many features on Charon are not yet understood, including what appears to be a mountain in the midst of a low-elevation region.

Footnotes

- 2In the process of convection, warmer material rises and colder material sinks, transferring energy. You are probably more familiar with convection in warm fluids, like the cells you see when you are heating some kinds of soup, but it can also happen (very slowly) in colder substances.

This page titled 10.6: Pluto and Charon is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.5: Pluto and Charon by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

10.7: Cosmic Influences on the Evolution of Earth

Learning Objectives

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

- Explain the scarcity of impact craters on Earth compared with other planets and moons
- Describe the evidence for recent impacts on Earth
- Detail how a massive impact changed the conditions for life on Earth, leading to the extinction of the dinosaurs
- Describe how impacts have influenced the evolution of life on Earth
- Discuss the search for objects that could potentially collide with our planet

In discussing Earth's geology earlier in this chapter, we dealt only with the effects of internal forces, expressed through the processes of plate tectonics and volcanism. On the Moon, in contrast, we see primarily craters, produced by the impacts of interplanetary debris such as asteroids and comets. Why don't we see more evidence on Earth of the kinds of impact craters that are so prominent on the Moon and other worlds?

Where Are the Craters on Earth?

It is not possible that Earth escaped being struck by the interplanetary debris that has pockmarked the Moon. From a cosmic perspective, the Moon is almost next door. Our atmosphere does make small pieces of cosmic debris burn up (which we see as *meteors*—commonly called shooting stars). But, the layers of our air provide no shield against the large impacts that form craters several kilometers in diameter and are common on the Moon.

In the course of its history, Earth must therefore have been impacted as heavily as the Moon. The difference is that, on Earth, these craters are destroyed by our active geology before they can accumulate. As plate tectonics constantly renews our crust, evidence of past cratering events is slowly erased. Only in the past few decades have geologists succeeded in identifying the eroded remnants of many impact craters (Figure 8.19). Even more recent is our realization that, over the history of Earth, these impacts have had an important influence on the evolution of life.

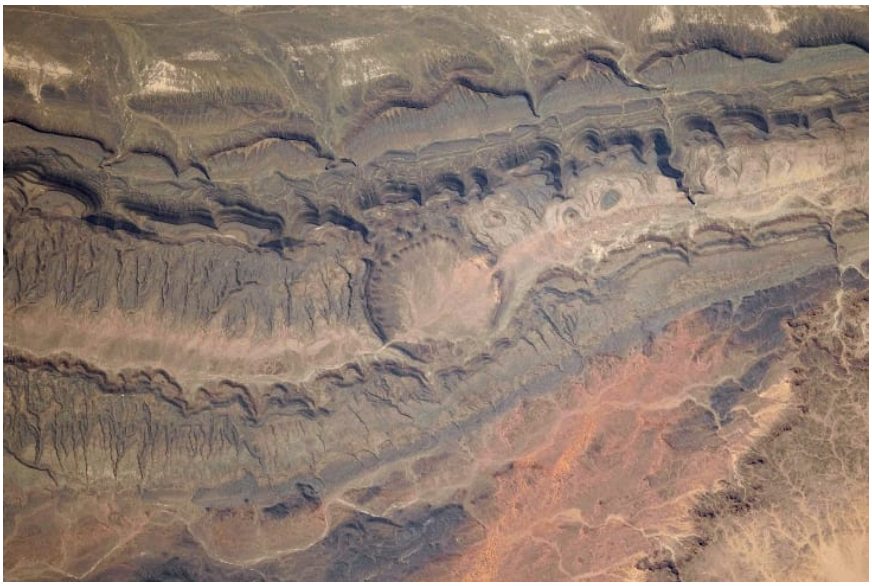


Figure 8.19 : Ouarkziz Impact Crater. Located in Algeria, this 4-km-wide crater (the round feature in the center) is the result of a meteor impact during the Cretaceous period. Although the crater has experienced heavy erosion, this image from the International Space Station shows the circular pattern resulting from impact. (credit: modification of work by NASA)

Recent Impacts

The collision of interplanetary debris with Earth is not a hypothetical idea. Evidence of relatively recent impacts can be found on our planet's surface. One well-studied historic collision took place on June 30, 1908, near the Tunguska River in Siberia. In this desolate region, there was a remarkable explosion in the atmosphere about 8 kilometers above the surface. The shock wave flattened more than a thousand square kilometers of forest (Figure 8.20). Herds of reindeer and other animals were killed, and a man at a trading post 80 kilometers from the blast was thrown from his chair and knocked unconscious. The blast wave spread around the world, as recorded by instruments designed to measure changes in atmospheric pressure.



Figure 8.20 : Aftermath of the Tunguska Explosion. This photograph, taken 19 years after the blast, shows a part of the forest that was destroyed by the 5-megaton explosion, resulting when a stony projectile about the size of a small office building (40 meters in diameter) collided with our planet. (credit: modification of work by Leonid Kulik)

Despite this violence, no craters were formed by the Tunguska explosion. Shattered by atmospheric pressure, the stony projectile with a diameter of approximately 60 meters disintegrated above our planet's surface to create a blast equivalent to a 10-megaton nuclear bomb. Had it been smaller or more fragile, the impacting body would have dissipated its energy at high altitude and probably attracted no attention. Today, such high-altitude atmospheric explosions are monitored regularly by military surveillance systems.

If it had been larger or made of stronger material (such as metal), the Tunguska projectile would have penetrated all the way to the surface of Earth and exploded to form a crater. Instead, only the heat and shock of the atmospheric explosion reached the surface, but the devastation it left behind in Siberia bore witness to the power of such impacts. Imagine if the same rocky impactor had exploded over New York City in 1908; history books might today record it as one of the most deadly events in human history.

Tens of thousands of people witnessed directly the explosion of a smaller (20-meter) projectile over the Russian city of Chelyabinsk on an early winter morning in 2013. It exploded at a height of 21 kilometers in a burst of light brighter than the Sun, and the shockwave of the 0.5-megaton explosion broke tens of thousands of windows and sent hundreds of people to the hospital. Rock fragments (meteorites) were easily collected by people in the area after the blast because they landed on fresh snow.

Link to Learning

Dr. David Morrison, one of the original authors of this textbook, provides a nontechnical talk about the Chelyabinsk explosion, and impacts in general.

The best-known recent crater on Earth was formed about 50,000 years ago in Arizona. The projectile in this case was a lump of iron about 40 meters in diameter. Now called *Meteor Crater* and a major tourist attraction on the way to the Grand Canyon, the

crater is about a mile across and has all the features associated with similar-size lunar impact craters (Figure 8.21). Meteor Crater is one of the few impact features on Earth that remains relatively intact; some older craters are so eroded that only a trained eye can distinguish them. Nevertheless, more than 150 have been identified. (See the list of suggested online sites at the end of this chapter if you want to find out more about these other impact scars.)



Figure 8.21 : Meteor Crater in Arizona. Here we see a 50,000-year-old impact crater made by the collision of a 40-meter lump of iron with our planet. Although impact craters are common on less active bodies such as the Moon, this is one of the very few well-preserved craters on Earth. (modification of work by D. Roddy/USGS)

Mass Extinction

The impact that produced Meteor Crater would have been dramatic indeed to any humans who witnessed it (from a safe distance) since the energy release was equivalent to a 10-megaton nuclear bomb. But such explosions are devastating only in their local areas; they have no *global* consequences. Much larger (and rarer) impacts, however, can disturb the ecological balance of the entire planet and thus influence the course of evolution.

The best-documented large impact took place 65 million years ago, at the end of what is now called the Cretaceous period of geological history. This time in the history of life on Earth was marked by a mass extinction, in which more than half of the species on our planet died out. There are a dozen or more mass extinctions in the geological record, but this particular event (nicknamed the “great dying”) has always intrigued paleontologists because it marks the end of the dinosaur age. For tens of millions of years these great creatures had flourished and dominated. Then, they suddenly disappeared (along with many other species), and thereafter mammals began the development and diversification that ultimately led to all of us.

The object that collided with Earth at the end of the Cretaceous period struck a shallow sea in what is now the Yucatán peninsula of Mexico. Its mass must have been more than a trillion tons, determined from study of a worldwide layer of sediment deposited from the dust cloud that enveloped the planet after its impact. First identified in 1979, this sediment layer is rich in the rare metal iridium and other elements that are relatively abundant in asteroids and comets, but exceedingly rare in Earth’s crust. Even though it was diluted by the material that the explosion excavated from the surface of Earth, this cosmic component can still be identified. In addition, this layer of sediment contains many minerals characteristic of the temperatures and pressures of a gigantic explosion.

The impact that led to the extinction of dinosaurs released energy equivalent to 5 billion Hiroshima-size nuclear bombs and excavated a crater 200 kilometers across and deep enough to penetrate through Earth’s crust. This large crater, named Chicxulub for a small town near its center, has subsequently been buried in sediment, but its outlines can still be identified (Figure 8.22). The explosion that created the Chicxulub crater lifted about 100 trillion tons of dust into the atmosphere. We can determine this amount by measuring the thickness of the sediment layer that formed when this dust settled to the surface.



Figure 8.22 : Site of the Chicxulub Crater. This map shows the location of the impact crater created 65 million years ago on Mexico's Yucatán peninsula. The crater is now buried under more than 500 meters of sediment. (credit: modification of work by "Carport"/Wikimedia)

Such a quantity of airborne material would have blocked sunlight completely, plunging Earth into a period of cold and darkness that lasted several months. Many plants dependent on sunlight would have died, leaving plant-eating animals without a food supply. Other worldwide effects included large-scale fires (started by the hot, flying debris from the explosion) that destroyed much of the planet's forests and grasslands, and a long period in which rainwater around the globe was acidic. It was these environmental effects, rather than the explosion itself, that were responsible for the mass extinction, including the demise of the dinosaurs.

Impacts and the Evolution of Life

It is becoming clear that many—perhaps most—mass extinctions in Earth's long history resulted from a variety of other causes, but in the case of the dinosaur killer, the cosmic impact certainly played a critical role and may have been the "final straw" in a series of climactic disturbances that resulted in the "great dying."

A catastrophe for one group of living things, however, may create opportunities for another group. Following each mass extinction, there is a sudden evolutionary burst as new species develop to fill the ecological niches opened by the event. Sixty-five million years ago, our ancestors, the mammals, began to thrive when so many other species died out. We are the lucky beneficiaries of this process.

Impacts by comets and asteroids represent the only mechanisms we know of that could cause truly global catastrophes and seriously influence the evolution of life all over the planet. As paleontologist Stephen Jay Gould of Harvard noted, such a perspective changes fundamentally our view of biological evolution. The central issues for the survival of a species must now include more than just its success in competing with other species and adapting to slowly changing environments, as envisioned by Darwin's idea of natural selection. Also required is an ability to survive random global catastrophes due to impacts.

Still earlier in its history, Earth was subject to even larger impacts from the leftover debris of planet formation. We know that the Moon was struck repeatedly by objects larger than 100 kilometers in diameter—1000 times more massive than the object that wiped out most terrestrial life 65 million years ago. Earth must have experienced similar large impacts during its first 700 million years of existence. Some of them were probably violent enough to strip the planet of most its atmosphere and to boil away its oceans. Such events would sterilize the planet, destroying any life that had begun. Life may have formed and been wiped out several times before our own microbial ancestors took hold sometime about 4 billion years ago.

The fact that the oldest surviving microbes on Earth are thermophiles (adapted to very high temperatures) can also be explained by such large impacts. An impact that was just a bit too small to sterilize the planet would still have destroyed anything that lived in what we consider "normal" environments, and only the creatures adapted to high temperatures would survive. Thus, the oldest surviving terrestrial lifeforms are probably the remnants of a sort of evolutionary bottleneck caused by repeated large impacts early in the planet's history.

Impacts in Our Future?

The impacts by asteroids and comets that have had such a major influence on life are not necessarily a thing of the past. In the full scope of planetary history, 65 million years ago was just yesterday. Earth actually orbits the Sun within a sort of cosmic shooting gallery, and although major impacts are rare, they are by no means over. Humanity could suffer the same fate as the dinosaurs, or lose a city to the much more frequent impacts like the one over Tunguska, unless we figure out a way to predict the next big impact and to protect our planet. The fact that our solar system is home to some very large planets in outer orbits may be beneficial to us; the gravitational fields of those planets can be very effective at pulling in cosmic debris and shielding us from larger, more frequent impacts.

Beginning in the 1990s, a few astronomers began to analyze the cosmic impact hazard and to persuade the government to support a search for potentially hazardous asteroids. Several small but sophisticated wide-field telescopes are now used for this search, which is called the NASA Spaceguard Survey. Already we know that there are currently no asteroids on a collision course with Earth that are as big (10–15 kilometers) as the one that killed the dinosaurs. The Spaceguard Survey now concentrates on finding smaller potential impactors. By 2021, the search had netted more than 98 percent of the hazardous near-Earth asteroids larger than 1 kilometer. The Spaceguard Survey will soon be augmented by the Vera Rubin Telescope that will scan the full sky on a regular basis, and an orbiting search telescope is under development by NASA. Of course, we cannot make a similar statement about the asteroids that have not yet been discovered, but these will be found and evaluated one by one for their potential hazard. These asteroid surveys are one of the few really life-and-death projects carried out by astronomers, with a potential to help to save our planet from future major impacts. Scientists are now studying how we could protect ourselves from an impact if a large asteroid were discovered on a collision course. The NASA spacecraft called DART (Double Asteroid Redirection Test), launched in late 2021, will demonstrate one example of such technology by crashing into the small asteroid Diomorphos to slightly alter its orbit.

Link to Learning

The Torino Impact Hazard Scale is a method for categorizing the impact hazard associated with near-Earth objects such as asteroids and comets. It is a communication tool for astronomers and the public to assess the seriousness of collision predictions by combining probability statistics and known kinetic damage potentials into a single threat value.

Purdue University's "Impact: Earth" calculator lets you input the characteristics of an approaching asteroid to determine the effect of its impact on our planet.

This page titled 10.7: Cosmic Influences on the Evolution of Earth is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 8.6: Cosmic Influences on the Evolution of Earth by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11: THE SUN, OUR STAR

11.1: Sources of Sunshine- Thermal and Gravitational Energy
11.2: Mass, Energy, and the Theory of Relativity
11.3: The Structure and Composition of the Sun
11.4: The Solar Interior - Theory
11.5: The Solar Interior - Observations
11.6: The Solar Cycle
11.7: Solar Activity above the Photosphere
11.8: Space Weather

This page titled 11: The Sun, Our Star is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

11.1: Sources of Sunshine- Thermal and Gravitational Energy

Learning Objectives

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

- Identify different forms of energy
- Understand the law of conservation of energy
- Explain ways that energy can be transformed

Energy is a challenging concept to grasp because it exists in so many different forms that it defies any single simple explanation. In many ways, comprehending energy is like comprehending wealth: There are very different forms of wealth and they follow different rules, depending on if they are the stock market, real estate, a collection of old comic books, great piles of cash, or one of the many other ways to make and lose money. It is easier to discuss one or two forms of wealth—or energy—than to discuss that concept in general.

When striving to understand how the Sun can put out so much energy for so long, scientists considered many different types of energy. Nineteenth-century scientists knew of two possible sources for the Sun's energy: chemical and gravitational energy. The source of chemical energy most familiar to them was the burning (the chemical term is *oxidation*) of wood, coal, gasoline, or other fuel. We know exactly how much energy the burning of these materials can produce. We can thus calculate that even if the immense mass of the Sun consisted of a burnable material like coal or wood, our star could not produce energy at its present rate for more than few thousand years. However, we know from geologic evidence that water was present on Earth's surface nearly 4 billion years ago, so the Sun must have been shining brightly (and making Earth warm) at least as long as that. Today, we also know that at the temperatures found in the Sun, nothing like solid wood or coal could survive.

Astronomy Basics: What's Watt?

Just a word about the units we are using. A watt (W) is a unit of *power*, which is energy used or given off per unit time. It is measured in joules per second (J/s). You know from your everyday experience that it is not just *how much* energy you expend, but *how long* you take to do it. (Burning 10 Calories in 10 minutes requires a very different kind of exercise than burning those 10 Calories in an hour.) Watts tell you the *rate* at which energy is being used; for example, a 100-watt bulb uses 100 joules (J) of energy every second.

And how big is a joule? A 73-kilogram (160-pound) astronomy instructor running at about 4.4 meters per second (10 miles per hour) because he is late for class has a motion energy of about 700 joules.

Conservation of Energy

Other nineteenth-century attempts to determine what makes the Sun shine used the law of conservation of energy. Simply stated, this law says that energy cannot be created or destroyed, but can be transformed from one type to another, such as from heat to mechanical energy. The steam engine, which was key to the Industrial Revolution, provides a good example. In this type of engine, the hot steam from a boiler drives the movement of a piston, converting heat energy into motion energy.

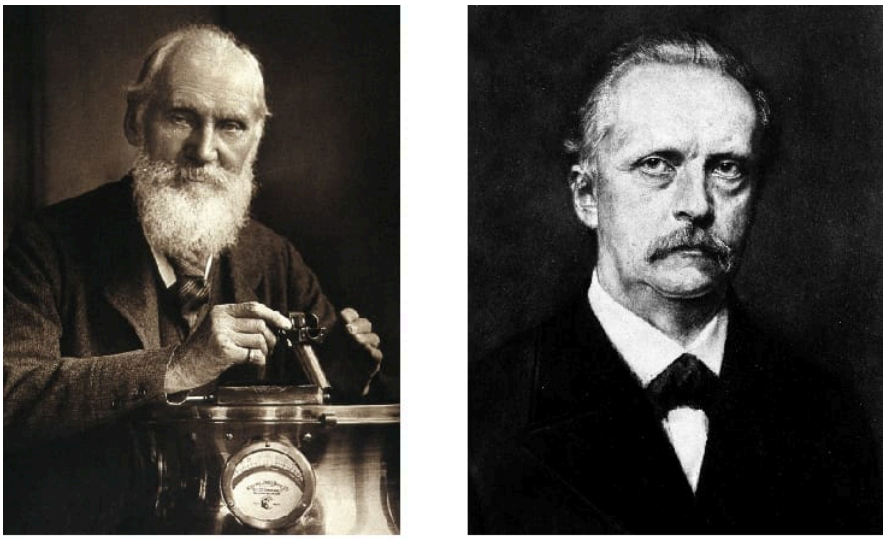
Conversely, motion can be transformed into heat. If you clap your hands vigorously at the end of an especially good astronomy lecture, your palms become hotter. If you rub ice on the surface of a table, the heat produced by friction melts the ice. The brakes on cars use friction to reduce speed, and in the process, transform motion energy into heat energy. That is why after bringing a car to a stop, the brakes can be very hot; this also explains why brakes can overheat when used carelessly while descending long mountain roads.

In the nineteenth century, scientists thought that the source of the Sun's heat might be the mechanical motion of meteorites falling into it. Their calculations showed, however, that in order to produce the total amount of energy emitted by the Sun, the mass in meteorites that would have to fall into the Sun every 100 years would equal the mass of Earth. The resulting increase in

the Sun's mass would, according to Kepler's third law, change the period of Earth's orbit by 2 seconds per year. Such a change would be easily measurable and was not, in fact, occurring. Scientists could then disprove this as the source of the Sun's energy.

Gravitational Contraction as a Source of Energy

Proposing an alternative explanation, British physicist Lord Kelvin and German scientist Hermann von Helmholtz (Figure 16.2), in about the middle of the nineteenth century, proposed that the Sun might produce energy by the conversion of gravitational energy into heat. They suggested that the outer layers of the Sun might be "falling" inward because of the force of gravity. In other words, they proposed that the Sun could be shrinking in size, staying hot and bright as a result.



(a) (b)

Figure 16.2 : Kelvin (1824–1907) and Helmholtz (1821–1894). (a) British physicist William Thomson (Lord Kelvin) and (b) German scientist Hermann von Helmholtz proposed that the contraction of the Sun under its own gravity might account for its energy. (credit a: modification of work by Wellcome Library, London; credit b: modification of work by Wellcome Library, London)

To imagine what would happen if this hypothesis were true, picture the outer layer of the Sun starting to fall inward. This outer layer is a gas made up of individual atoms, all moving about in random directions. If a layer falls inward, the atoms acquire an additional speed because of falling motion. As the outer layer falls inward, it also contracts, moving the atoms closer together. Collisions become more likely, and some of them transfer the extra speed associated with the falling motion to other atoms. This, in turn, increases the speeds of those atoms. The temperature of a gas is a measure of the kinetic energy (motion) of the atoms within it; hence, the temperature of this layer of the Sun increases. Collisions also excite electrons within the atoms to higher-energy orbits. When these electrons return to their normal orbits, they emit photons, which can then escape from the Sun (see Radiation and Spectra).

Kelvin and Helmholtz calculated that a contraction of the Sun at a rate of only about 40 meters per year would be enough to produce the amount of energy that it is now radiating. Over the span of human history, the decrease in the Sun's size from such a slow contraction would be undetectable.

If we assume that the Sun began its life as a large, diffuse cloud of gas, then we can calculate how much energy has been radiated by the Sun during its entire lifetime as it has contracted from a very large diameter to its present size. The amount of energy is on the order of 10^{42} joules. Since the solar luminosity is 4×10^{26} watts (joules/second) or about 10^{34} joules per year, contraction could keep the Sun shining at its present rate for roughly 100 million years.

In the nineteenth century, 100 million years at first seemed plenty long enough, since Earth was then widely thought to be much younger than this. But toward the end of that century and into the twentieth, geologists and physicists showed that Earth (and, hence, the Sun) is actually much older. Contraction therefore cannot be the primary source of solar energy (although, as we shall see in The Birth of Stars and the Discovery of Planets Outside the Solar System, contraction is an important source of energy for a while in stars that are just being born). Scientists were thus confronted with a puzzle of enormous proportions. Either an unknown type of energy was responsible for the most important energy source known to humanity, or estimates of the age of the solar system (and life on Earth) had to be seriously modified. Charles Darwin, whose theory of evolution required a longer time

span than the theories of the Sun seemed to permit, was discouraged by these results and continued to worry about them until his death in 1882.

It was only in the twentieth century that the true source of the Sun's energy was identified. The two key pieces of information required to solve the puzzle were the structure of the nucleus of the atom and the fact that mass can be converted into energy.

This page titled 11.1: Sources of Sunshine- Thermal and Gravitational Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.2: Sources of Sunshine- Thermal and Gravitational Energy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.2: Mass, Energy, and the Theory of Relativity

Learning Objectives

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

- Explain how matter can be converted into energy
- Describe the particles that make up atoms
- Describe the nucleus of an atom
- Understand the nuclear forces that hold atoms together
- Trace the nuclear reactions in the solar interior

As we have seen, energy cannot be created or destroyed, but only converted from one form to another. One of the remarkable conclusions derived by Albert Einstein (see Albert Einstein) when he developed his theory of relativity is that matter can be considered a form of energy too and can be converted into energy. Furthermore, energy can also be converted into matter. This seemed to contradict what humans had learned over thousands of years by studying nature. Matter is something we can see and touch, whereas energy is something objects have when they do things like move or heat up. The idea that matter or energy can be converted into each other seemed as outrageous as saying you could accelerate a car by turning the bumper into more speed, or that you could create a bigger front seat by slowing down your car. That would be pretty difficult to believe; yet, the universe actually works somewhat like that.

Converting Matter into Energy

The remarkable equivalence between matter and energy is given in one of the most famous equations:

$$E =$$

In this equation, *E* stands for energy, *m* stands for mass, and *c*, the constant that relates the two, is the speed of light $\left(3 \times 10^{8} \text{meters per second}\right)$. Note that mass is a measure of the quantity of matter, so the significance of this equation is that matter can be converted into energy and energy can be converted into matter. Let's compare this equation of converting matter and energy to some common conversion equations that have the same form:

$$\text{inches} = \text{feet} \times 12, \text{or cents} =$$

Just as each conversion formula allows you to calculate the conversion of one thing into another, when we convert matter into energy, we consider how much mass the matter has. The conversion factor in this case turns out not to be either 12 or 100, as in our examples, but another constant quantity: the speed of light squared. Note that matter does not have to travel at the speed of light (or the speed of light squared) for this conversion to occur. The factor of c^2 is just the number that Einstein showed must be used to relate mass and energy.

Notice that this formula does not tell us *how* to convert mass into energy, just as the formula for cents does not tell us where to exchange coins for a dollar bill. The formulas merely tell us what the equivalent values are if we succeed in making the conversion. When Einstein first derived his formula in 1905, no one had the faintest idea how to convert mass into energy in any practical way. Einstein himself tried to discourage speculation that the large-scale conversion of atomic mass into energy would be feasible in the near future. Today, as a result of developments in nuclear physics, we regularly convert mass into energy in power plants, nuclear weapons, and high-energy physics experiments in particle accelerators.

Because the speed of light squared (c^2) is a very large quantity, the conversion of even a small amount of mass results in a very large amount of energy. For example, the complete conversion of 1 gram of matter (about 1/28 ounce, or approximately 1 paperclip) would produce as much energy as the burning of 15,000 barrels of oil.

Scientists soon realized that the conversion of mass into energy is the source of the Sun's heat and light. With Einstein's equation of $E = mc^2$, we can calculate that the amount of energy radiated by the Sun could be produced by the complete conversion of about 4 million tons of matter into energy inside the Sun each second. Destroying 4 million tons per second sounds like a lot

when compared to earthly things, but bear in mind that the Sun is a very big reservoir of matter. In fact, we will see that the Sun contains more than enough mass to destroy such huge amounts of matter and still continue shining at its present rate for billions of years.

But knowing all that still does not tell us *how* mass can be converted into energy. To understand the process that actually occurs in the Sun, we need to explore the structure of the atom a bit further.

Voyagers in Astronomy: Albert Einstein

For a large part of his life, Albert Einstein (Figure 16.3) was one of the most recognized celebrities of his day. Strangers stopped him on the street, and people all over the world asked him for endorsements, advice, and assistance. In fact, when Einstein and the great film star Charlie Chaplin met in California, they found they shared similar feelings about the loss of privacy that came with fame. Einstein's name was a household word despite the fact that most people did not understand the ideas that had made him famous.

Einstein was born in 1879 in Ulm, Germany. Legend has it that he did not do well in school (even in arithmetic), and thousands of students have since attempted to justify a bad grade by referring to this story. Alas, like many legends, this one is not true. Records indicate that although he tended to rebel against the authoritarian teaching style in vogue in Germany at that time, Einstein was a good student.

After graduating from the Federal Polytechnic Institute in Zurich, Switzerland, Einstein at first had trouble getting a job (even as a high school teacher), but he eventually became an examiner in the Swiss Patent Office. Working in his spare time, without the benefit of a university environment but using his superb physical intuition, he wrote four papers in 1905 that would ultimately transform the way physicists looked at the world.

One of these, which earned Einstein the Nobel Prize in 1921, set part of the foundation of *quantum mechanics*—the rich, puzzling, and remarkable theory of the subatomic realm. But his most important paper presented the *special theory of relativity*, a reexamination of space, time, and motion that added a whole new level of sophistication to our understanding of those concepts. The famed equation $E = mc^2$ was actually a relatively minor part of this theory, added in a later paper.

In 1916, Einstein published his *general theory of relativity*, which was, among other things, a fundamentally new description of gravity (see Black Holes and Curved Spacetime). When this theory was confirmed by measurements of the "bending of starlight" during a 1919 eclipse (*The New York Times* headline read, "Lights All Askew in the Heavens"), Einstein became world famous.

In 1933, to escape Nazi persecution, Einstein left his professorship in Berlin and settled in the United States at the newly created Institute for Advanced Studies at Princeton. He remained there until his death in 1955, writing, lecturing, and espousing a variety of intellectual and political causes. For example, he agreed to sign a letter written by Leo Szilard and other scientists in 1939, alerting President Roosevelt to the dangers of allowing Nazi Germany to develop the atomic bomb first. And in 1952, Einstein was offered the second presidency of Israel. In declining the position, he said, "I know a little about nature and hardly anything about men."



Figure 16.3 : Albert Einstein (1879–1955). This portrait of Einstein was taken in 1912. (credit: modification of work by J. F. Langhans)

Elementary Particles

The fundamental components of atoms are the proton, neutron, and electron (see The Structure of the Atom).

Protons, neutrons, and electrons are by no means all the particles that exist. First, for each kind of particle, there is a corresponding but opposite *antiparticle*. If the particle carries a charge, its antiparticle has the opposite charge. The antielectron is the *positron*, which has the same mass as the electron but is positively charged. Similarly, the antiproton has a negative charge. The remarkable thing about such *antimatter* is that when a particle comes into contact with its antiparticle, the original particles are annihilated, and substantial amounts of energy in the form of photons are produced.

Since our world is made exclusively of ordinary particles of matter, antimatter cannot survive for very long. But individual antiparticles are found in cosmic rays (particles that arrive at the top of Earth's atmosphere from space) and can be created in particle accelerators. And, as we will see in a moment, antimatter is created in the core of the Sun and other stars.

Science fiction fans may be familiar with antimatter from the *Star Trek* television series and films. The Starship Enterprise is propelled by the careful combining of matter and antimatter in the ship's engine room. According to $E = mc^2$, the annihilation of matter and antimatter can produce a huge amount of energy, but keeping the antimatter fuel from touching the ship before it is needed must be a big problem. No wonder Scotty, the chief engineer in the original TV show, always looked worried!

In 1933, physicist Wolfgang Pauli (Figure 16.4) suggested that there might be another type of elementary particle. Energy seemed to disappear when certain types of nuclear reactions took place, violating the law of conservation of energy. Pauli was reluctant to accept the idea that one of the basic laws of physics was wrong, and he suggested a "desperate remedy." Perhaps a so-far-undetected particle, which was given the name neutrino ("little neutral one"), carried away the "missing" energy. He suggested that neutrinos were particles with zero mass, and that like photons, they moved with the speed of light.



Figure 16.4 : Wolfgang Pauli in 1945. Pauli is considered the “father” of the neutrino, having conceived of it in 1933.

The elusive neutrino was not detected until 1956. The reason it was so hard to find is that neutrinos interact very weakly with other matter and therefore are very difficult to detect. Earth is more transparent to a neutrino than the thinnest and cleanest pane of glass is to a photon of light. In fact, most neutrinos can pass completely through a star or planet without being absorbed. As we shall see, this behavior of neutrinos makes them a very important tool for studying the Sun. Since Pauli’s prediction, scientists have learned a lot more about the neutrino. We now know that there are three different types of neutrinos, and in 1998, neutrinos were discovered to have a tiny amount of mass. Indeed, it is so tiny that electrons are at least 500,000 times more massive. Ongoing research is focused on determining the mass of neutrinos more precisely, and it may still turn out that one of the three types is massless. We will return to the subject of neutrinos later in this chapter.

Some of the properties of the proton, electron, neutron, and neutrino are summarized in Table 16.1. (Other subatomic particles have been produced by experiments with particle accelerators, but they do not play a role in the generation of solar energy.)

Table 16.1: Properties of Some Common Particles

Particle	Mass (kg)	Charge
Proton	1.67265×10^{-27}	+1
Neutron	1.67495×10^{-27}	0
Electron	9.11×10^{-31}	–1
Neutrino	$< 2 \times 10^{-36}$ (uncertain)	0

The Atomic Nucleus

The nucleus of an atom is not just a loose collection of elementary particles. Inside the nucleus, particles are held together by a very powerful force called the strong nuclear force. This is short-range force, only capable of acting over distances about the size of the atomic nucleus. A quick thought experiment shows how important this force is. Take a look at your finger and consider the atoms composing it. Among them is carbon, one of the basic elements of life. Focus your imagination on the nucleus of one of your carbon atoms. It contains six protons, which have a positive charge, and six neutrons, which are neutral. Thus, the nucleus has a net charge of six positives. If only the electrical force were acting, the protons in this and every carbon atom would find each other very repulsive and fly apart.

The strong nuclear force is an attractive force, stronger than the electrical force, and it keeps the particles of the nucleus tightly bound together. We saw earlier that if under the force of gravity a star “shrinks”—bringing its atoms closer together—gravitational energy is released. In the same way, if particles come together under the strong nuclear force and unite to form an atomic nucleus, some of the nuclear energy is released. The energy given up in such a process is called the *binding energy* of the nucleus.

When such binding energy is released, the resulting nucleus has slightly less mass than the sum of the masses of the particles that came together to form it. In other words, the energy comes from the loss of mass. This slight deficit in mass is only a small fraction of the mass of one proton. But because each bit of lost mass can provide a lot of energy (remember, $E = mc^2$), this nuclear energy release can be quite substantial.

Measurements show that the binding energy is greatest for atoms with a mass near that of the iron nucleus (with a combined number of protons and neutrons equal to 56) and less for both the lighter and the heavier nuclei. Iron, therefore, is the most stable element: since it gives up the most energy when it forms, it would require the most energy to break it back down into its component particles.

What this means is that, in general, when light atomic nuclei come together to form a heavier one (up to iron), mass is lost and energy is released. This joining together of atomic nuclei is called nuclear fusion.

Energy can also be produced by breaking up heavy atomic nuclei into lighter ones (down to iron); this process is called nuclear fission. Nuclear fission was the process we learned to use first—in atomic bombs and in nuclear reactors used to generate electrical power—and it may therefore be more familiar to you. Fission also sometimes occurs spontaneously in some unstable nuclei through the process of natural radioactivity. But fission requires big, complex nuclei, whereas we know that the stars are made up predominantly of small, simple nuclei. So we must look to fusion first to explain the energy of the Sun and the stars (Figure 16.5).

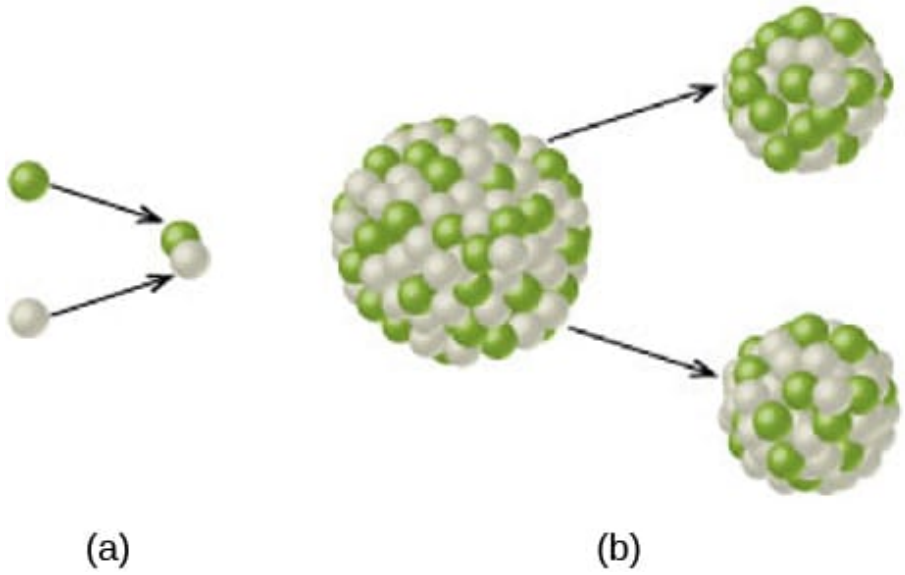


Figure 16.5 : Fusion and Fission. (a) In fusion, light atomic nuclei join together to form a heavier nuclei, releasing energy in the process. (b) In fission, energy is produced by the breaking up of heavy, complex nuclei into lighter ones.

Nuclear Attraction versus Electrical Repulsion

So far, we seem to have a very attractive prescription for producing the energy emitted by the Sun: “roll” some nuclei together and join them via nuclear fusion. This will cause them to lose some of their mass, which then turns into energy. However, every nucleus beyond simple hydrogen has two or more protons—and protons all have positive charges. Since like charges repel via the electrical force, the closer we get two nuclei to each other, the more they repel. It’s true that if we can get them within “striking distance” of the nuclear force, they will then come together with a much stronger attraction. But that striking distance is very tiny, about the size of a nucleus. How can we get nuclei close enough to participate in fusion?

The answer turns out to be heat—tremendous heat—which speeds the protons up enough to overcome the electrical forces that try to keep protons apart. Inside the Sun, as we saw, the most common element is hydrogen, whose nucleus contains only a single proton. Two protons can fuse only in regions where the temperature is greater than about 12 million K, and the speed of the protons average around 1000 kilometers per second or more. (In old-fashioned units, that’s over 2 million miles per hour!)

In our Sun, such extreme temperatures are reached only in the regions near its center, which has a temperature of 15 million K. Calculations show that nearly all of the Sun’s energy is generated within about 150,000 kilometers of its core, or within less than

10% of its total volume.

Even at these high temperatures, it is exceedingly difficult to force two protons to combine. On average, a proton will rebound from other protons in the Sun's crowded core for about 14 billion years, at the rate of *100 million collisions per second*, before it fuses with a second proton. This is, however, only the *average* waiting time. Some of the enormous numbers of protons in the Sun's inner region are "lucky" and take only a few collisions to achieve a fusion reaction: they are the protons responsible for producing the energy radiated by the Sun. Since the Sun is about 4.5 billion years old, most of its protons have not yet been involved in fusion reactions.

Nuclear Reactions in the Sun's Interior

The Sun, then, taps the energy contained in the nuclei of atoms through nuclear fusion. Let's look at what happens in more detail. Deep inside the Sun, a three-step process takes four hydrogen nuclei and fuses them together to form a single helium nucleus. The helium nucleus is slightly less massive than the four hydrogen nuclei that combine to form it, and that mass is converted into energy.

The initial step required to form one helium nucleus from four hydrogen nuclei is shown in Figure 16.6. At the high temperatures inside the Sun's core, two protons combine to make a *deuterium* nucleus, which is an isotope (or version) of hydrogen that contains one proton and one neutron. In effect, one of the original protons has been converted into a neutron in the fusion reaction. Electric charge has to be conserved in nuclear reactions, and it is conserved in this one. A positron (antimatter electron) emerges from the reaction and carries away the positive charge originally associated with one of the protons.

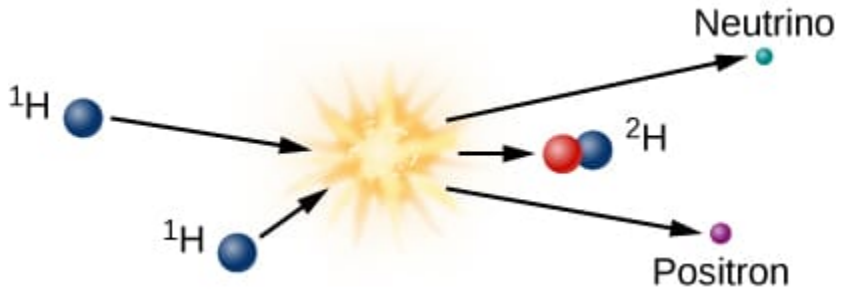


Figure 16.6 : Proton-Proton Chain, Step 1. This is the first step in the process of fusing hydrogen into helium in the Sun. High temperatures are required because this reaction starts with two hydrogen nuclei, which are protons (shown in blue at left) that must overcome electrical repulsion to combine, forming a hydrogen nucleus with a proton and a neutron (shown in red). Note that hydrogen containing one proton and one neutron is given its own name: deuterium. Also produced in this reaction are a positron, which is an antielectron, and an elusive particle named the neutrino.

Since it is antimatter, this positron will instantly collide with a nearby electron, and both will be annihilated, producing electromagnetic energy in the form of gamma-ray photons. This gamma ray, which has been created in the center of the Sun, finds itself in a world crammed full of fast-moving nuclei and electrons. The gamma ray collides with particles of matter and transfers its energy to one of them. The particle later emits another gamma-ray photon, but often the emitted photon has a bit less energy than the one that was absorbed.

Such interactions happen to gamma rays again and again and again as they make their way slowly toward the outer layers of the Sun, until their energy becomes so reduced that they are no longer gamma rays but X-rays (recall what you learned in The Electromagnetic Spectrum). Later, as the photons lose still more energy through collisions in the crowded center of the Sun, they become ultraviolet photons.

By the time they reach the Sun's surface, most of the photons have given up enough energy to be ordinary light—and they are the sunlight we see coming from our star. (To be precise, each gamma-ray photon is ultimately converted into many separate lower-energy photons of sunlight.) So, the sunlight given off by the Sun today had its origin as a gamma ray produced by nuclear reactions deep in the Sun's core. The length of time that photons require to reach the surface depends on how far a photon on average travels between collisions, and the travel time depends on what model of the complicated solar interior we accept. Estimates are somewhat uncertain but indicate that the emission of energy from the surface of the Sun can lag its production in the interior by 100,000 years to as much as 1,000,000 years.

In addition to the positron, the fusion of two hydrogen atoms to form deuterium results in the emission of a neutrino. Because

neutrinos interact so little with ordinary matter, those produced by fusion reactions near the center of the Sun travel directly to the Sun's surface and then out into space, in all directions. Neutrinos move at nearly the speed of light, and they escape the Sun about two seconds after they are created.

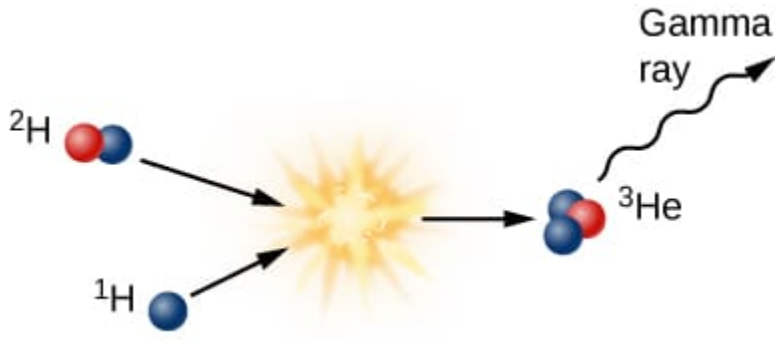


Figure 16.7 : Proton-Proton Chain, Step 2. This is the second step of the proton-proton chain, the fusion reaction that converts hydrogen into helium in the Sun. This step combines one hydrogen nucleus, which is a proton (shown in blue), with the deuterium nucleus from the previous step (shown as a red and blue particle). The product of this is an isotope of helium with two protons (blue) and one neutron (red) and energy in the form of gamma-ray radiation.

The second step in forming helium from hydrogen is to add another proton to the deuterium nucleus to create a helium nucleus that contains two protons and one neutron (Figure 16.7). In the process, some mass is again lost and more gamma radiation is emitted. Such a nucleus is helium because an element is defined by its number of protons; any nucleus with two protons is called helium. But this form of helium, which we call helium-3 (and write in shorthand as ^{3}He) is not the isotope we see in the Sun's atmosphere or on Earth. That helium has two neutrons and two protons and hence is called helium-4 (^{4}He).

To get to helium-4 in the Sun, helium-3 must combine with another helium-3 in the third step of fusion (illustrated in Figure 16.8). Note that two energetic protons are left over from this step; each of them comes out of the reaction ready to collide with other protons and to start step 1 in the chain of reactions all over again.

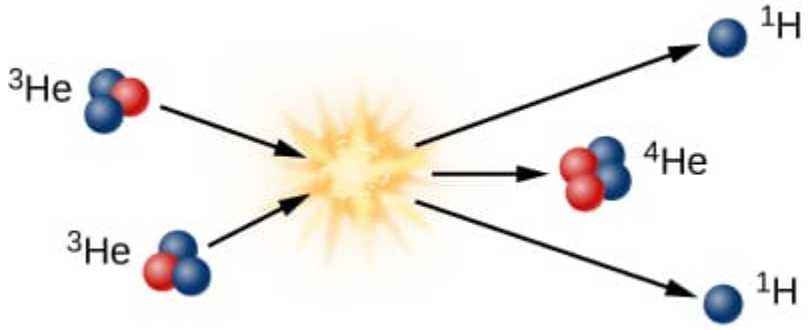


Figure 16.8 : Proton-Proton Chain, Step 3. This is the third step in the fusion of hydrogen into helium in the Sun. Note that the two helium-3 nuclei from the second step (see Figure 16.7) must combine before the third step becomes possible. The two protons that come out of this step have the energy to collide with other protons in the Sun and start step one again.

Link to Learning

Visit the Tokamak Fusion Reactor at the General Atomics Lab in San Diego, CA, for an 8-minute tour.

The Proton-Proton Chain

The nuclear reactions in the Sun that we have been discussing can be described succinctly through the following nuclear formulas:

$$^1\mathrm{H} + {}^1\mathrm{H} \longrightarrow {}^2\mathrm{H} + \mathrm{e}^+ + v$$
$$^2\mathrm{H} + {}^1\mathrm{H} \longrightarrow {}^3\mathrm{He} + \gamma$$
$$^3\mathrm{He} + {}^3\mathrm{He} \longrightarrow {}^4\mathrm{He} + {}^1\mathrm{H} + {}^1\mathrm{H}$$

Here, the superscripts indicate the total number of neutrons plus protons in the nucleus, e^+ is the positron, *v* is the neutrino, and *y* indicates that gamma rays are emitted. Note that the third step requires two helium-3 nuclei to start; the first two steps must happen twice before the third step can occur.

Although, as we discussed, the first step in this chain of reactions is very difficult and generally takes a long time, the other steps happen more quickly. After the deuterium nucleus is formed, it survives an average of only about 6 seconds before being converted into ^{3}He. About a million years after that (on average), the ^{3}He nucleus will combine with another to form ^{4}He.

We can compute the amount of energy these reactions generate by calculating the difference in the initial and final masses. The masses of hydrogen and helium atoms in the units normally used by scientists are 1.007825 *u* and 4.00268 *u*, respectively. (The unit of mass, *u*, is defined to be 1/12 the mass of an atom of carbon, or approximately the mass of a proton.) Here, we include the mass of the entire atom, not just the nucleus, because electrons are involved as well. When hydrogen is converted into helium, two positrons are created (remember, the first step happens twice), and these are annihilated with two free electrons, adding to the energy produced.

$$\begin{aligned} 4 \times 1.007825u &= 4.03130\,u\ (\text{mass of initial hydrogen atoms}) \\ &\quad - 4.00268\,u\ (\text{mass of final helium atoms}) \\ &= 0.02862\,u\ (\text{mass lost in the transformation}) \end{aligned}$$

The mass lost, 0.02862 *u*, is 0.71% of the mass of the initial hydrogen. Thus, if 1 kilogram of hydrogen is converted into helium, then the mass of the helium is only 0.9929 kilogram, and 0.0071 kilogram of material is converted into energy. The speed of light (*c*) is 3 × 10^8 meters per second, so the energy released by the conversion of just 1 kilogram of hydrogen into helium is:

$$E = mc^2$$
$$E = 0.0071\ \text{kg} \times \left(3 \times 10^8 \text{m/s}\right)^2 = 6.4 \times 10^{14}\ \text{J}$$

This amount, the energy released when a single kilogram (2.2 pounds) of hydrogen undergoes fusion, would supply all the electricity used by the typical U.S. household for roughly 17,000 years.

To produce the Sun's luminosity of 4 × 10^{26} watts, some 600 million tons of hydrogen must be converted into helium *each second*, of which about 4 million tons are converted from matter into energy. As large as these numbers are, the store of hydrogen (and thus of nuclear energy) in the Sun is still *more* enormous, and can last a long time—billions of years, in fact.

At the temperatures inside the stars with masses smaller than about 1.2 times the mass of our Sun (a category that includes the Sun itself), most of the energy is produced by the reactions we have just described, and this set of reactions is called the proton-proton chain (or sometimes, the p-p chain). In the proton-proton chain, protons collide directly with other protons to form helium nuclei.

In hotter stars, another set of reactions, called the carbon-nitrogen-oxygen (CNO) cycle, accomplishes the same net result. In the CNO cycle, carbon and hydrogen nuclei collide to initiate a series of reactions that form nitrogen, oxygen, and ultimately, helium. The nitrogen and oxygen nuclei do not survive but interact to form carbon again. Therefore, the outcome is the same as in the proton-proton chain: four hydrogen atoms disappear, and in their place, a single helium atom is created. The CNO cycle plays only a minor role in the Sun but is the main source of energy for stars with masses greater than about the mass of the Sun.

So you can see that we have solved the puzzle that so worried scientists at the end of the nineteenth century. The Sun can maintain its high temperature and energy output for billions of years through the fusion of the simplest element in the universe, hydrogen. Because most of the Sun (and the other stars) is made of hydrogen, it is an ideal "fuel" for powering a star. As will be discussed in the following chapters, we can define a star as a ball of gas capable of getting its core hot enough to initiate the fusion of hydrogen. There are balls of gas that lack the mass required to do this (Jupiter is a local example); like so many hopefuls in Hollywood, they will never be stars.

Making Connections: Fusion on Earth

Wouldn't it be wonderful if we could duplicate the Sun's energy mechanism in a controlled way on Earth? (We have already duplicated it in an uncontrolled way in hydrogen bombs, but we hope our storehouses of these will never be used.) Fusion energy would have many advantages: it would use hydrogen (or deuterium, which is heavy hydrogen) as fuel, and there is

abundant hydrogen in Earth's lakes and oceans. Water is much more evenly distributed around the world than oil or uranium, meaning that a few countries would no longer hold an energy advantage over the others. And unlike fission, which leaves dangerous byproducts, the nuclei that result from fusion are perfectly safe.

The problem is that, as we saw, it takes extremely high temperatures for nuclei to overcome their electrical repulsion and undergo fusion. When the first hydrogen bombs were exploded in tests in the 1950s, the "fuses" to get them hot enough were fission bombs. Interactions at such temperatures are difficult to sustain and control. To make fusion power on Earth, after all, we have to do what the Sun does: produce temperatures and pressures high enough to get hydrogen nuclei on intimate terms with one another.

The European Union, the United States, South Korea, Japan, China, Russia, Switzerland, and India are collaborating on the International Thermonuclear Experimental Reactor (ITER), a project to demonstrate the feasibility of controlled fusion (Figure 16.9). The facility is being built in France. Construction will require over 10,000,000 components and 2000 workers for assembly. The date for the start of operations is yet to be determined.

ITER is based on the Tokamak design, in which a large doughnut-shaped container is surrounded by superconducting magnets to confine and control the hydrogen nuclei in a strong magnetic field. Previous fusion experiments have produced about 15 million watts of energy, but only for a second or two, and they have required 100 million watts to produce the conditions necessary to achieve fusion. The goal of ITER is to build the first fusion device capable of producing 500 million watts of fusion energy for up to 1000 seconds. The challenge is keeping the deuterium and tritium—which will participate in fusion reactions—hot enough and dense enough, for a long enough time to produce energy.

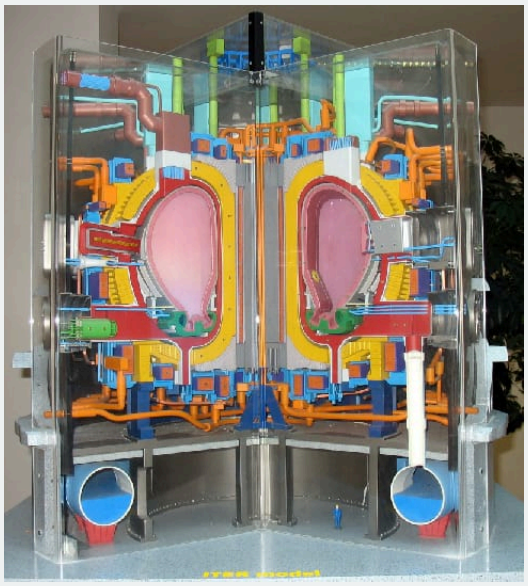


Figure 16.9 : ITER Design. The bright yellow areas in this model show where the superconducting magnets will circle the chamber within which fusion will take place. A huge magnet will keep the charged nuclei of heavy hydrogen confined. The goal is to produce 500 megawatts of energy. (credit: modification of work by Stephan Mosel)

This page titled 11.2: Mass, Energy, and the Theory of Relativity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.3: Mass, Energy, and the Theory of Relativity by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.3: The Structure and Composition of the Sun

Learning Objectives

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

- Explain how the composition of the Sun differs from that of Earth
- Describe the various layers of the Sun and their functions
- Explain what happens in the different parts of the Sun's atmosphere

The Sun, like all stars, is an enormous ball of extremely hot, largely ionized gas, shining under its own power. And we do mean enormous. The Sun could fit 109 Earths side-by-side across its diameter, and it has enough volume (takes up enough space) to hold about 1.3 million Earths.

The Sun does not have a solid surface or continents like Earth, nor does it have a solid core (Figure 15.2). However, it does have a lot of structure and can be discussed as a series of layers, not unlike an onion. In this section, we describe the huge changes that occur in the Sun's extensive interior and atmosphere, and the dynamic and violent eruptions that occur daily in its outer layers.



Figure 15.2 : Earth and the Sun. Here, Earth is shown to scale with part of the Sun and a giant loop of hot gas erupting from its surface. The inset shows the entire Sun, smaller. (credit: modification of work by SOHO/EIT/ESA)

Some of the basic characteristics of the Sun are listed in Table 15.1. Although some of the terms in that table may be unfamiliar to you right now, you will get to know them as you read further.

Table 15.1: Characteristics of the Sun

Characteristic	How Found	Value
Mean distance	Radar reflection from planets	1 AU (149,597,892 km)
Maximum distance from Earth		1.521×10^8 km
Minimum distance from Earth		1.471×10^8 km
Mass	Orbit of Earth	333,400 Earth masses (1.99×10^{30} kg)
Mean angular diameter	Direct measure	31′59′′.3

Characteristic	How Found	Value
Diameter of photosphere	Angular size and distance	109.3 × Earth diameter (1.39×10^6 km)
Mean density	Mass/volume	1.41 g/cm^3 (1400 kg/m^3)
Gravitational acceleration at photosphere (surface gravity)	GM/R^2	27.9 × Earth surface gravity = 273 m/s^2
Solar constant	Instrument sensitive to radiation at all wavelengths	1370 W/m^2
Luminosity	Solar constant × area of spherical surface 1 AU in radius	3.8×10^{26} W
Spectral class	Spectrum	G2V
Effective temperature	Derived from luminosity and radius of the Sun	5800 K
Rotation period at equator	Sunspots and Doppler shift in spectra taken at the edge of the Sun	24 days 16 hours
Inclination of equator to ecliptic	Motions of sunspots	7°10′.5

Composition of the Sun's Atmosphere

Let's begin by asking what the solar atmosphere is made of. As explained in Radiation and Spectra, we can use a star's *absorption line spectrum* to determine what elements are present. It turns out that the Sun contains the same elements as Earth but *not* in the same proportions. About 73% of the Sun's mass is hydrogen, and another 25% is helium. All the other chemical elements (including those we know and love in our own bodies, such as carbon, oxygen, and nitrogen) make up only 2% of our star. The 10 most abundant gases in the Sun's visible surface layer are listed in Table 15.2. Examine that table and notice that the composition of the Sun's outer layer is very different from Earth's crust, where we live. (In our planet's crust, the three most abundant elements are oxygen, silicon, and aluminum.) Although not like our planet's, the makeup of the Sun is quite typical of stars in general.

Table 15.2: The Abundance of Elements in the Sun

Element	Percentage by Number of Atoms	Percentage By Mass
Hydrogen	92.0	73.4
Helium	7.8	25.0
Carbon	0.02	0.20
Nitrogen	0.008	0.09
Oxygen	0.06	0.80
Neon	0.01	0.16
Magnesium	0.003	0.06
Silicon	0.004	0.09
Sulfur	0.002	0.05
Iron	0.003	0.14

The fact that our Sun and the stars all have similar compositions and are made up of mostly hydrogen and helium was first shown in a brilliant thesis in 1925 by Cecilia Payne-Gaposchkin, the first woman to get a PhD in astronomy in the United States (Figure 15.3). However, the idea that the simplest light gases—hydrogen and helium—were the most abundant elements in stars was so unexpected and so shocking that she was persuaded her analysis of the data must be wrong. At the time, she wrote, "The enormous abundance derived for these elements in the stellar atmosphere is almost certainly not real." Even scientists sometimes find it hard to accept new ideas that do not agree with what everyone "knows" to be right.



Figure 15.3 : Cecilia Payne-Gaposchkin (1900–1979). Her 1925 doctoral thesis laid the foundations for understanding the composition of the Sun and the stars. Yet, being a woman, she was not given a formal appointment at Harvard, where she worked, until 1938 and was not appointed a professor until 1956. (credit: Smithsonian Institution)

Before Payne-Gaposchkin's work, everyone assumed that the composition of the Sun and stars would be much like that of Earth. It was 3 years after her thesis that other studies proved beyond a doubt that the enormous abundance of hydrogen and helium in the Sun is indeed real. (And, as we will see, the composition of the Sun and the stars is much more typical of the makeup of the universe than the odd concentration of heavier elements that characterizes our planet.)

Most of the elements found in the Sun are in the form of atoms, with a small number of molecules, all in the form of gases: the Sun is so hot that no matter can survive as a liquid or a solid. In fact, the Sun is so hot that many of the atoms in it are *ionized*, that is, stripped of one or more of their electrons. This removal of electrons from their atoms means that there is a large quantity of free electrons and positively charged ions in the Sun, making it an electrically charged environment—quite different from the neutral one in which you are reading this text. (Scientists call such a hot ionized gas a plasma.)

In the nineteenth century, scientists observed a spectral line at 530.3 nanometers in the Sun's outer atmosphere, called the corona (a layer we will discuss in a minute.) This line had never been seen before, and so it was assumed that this line was the result of a new element found in the corona, quickly named coronium. It was not until 60 years later that astronomers discovered that this emission was in fact due to highly ionized iron—iron with 13 of its electrons stripped off. This is how we first discovered that the Sun's atmosphere had a temperature of more than a million degrees.

The Layers of the Sun beneath the Visible Surface

Figure 15.4 shows what the Sun would look like if we could see all parts of it from the center to its outer atmosphere; the terms in the figure will become familiar to you as you read on.

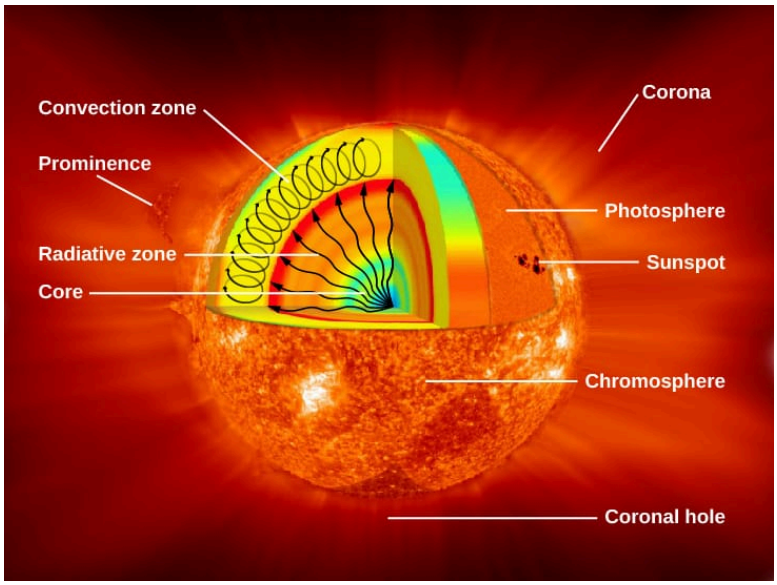


Figure 15.4 : Parts of the Sun. This illustration shows the different parts of the Sun, from the hot core where the energy is generated through regions where energy is transported outward, first by radiation, then by convection, and then out through the solar atmosphere. The parts of the atmosphere are also labeled the photosphere, chromosphere, and corona. Some typical features in the atmosphere are shown, such as coronal holes and prominences. (credit: modification of work by NASA/Goddard)

The Sun's layers are different from each other, and each plays a part in producing the energy that the Sun ultimately emits. We will begin with the core and work our way out through the layers. The Sun's *core* is extremely dense and is the source of all of its energy. Inside the core, nuclear energy is being released (in ways we will discuss in The Sun: A Nuclear Powerhouse). The core is approximately 20% of the size of the solar interior and is thought to have a temperature of approximately 15 million K, making it the hottest part of the Sun.

Above the core is a region known as the *radiative zone*—named for the primary mode of transporting energy across it. This region starts at about 25% of the distance to the solar surface and extends up to about 70% of the way to the surface. The light generated in the core is transported through the radiative zone very slowly, since the high density of matter in this region means a photon cannot travel too far without encountering a particle, causing it to change direction and lose some energy.

The *convective zone* is the outermost layer of the solar interior. It is a thick layer approximately 200,000 kilometers deep that transports energy from the edge of the radiative zone to the surface through giant convection cells, similar to a pot of boiling oatmeal. The plasma at the bottom of the convective zone is extremely hot, and it bubbles to the surface where it loses its heat to space. Once the plasma cools, it sinks back to the bottom of the convective zone.

Now that we have given a quick overview of the structure of the whole Sun, in this section, we will embark on a journey through the visible layers of the Sun, beginning with the photosphere—the visible surface.

The Solar Photosphere

Earth's air is generally transparent. But on a smoggy day in many cities, it can become opaque, which prevents us from seeing through it past a certain point. Something similar happens in the Sun. Its outer atmosphere is transparent, allowing us to look a short distance through it. But when we try to look through the atmosphere deeper into the Sun, our view is blocked. The photosphere is the layer where the Sun becomes opaque and marks the boundary past which we cannot see (Figure 15.5).

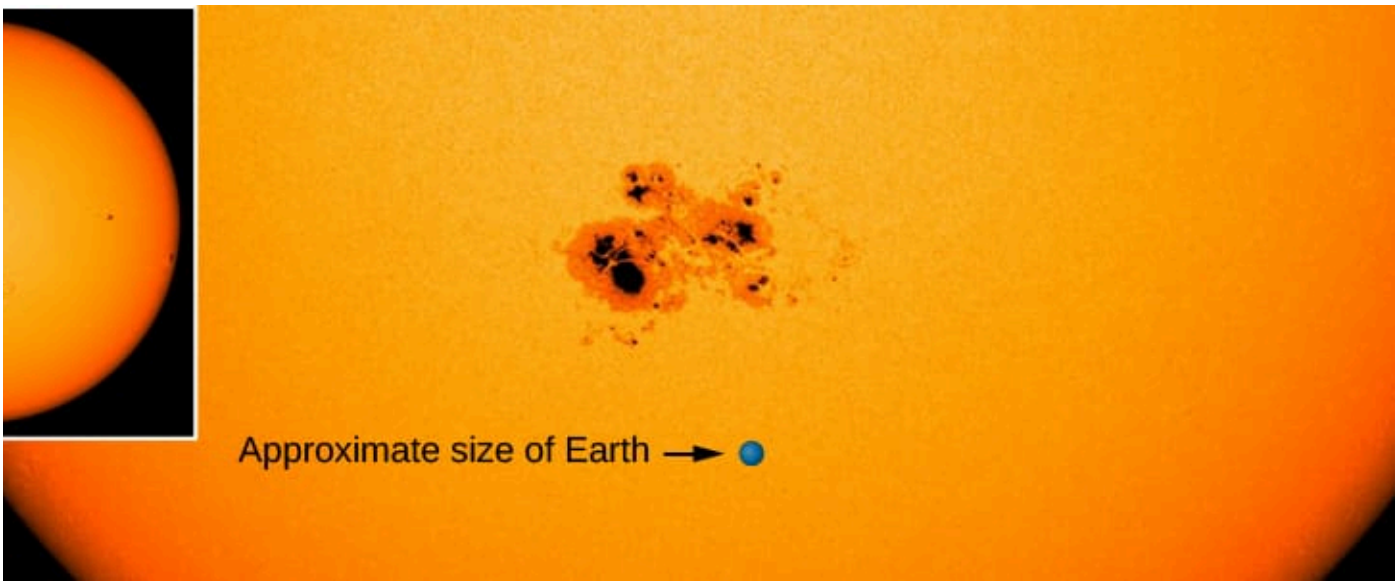


Figure 15.5 : Solar Photosphere plus Sunspots. This photograph shows the photosphere—the visible surface of the Sun. Also shown is an enlarged image of a group of sunspots; the size of Earth is shown for comparison. Sunspots appear darker because they are cooler than their surroundings. The typical temperature at the center of a large sunspot is about 3800 K, whereas the photosphere has a temperature of about 5800 K. (credit: modification of work by NASA/SDO)

As we saw, the energy that emerges from the photosphere was originally generated deep inside the Sun (more on this in The Sun: A Nuclear Powerhouse). This energy is in the form of photons, which make their way slowly toward the solar surface. Outside the Sun, we can observe *only* those photons that are emitted into the solar photosphere, where the density of atoms is sufficiently low and the photons can finally escape from the Sun without colliding with another atom or ion.

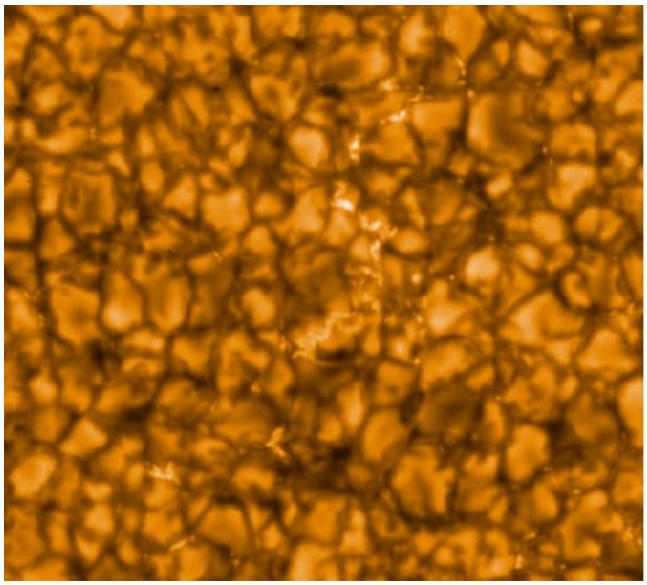
As an analogy, imagine that you are attending a big campus rally and have found a prime spot near the center of the action. Your friend arrives late and calls you on your cell phone to ask you to join her at the edge of the crowd. You decide that friendship is worth more than a prime spot, and so you work your way out through the dense crowd to meet her. You can move only a short distance before bumping into someone, changing direction, and trying again, making your way slowly to the outside edge of the crowd. All this while, your efforts are not visible to your waiting friend at the edge. Your friend can't see you until you get very close to the edge because of all the bodies in the way. So too photons making their way through the Sun are constantly bumping into atoms, changing direction, working their way slowly outward, and becoming visible only when they reach the atmosphere of the Sun where the density of atoms is too low to block their outward progress.

Astronomers have found that the solar atmosphere changes from almost perfectly transparent to almost completely opaque in a distance of just over 400 kilometers; it is this thin region that we call the *photosphere*, a word that comes from the Greek for "light sphere." When astronomers speak of the "diameter" of the Sun, they mean the size of the region surrounded by the photosphere.

The photosphere looks sharp only from a distance. If you were falling into the Sun, you would not feel any surface but would just sense a gradual increase in the density of the gas surrounding you. It is much the same as falling through a cloud while skydiving. From far away, the cloud looks as if it has a sharp surface, but you do not feel a surface as you fall into it. (One big difference between these two scenarios, however, is temperature. The Sun is so hot that you would be vaporized long before you reached the photosphere. Skydiving in Earth's atmosphere is much safer.)

We might note that the atmosphere of the Sun is not a very dense layer compared to the air in the room where you are reading this text. At a typical point in the photosphere, the pressure is less than 10% of Earth's pressure at sea level, and the density is about one ten-thousandth of Earth's atmospheric density at sea level.

Observations with telescopes show that the photosphere has a mottled appearance, resembling grains of rice spilled on a dark tablecloth or a pot of boiling oatmeal. This structure of the photosphere is called granulation (see Figure 15.6). Granules, which are typically 700 to 1000 kilometers in diameter (about the width of Texas), appear as bright areas surrounded by narrow, darker (cooler) regions. The lifetime of an individual granule is only 5 to 10 minutes. Even larger are supergranules, which are about 35,000 kilometers across (about the size of two Earths) and last about 24 hours.



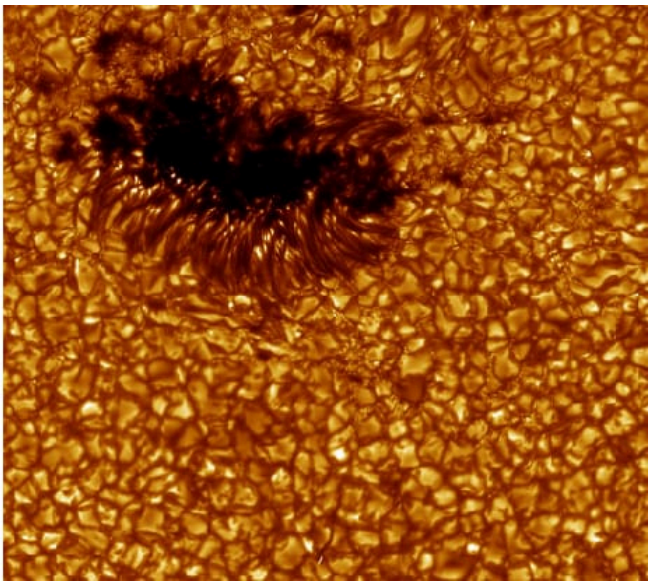


Figure 15.6 : Granulation Pattern. The surface markings of the convection cells create a granulation pattern on this dramatic image (left) taken from the Japanese Hinode spacecraft. You can see the same pattern when you heat up miso soup. The right image shows an irregular-shaped sunspot and granules on the Sun's surface, seen with the Swedish Solar Telescope on August 22, 2003. (credit left: modification of work by Hinode JAXA/NASA/PPARC; credit right: ISP/SST/Oddbjorn Engvold, Jun Elin Wiik, Luc Rouppe van der Voort)

The motions of the granules can be studied by examining the Doppler shifts in the spectra of gases just above them (see The Doppler Effect). The bright granules are columns of hotter gases rising at speeds of 2 to 3 kilometers per second from below the photosphere. As this rising gas reaches the photosphere, it spreads out, cools, and sinks down again into the darker regions between the granules. Measurements show that the centers of the granules are hotter than the intergranular regions by 50 to 100 K.

Link to Learning

See the "boiling" action of granulation in this 30-second time-lapse video from the Swedish Institute for Solar Physics.

The Chromosphere

The Sun's outer gases extend far beyond the photosphere (Figure 15.7). Because they are transparent to most visible radiation and emit only a small amount of light, these outer layers are difficult to observe. The region of the Sun's atmosphere that lies immediately above the photosphere is called the chromosphere. Until this century, the chromosphere was visible only when the photosphere was concealed by the Moon during a total solar eclipse (see the chapter on Earth, Moon, and Sky). In the seventeenth century, several observers described what appeared to them as a narrow red "streak" or "fringe" around the edge of the Moon during a brief instant after the Sun's photosphere had been covered. The name *chromosphere*, from the Greek for "colored sphere," was given to this red streak.

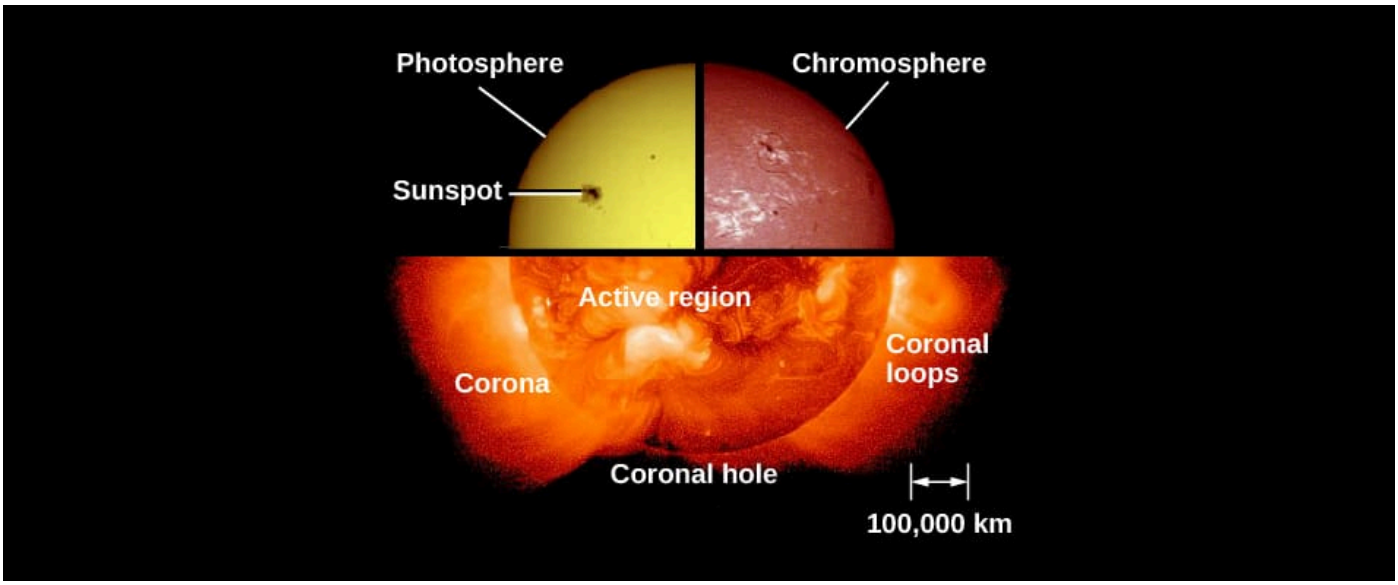


Figure 15.7 : The Sun's Atmosphere. Composite image showing the three components of the solar atmosphere: the photosphere or surface of the Sun taken in ordinary light; the chromosphere, imaged in the light of the strong red spectral line of hydrogen (H-alpha); and the corona as seen with X-rays. (credit: modification of work by NASA)

Observations made during eclipses show that the chromosphere is about 2000 to 3000 kilometers thick, and its spectrum consists of bright emission lines, indicating that this layer is composed of hot gases emitting light at discrete wavelengths. The reddish color of the chromosphere arises from one of the strongest emission lines in the visible part of its spectrum—the bright red line caused by hydrogen, the element that, as we have already seen, dominates the composition of the Sun.

In 1868, observations of the chromospheric spectrum revealed a yellow emission line that did not correspond to any previously known element on Earth. Scientists quickly realized they had found a new element and named it *helium* (after *helio*s, the Greek word for "Sun"). It took until 1895 for helium to be discovered on our planet. Today, students are probably most familiar with it as the light gas used to inflate balloons, although it turns out to be the second-most abundant element in the universe.

The temperature of the chromosphere is about 10,000 K. This means that the chromosphere is hotter than the photosphere, which should seem surprising. In all the situations we are familiar with, temperatures fall as one moves away from the source of heat, and the chromosphere is farther from the center of the Sun than the photosphere is.

The Transition Region

The increase in temperature does not stop with the chromosphere. Above it is a region in the solar atmosphere where the temperature changes from 10,000 K (typical of the chromosphere) to nearly a million degrees. The hottest part of the solar atmosphere, which has a temperature of a million degrees or more, is called the corona. Appropriately, the part of the Sun where the rapid temperature rise occurs is called the transition region. It is probably only a few tens of kilometers thick. Figure 15.8 summarizes how the temperature of the solar atmosphere changes from the photosphere outward.

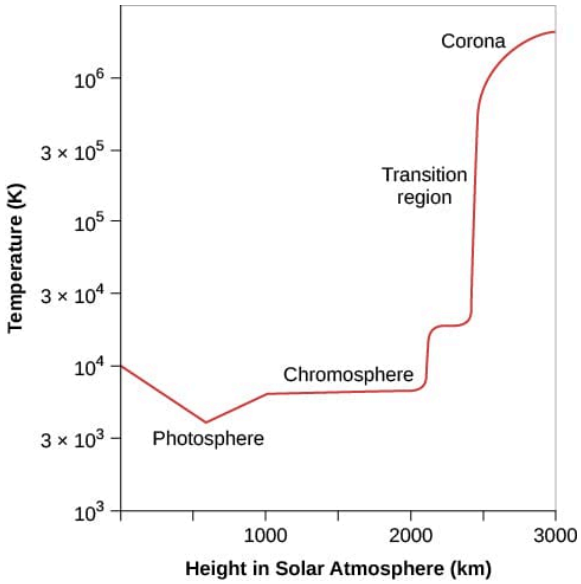


Figure 15.8 : Temperatures in the Solar Atmosphere. On this graph, temperature is shown increasing upward, and height above the photosphere is shown increasing to the right. Note the very rapid increase in temperature over a very short distance in the transition region between the chromosphere and the corona.

In 2013, NASA launched the Interface Region Imaging Spectrograph (IRIS) to study the transition region to understand better how and why this sharp temperature increase occurs. IRIS is the first space mission that is able to obtain high spatial resolution images of the different features produced over this wide temperature range and to see how they change with time and location (Figure 15.9).

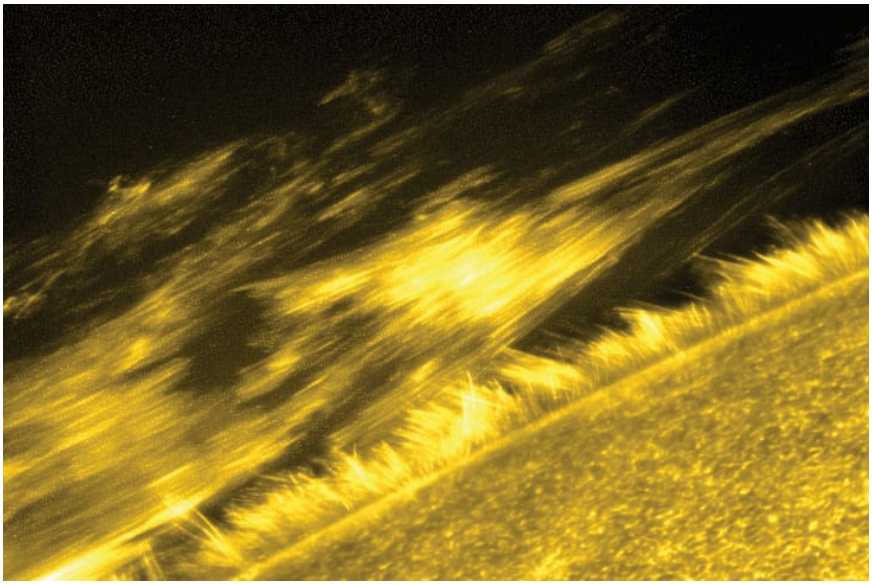


Figure 15.9 : Portion of the Transition Region. This image shows a giant ribbon of relatively cool gas threading through the lower portion of the hot corona. This ribbon (the technical term is filament) is made up of many individual threads. Time-lapse movies of this filament showed that it gradually heated as it moved through the corona. Scientists study events like this in order to try to understand what heats the chromosphere and corona to high temperatures. The “whiskers” at the edge of the Sun are spicules, jets of gas that shoot material up from the Sun’s surface and disappear after only a few minutes. This single image gives a hint of just how complicated it is to construct a model of the all the different structures and heating mechanisms in the solar atmosphere. (credit: JAXA/NASA/Hinode)

Figure 15.4 and the red graph in Figure 15.8 make the Sun seem rather like an onion, with smooth spherical shells, each one with a different temperature. For a long time, astronomers did indeed think of the Sun this way. However, we now know that while this idea of layers—photosphere, chromosphere, transition region, corona—describes the big picture fairly well, the Sun’s atmosphere is really more complicated, with hot and cool regions intermixed. For example, clouds of carbon monoxide gas with temperatures colder than 4000 K have now been found at the same height above the photosphere as the much hotter gas of the chromosphere.

The Corona

The outermost part of the Sun's atmosphere is called the *corona*. Like the chromosphere, the corona was first observed during total eclipses (Figure 15.10). Unlike the chromosphere, the corona has been known for many centuries: it was referred to by the Roman historian Plutarch and was discussed in some detail by Kepler.

The corona extends millions of kilometers above the photosphere and emits about half as much light as the full moon. The reason we don't see this light until an eclipse occurs is the overpowering brilliance of the photosphere. Just as bright city lights make it difficult to see faint starlight, so too does the intense light from the photosphere hide the faint light from the corona. While the best time to see the corona from Earth is during a total solar eclipse, it can be observed easily from orbiting spacecraft. Its brighter parts can now be photographed with a special instrument—a coronagraph—that removes the Sun's glare from the image with an occulting disk (a circular piece of material held so it is just in front of the Sun).

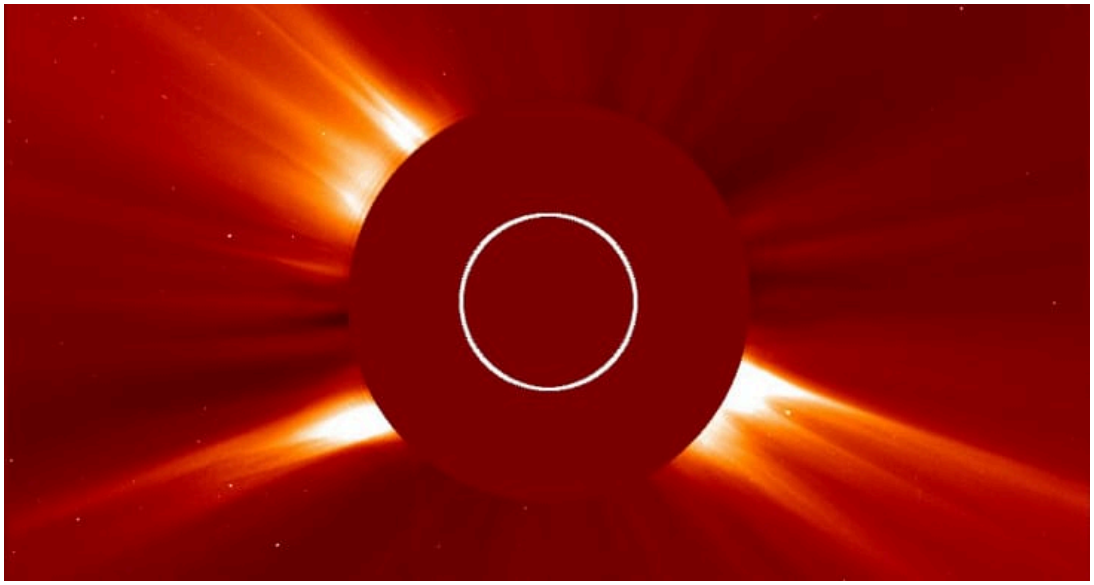


Figure 15.10 : Coronagraph. This image of the Sun was taken March 2, 2016. The larger dark circle in the center is the disk the blocks the Sun's glare, allowing us to see the corona. The smaller inner circle is where the Sun would be if it were visible in this image. (credit: modification of work by NASA/SOHO)

Studies of its spectrum show the corona to be very low in density. At the bottom of the corona, there are only about 10^9 atoms per cubic centimeter, compared with about 10^{16} atoms per cubic centimeter in the upper photosphere and 10^{19} molecules per cubic centimeter at sea level in Earth's atmosphere. The corona thins out very rapidly at greater heights, where it corresponds to a high vacuum by Earth laboratory standards. The corona extends so far into space—far past Earth—that here on our planet, we are technically living in the Sun's atmosphere.

The Solar Wind

One of the most remarkable discoveries about the Sun's atmosphere is that it produces a stream of charged particles (mainly protons and electrons) that we call the solar wind. These particles flow outward from the Sun into the solar system at a speed of about 400 kilometers per second (almost 1 million miles per hour)! The solar wind exists because the gases in the corona are so hot and moving so rapidly that they cannot be held back by solar gravity. (This wind was actually discovered by its effects on the charged tails of comets; in a sense, we can see the comet tails blow in the solar breeze the way wind socks at an airport or curtains in an open window flutter on Earth.)

Although the solar wind material is very, very rarified (i.e., *extremely* low density), the Sun has an enormous surface area. Astronomers estimate that the Sun is losing about 1–2 million tons of material each second through this wind. Although this sounds like a lot, it's so trivial compared to the enormous mass of the Sun that it can be neglected as we study the Sun.

From where in the Sun does the solar wind emerge? In visible photographs, the solar corona appears fairly uniform and smooth. X-ray and extreme ultraviolet pictures, however, show that the corona has loops, plumes, and both bright and dark regions. Large

dark regions of the corona that are relatively cool and quiet are called coronal holes (Figure 15.11). In these regions, magnetic field lines stretch far out into space away from the Sun, rather than looping back to the surface. The solar wind comes predominantly from coronal holes, where gas can stream away from the Sun into space unhindered by magnetic fields. Hot coronal gas, on the other hand, is present mainly where magnetic fields have trapped and concentrated it.

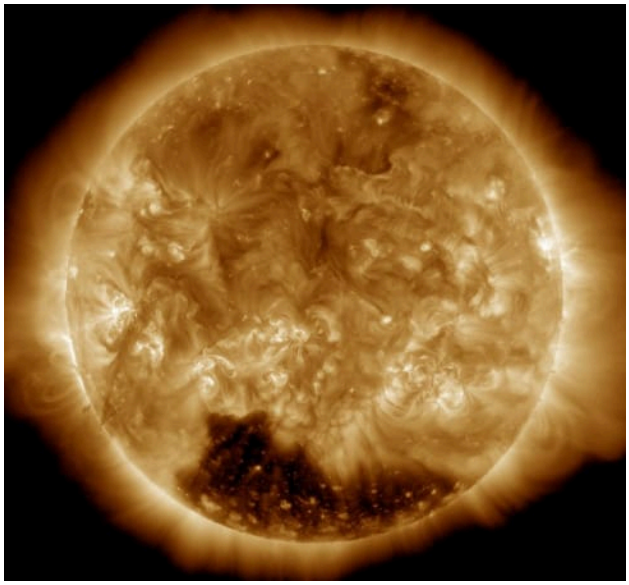


Figure 15.11 Coronal Hole. The dark area visible near the Sun's south pole on this Solar Dynamics Observer spacecraft image is a coronal hole. (credit: modification of work by NASA/SDO)

At the surface of Earth, we are protected to some degree from the solar wind by our atmosphere and Earth's magnetic field (see Earth as a Planet). However, the magnetic field lines come into Earth at the north and south magnetic poles. Here, charged particles accelerated by the solar wind can follow the field down into our atmosphere. As the particles strike molecules of air, they cause them to glow, producing beautiful curtains of light called the auroras, or the northern and southern lights (Figure 15.12).



Figure 15.12 : Aurora. The colorful glow in the sky results from charged particles in a solar wind interacting with Earth's magnetic fields. The stunning display captured here occurred over Jokulsarlon Lake in Iceland in 2013. (credit: Moyan Brenn)

Link to Learning

This NASA video explains and demonstrates the nature of the auroras and their relationship to Earth's magnetic field.

This page titled 11.3: The Structure and Composition of the Sun is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 15.2: The Structure and Composition of the Sun by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.4: The Solar Interior - Theory

Learning Objectives

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

- Describe the state of equilibrium of the Sun
- Understand the energy balance of the Sun
- Explain how energy moves outward through the Sun
- Describe the structure of the solar interior

Fusion of protons can occur in the center of the Sun only if the temperature exceeds 12 million K. How do we know that the Sun is actually this hot? To determine what the interior of the Sun might be like, it is necessary to resort to complex calculations. Since we can't see the interior of the Sun, we have to use our understanding of physics, combined with what we see at the surface, to construct a mathematical model of what must be happening in the interior. Astronomers use observations to build a computer program containing everything they think they know about the physical processes going on in the Sun's interior. The computer then calculates the temperature and pressure at every point inside the Sun and determines what nuclear reactions, if any, are taking place. For some calculations, we can use observations to determine whether the computer program is producing results that match what we see. In this way, the program evolves with ever-improving observations.

The computer program can also calculate how the Sun will change with time. After all, the Sun must change. In its center, the Sun is slowly depleting its supply of hydrogen and creating helium instead. Will the Sun get hotter? Cooler? Larger? Smaller? Brighter? Fainter? Ultimately, the changes in the center could be catastrophic, since eventually all the hydrogen fuel hot enough for fusion will be exhausted. Either a new source of energy must be found, or the Sun will cease to shine. We will describe the ultimate fate of the Sun in later chapters. For now, let's look at some of the things we must teach the computer about the Sun in order to carry out such calculations.

The Sun Is a Plasma

The Sun is so hot that all of the material in it is in the form of an ionized gas, called a plasma. Plasma acts much like a hot gas, which is easier to describe mathematically than either liquids or solids. The particles that constitute a gas are in rapid motion, frequently colliding with one another. This constant bombardment is the *pressure* of the gas (Figure 16.10).

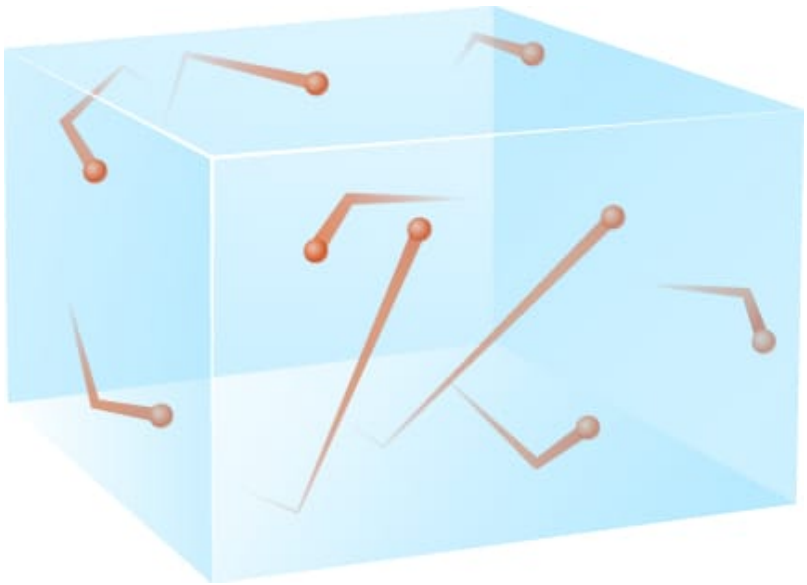


Figure 16.10 : Gas Pressure. The particles in a gas are in rapid motion and produce pressure through collisions with the surrounding material. Here, particles are shown bombarding the sides of an imaginary container.

More particles within a given volume of gas produce more pressure because the combined impact of the moving particles increases with their number. The pressure is also greater when the molecules or atoms are moving faster. Since the molecules move faster when the temperature is hotter, higher temperatures produce higher pressure.

The Sun Is Stable

The Sun, like the majority of other stars, is stable; it is neither expanding nor contracting. Such a star is said to be in a condition of *equilibrium*. All the forces within it are balanced, so that at each point within the star, the temperature, pressure, density, and so on are maintained at constant values. We will see in later chapters that even these stable stars, including the Sun, are changing as they evolve, but such evolutionary changes are so gradual that, for all intents and purposes, the stars are still in a state of equilibrium at any given time.

The mutual gravitational attraction between the masses of various regions within the Sun produces tremendous forces that tend to collapse the Sun toward its center. Yet we know from the history of Earth that the Sun has been emitting roughly the same amount of energy for billions of years, so clearly it has managed to resist collapse for a very long time. The gravitational forces must therefore be counterbalanced by some other force. That force is due to the pressure of gases within the Sun (Figure 16.11). Calculations show that, in order to exert enough pressure to prevent the Sun from collapsing due to the force of gravity, the gases at its center must be maintained at a temperature of 15 million K. Think about what this tells us. Just from the fact that the Sun is not contracting, we can conclude that its temperature must indeed be high enough at the center for protons to undergo fusion.

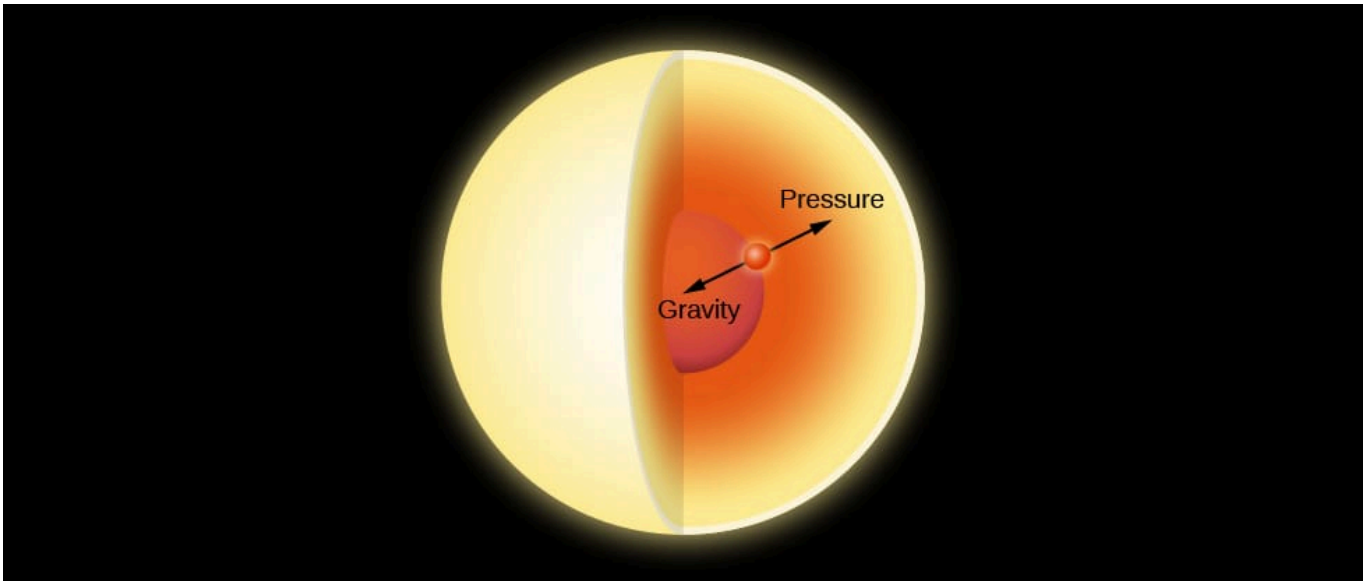


Figure 16.11 : Hydrostatic Equilibrium. In the interior of a star, the inward force of gravity is exactly balanced at each point by the outward force of gas pressure.

The Sun maintains its stability in the following way. If the internal pressure in such a star were not great enough to balance the weight of its outer parts, the star would collapse somewhat, contracting and building up the pressure inside. On the other hand, if the pressure were greater than the weight of the overlying layers, the star would expand, thus decreasing the internal pressure. Expansion would stop, and equilibrium would again be reached when the pressure at every internal point equaled the weight of the stellar layers above that point. An analogy is an inflated balloon, which will expand or contract until an equilibrium is reached between the pressure of the air inside and outside. The technical term for this condition is hydrostatic equilibrium. Stable stars are all in hydrostatic equilibrium; so are the oceans of Earth as well as Earth's atmosphere. The air's own pressure keeps it from falling to the ground.

The Sun Is Not Cooling Down

As everyone who has ever left a window open on a cold winter night knows, heat always flows from hotter to cooler regions. As energy filters outward toward the surface of a star, it must be flowing from inner, hotter regions. The temperature cannot ordinarily get cooler as we go inward in a star, or energy would flow in and heat up those regions until they were at least as hot as the outer ones. Scientists conclude that the temperature is highest at the center of a star, dropping to lower and lower values toward

the stellar surface. (The high temperature of the Sun's chromosphere and corona may therefore appear to be a paradox. But remember from The Sun: A Garden-Variety Star that these high temperatures are maintained by magnetic effects, which occur in the Sun's atmosphere.)

The outward flow of energy through a star robs it of its internal heat, and the star would cool down if that energy were not replaced. Similarly, a hot iron begins to cool as soon as it is unplugged from its source of electric energy. Therefore, a source of fresh energy must exist within each star. In the Sun's case, we have seen that this energy source is the ongoing fusion of hydrogen to form helium.

Heat Transfer in a Star

Since the nuclear reactions that generate the Sun's energy occur deep within it, the energy must be transported from the center of the Sun to its surface—where we see it in the form of both heat and light. There are three ways in which energy can be transferred from one place to another. In conduction, atoms or molecules pass on their energy by colliding with others nearby. This happens, for example, when the handle of a metal spoon heats up as you stir a cup of hot coffee. In convection, currents of warm material rise, carrying their energy with them to cooler layers. A good example is hot air rising from a fireplace. In radiation, energetic photons move away from hot material and are absorbed by some material to which they convey some or all of their energy. You can feel this when you put your hand close to the coils of an electric heater, allowing infrared photons to heat up your hand. Conduction and convection are both important in the interiors of planets. In stars, which are much more transparent, radiation and convection are important, whereas conduction can usually be ignored.

Stellar *convection* occurs as currents of hot gas flow up and down through the star (Figure 16.12). Such currents travel at moderate speeds and do not upset the overall stability of the star. They don't even result in a net transfer of mass either inward or outward because, as hot material rises, cool material falls and replaces it. This results in a convective circulation of rising and falling cells as seen in Figure 16.12. In much the same way, heat from a fireplace can stir up air currents in a room, some rising and some falling, without driving any air into or out the room. Convection currents carry heat very efficiently outward through a star. In the Sun, convection turns out to be important in the central regions and near the surface.

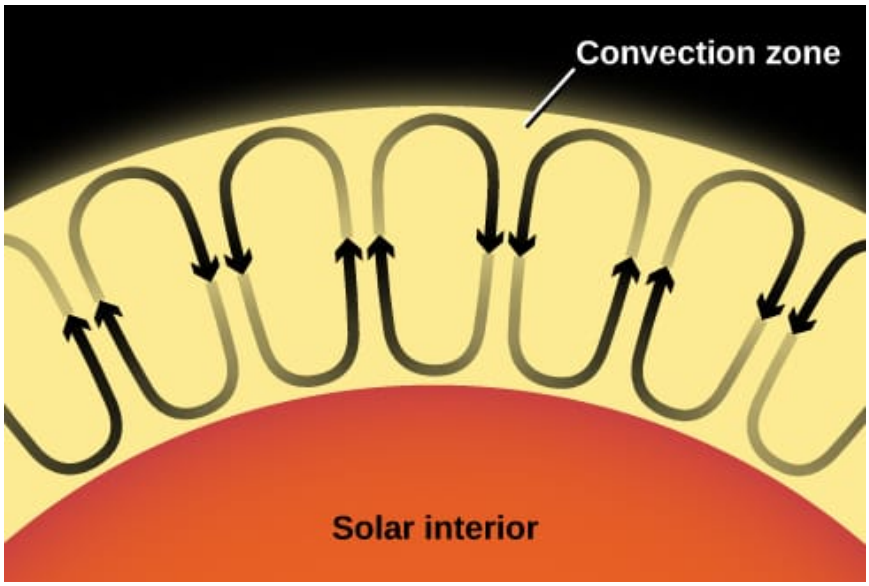


Figure 16.12 : Convection. Rising convection currents carry heat from the Sun's interior to its surface, whereas cooler material sinks downward. Of course, nothing in a real star is as simple as diagrams in textbooks suggest.

Unless convection occurs, the only significant mode of energy transport through a star is by electromagnetic radiation. Radiation is not an efficient means of energy transport in stars because gases in stellar interiors are very opaque, that is, a photon does not go far (in the Sun, typically about 0.01 meter) before it is absorbed. (The processes by which atoms and ions can interrupt the outward flow of photons—such as becoming ionized—were discussed in the section on the Formation of Spectral Lines.) The absorbed energy is always reemitted, but it can be reemitted in any direction. A photon absorbed when traveling outward in a star has almost as good a chance of being radiated back toward the center of the star as toward its surface.

A particular quantity of energy, therefore, zigzags around in an almost random manner and takes a long time to work its way from

the center of a star to its surface (Figure 16.13). Estimates are somewhat uncertain, but in the Sun, as we saw, the time required is probably between 100,000 and 1,000,000 years. If the photons were not absorbed and reemitted along the way, they would travel at the speed of light and could reach the surface in a little over 2 seconds, just as neutrinos do (Figure 16.14).

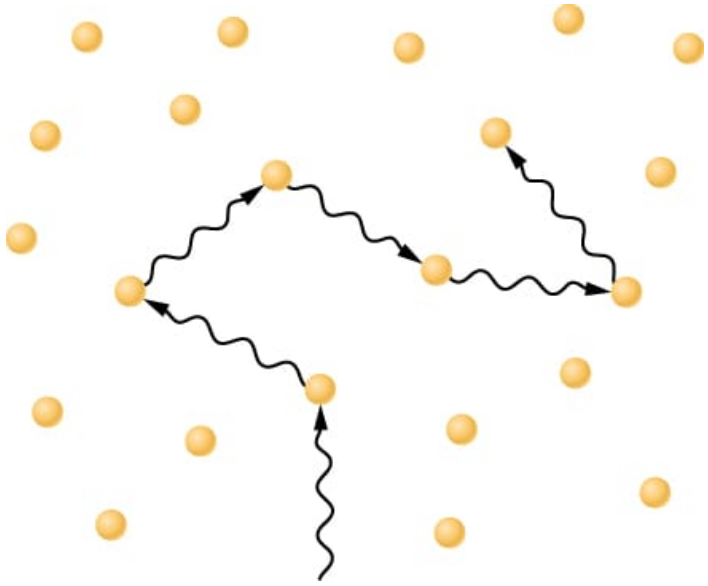


Figure 16.13 : Photons Deep in the Sun. A photon moving through the dense gases in the solar interior travels only a short distance before it interacts with one of the surrounding atoms. The resulting photon usually has a lower energy after each interaction and may then travel in any random direction.

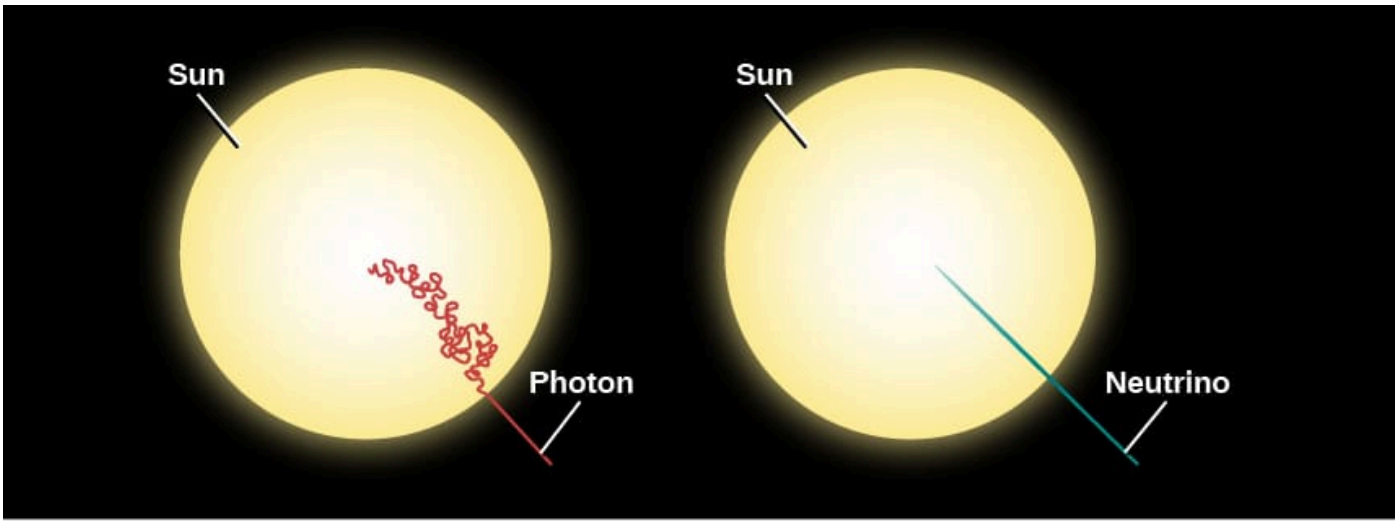


Figure 16.14 : Photon and Neutrino Paths in the Sun. (a) Because photons generated by fusion reactions in the solar interior travel only a short distance before being absorbed or scattered by atoms and sent off in random directions, estimates are that it takes between 100,000 and 1,000,000 years for energy to make its way from the center of the Sun to its surface. (b) In contrast, neutrinos do not interact with matter but traverse straight through the Sun at the speed of light, reaching the surface in only a little more than 2 seconds.

Making Connections: Heat Transfer and Cooking

The three ways that heat energy moves from higher-temperature regions to cooler regions are all used in cooking, and this is important to all of us who enjoy making or eating food.

Conduction is heat transfer by physical contact during which the energetic motion of particles in one region spread to other regions and even to adjacent objects in close contact. A tasty example of this is cooking a steak on a hot iron skillet. When a flame makes the bottom of a skillet hot, the particles in it vibrate actively and collide with neighboring particles, spreading the heat energy throughout the skillet (the ability to spread heat uniformly is a key criterion for selecting materials for cookware). A steak sitting on the surface of the skillet picks up heat energy by the particles in the surface of the skillet colliding with particles on the surface of the steak. Many cooks will put a little oil on the pan, and this layer of oil, besides preventing sticking, increases heat transfer by filling in gaps and increasing the contact surface area.

Convection is heat transfer by the motion of matter that rises because it is hot and less dense. Heating a fluid makes it expand, which makes it less dense, so it rises. An oven is a great example of this: the fire is at the bottom of the oven and heats the air down there, causing it to expand (becoming less dense), so it rises up to where the food is. The rising hot air carries the heat from the fire to the food by convection. This is how conventional ovens work. You may also be familiar with convection ovens that use a fan to circulate hot air for more even cooking. A scientist would object to that name because normal non-fan ovens that rely on hot air rising to circulate the heat are convection ovens; technically, the ovens that use fans to help move heat are “advection” ovens. (You may not have heard about this because the scientists who complain loudly about misusing the terms convection and advection don’t get out much.)

Radiation is the transfer of heat energy by electromagnetic radiation. Although microwave ovens are an obvious example of using radiation to heat food, a simpler example is a toy oven. Toy ovens are powered by a very bright light bulb. The child-chefs prepare a mix for brownies or cookies, put it into a tray, and place it in the toy oven under the bright light bulb. The light and heat from the bulb hit the brownie mix and cook it. If you have ever put your hand near a bright light, you have undoubtedly noticed your hand getting warmed by the light.

Model Stars

Scientists use the principles we have just described to calculate what the Sun’s interior is like. These physical ideas are expressed as mathematical equations that are solved to determine the values of temperature, pressure, density, the efficiency with which photons are absorbed, and other physical quantities throughout the Sun. The solutions obtained, based on a specific set of physical assumptions, provide a theoretical model for the interior of the Sun.

Figure 16.15 schematically illustrates the predictions of a theoretical model for the Sun’s interior. Energy is generated through fusion in the core of the Sun, which extends only about one-quarter of the way to the surface but contains about one-third of the total mass of the Sun. At the center, the temperature reaches a maximum of approximately 15 million K, and the density is nearly 150 times that of water. The energy generated in the core is transported toward the surface by radiation until it reaches a point about 70% of the distance from the center to the surface. At this point, convection begins, and energy is transported the rest of the way, primarily by rising columns of hot gas.

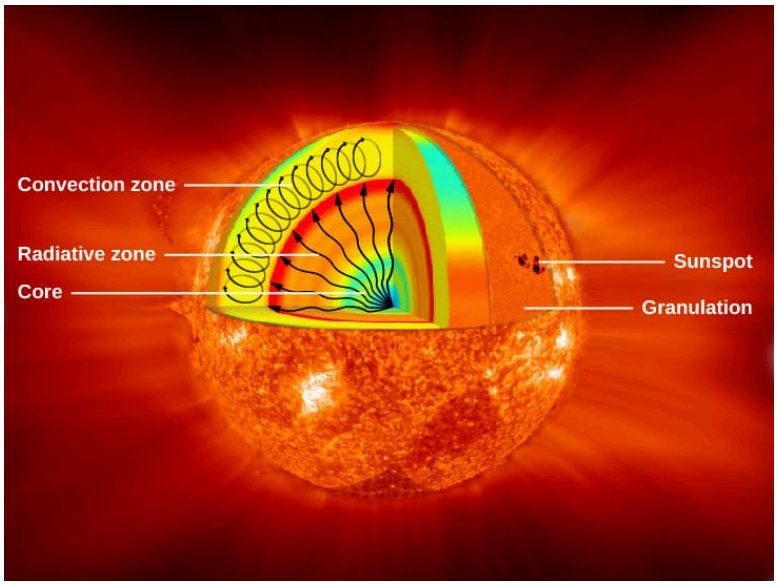


Figure 16.15 : Interior Structure of the Sun. Energy is generated in the core by the fusion of hydrogen to form helium. This energy is transmitted outward by radiation—that is, by the absorption and reemission of photons. In the outermost layers, energy is transported mainly by convection. (credit: modification of work by NASA/Goddard)

Figure 16.16 shows how the temperature, density, rate of energy generation, and composition vary from the center of the Sun to its surface.

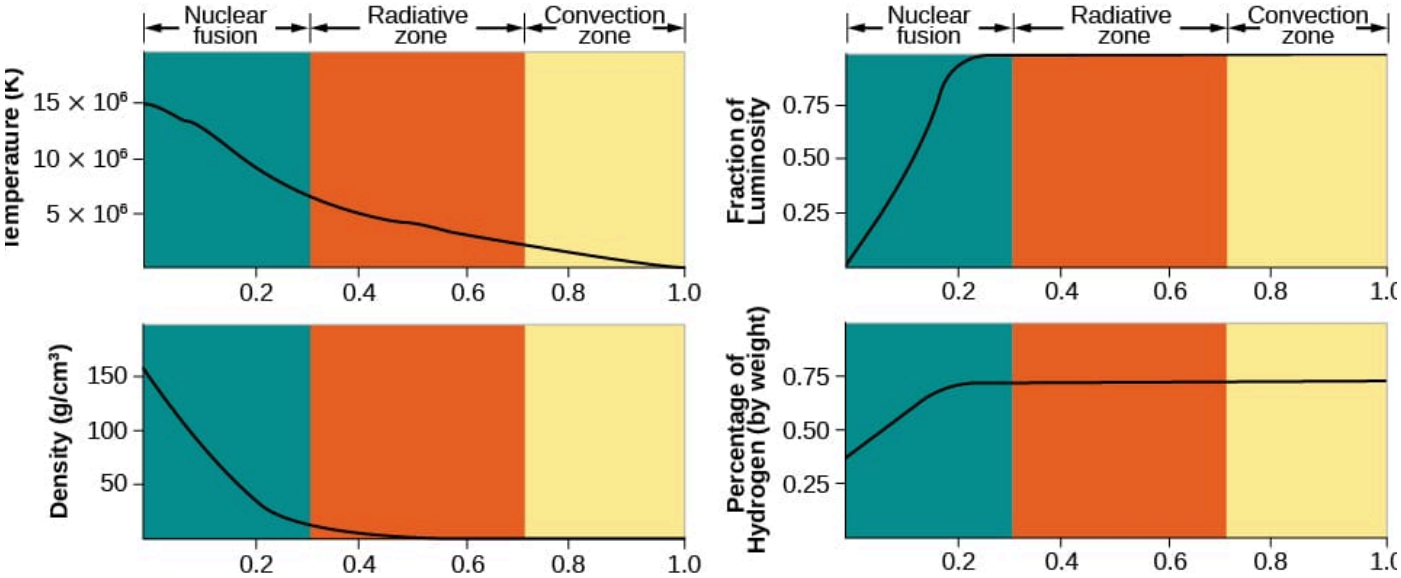


Figure 16.16 : Interior of the Sun. Diagrams showing how temperature, density, rate of energy generation, and the percentage (by mass) abundance of hydrogen vary inside the Sun. The horizontal scale shows the fraction of the Sun's radius: the left edge is the very center, and the right edge is the visible surface of the Sun, which is called the photosphere.

This page titled 11.4: The Solar Interior - Theory is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.4: The Solar Interior - Theory by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.5: The Solar Interior - Observations

Learning Objectives

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

- Explain how the Sun pulsates
- Explain what helioseismology is and what it can tell us about the solar interior
- Discuss how studying neutrinos from the Sun has helped understand neutrinos

Recall that when we observe the Sun's photosphere (the surface layer we see from the outside), we are not seeing very deeply into our star, certainly not into the regions where energy is generated. That's why the title of this section—observations of the solar interior—should seem very surprising. However, astronomers have indeed devised two types of measurements that can be used to obtain information about the inner parts of the Sun. One technique involves the analysis of tiny changes in the motion of small regions at the Sun's surface. The other relies on the measurement of the neutrinos emitted by the Sun.

Solar Pulsations

Astronomers discovered that the Sun pulsates—that is, it alternately expands and contracts—just as your chest expands and contracts as you breathe. This pulsation is very slight, but it can be detected by measuring the *radial velocity* of the solar surface—the speed with which it moves toward or away from us. The velocities of small regions on the Sun are observed to change in a regular way, first toward Earth, then away, then toward, and so on. It is as if the Sun were "breathing" through thousands of individual lungs, each having a size in the range of 4000 to 15,000 kilometers, each fluctuating back and forth (Figure 16.17).

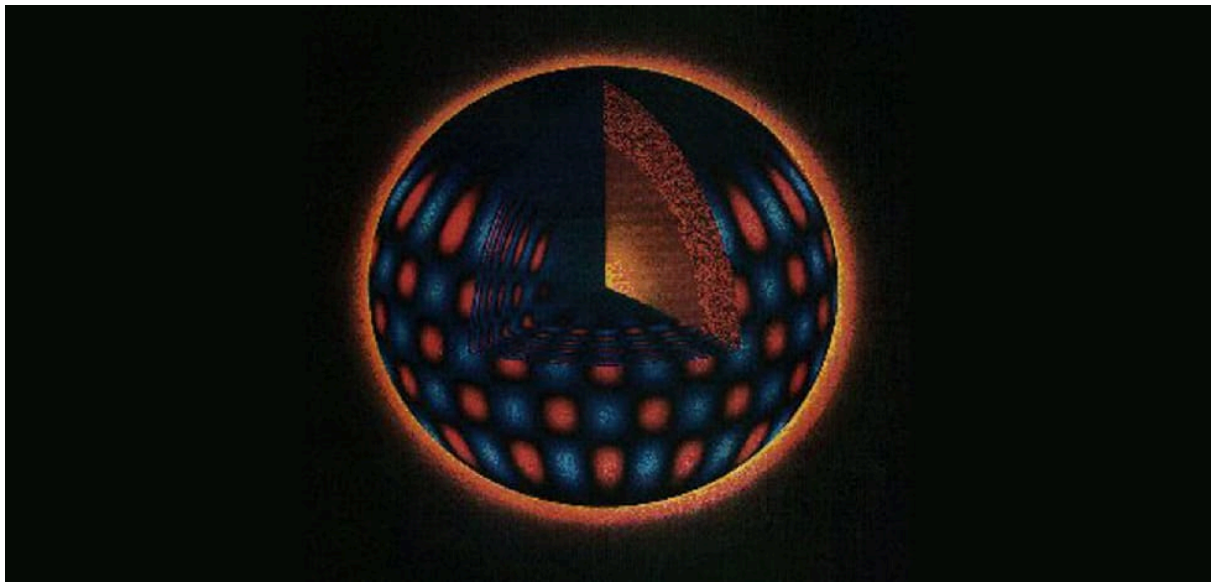


Figure 16.17 : Oscillations in the Sun. New observational techniques permit astronomers to measure small differences in velocity at the Sun's surface to infer what the deep solar interior is like. In this computer simulation, red shows surface regions that are moving away from the observer (inward motion); blue marks regions moving toward the observer (outward motion). Note that the velocity changes penetrate deep into the Sun's interior. (credit: modification of work by GONG, NOAO)

The typical velocity of one of the oscillating regions on the Sun is only a few hundred meters per second, and it takes about 5 minutes to complete a full cycle from maximum to minimum velocity and back again. The change in the size of the Sun measured at any given point is no more than a few kilometers.

The remarkable thing is that these small velocity variations can be used to determine what the interior of the Sun is like. The motion of the Sun's surface is caused by waves that reach it from deep in the interior. Study of the amplitude and cycle

length of velocity changes provides information about the temperature, density, and composition of the layers through which the waves passed before they reached the surface. The situation is somewhat analogous to the use of seismic waves generated by earthquakes to infer the properties of Earth's interior. For this reason, studies of solar oscillations (back-and-forth motions) are referred to as helioseismology.

It takes a little over an hour for waves to traverse the Sun from center to surface, so the waves, like neutrinos, provide information about what the solar interior is like at the present time. In contrast, remember that the sunlight we see today emerging from the Sun was actually generated in the core several hundred thousand years ago.

Helioseismology has shown that convection extends inward from the surface 30% of the way toward the center; we have used this information in drawing Figure 16.15. Pulsation measurements also show that the *differential rotation* that we see at the Sun's surface, with the fastest rotation occurring at the equator, persists down through the convection zone. Below the convection zone, however, the Sun, even though it is gaseous throughout, rotates as if it were a solid body like a bowling ball. Another finding from helioseismology is that the abundance of helium inside the Sun, except in the center where nuclear reactions have converted hydrogen into helium, is about the same as at its surface. That result is important to astronomers because it means we are correct when we use the abundance of the elements measured in the solar atmosphere to construct models of the solar interior.

Helioseismology also allows scientists to look beneath a sunspot and see how it works. In The Sun: A Garden-Variety Star, we said that sunspots are cool because strong magnetic fields block the outward flow of energy. Figure 16.18 shows how gas moves around underneath a sunspot. Cool material from the sunspot flows downward, and material surrounding the sunspot is pulled inward, carrying magnetic field with it and thus maintaining the strong field that is necessary to form a sunspot. As the new material enters the sunspot region, it too cools, becomes denser, and sinks, thus setting up a self-perpetuating cycle that can last for weeks.

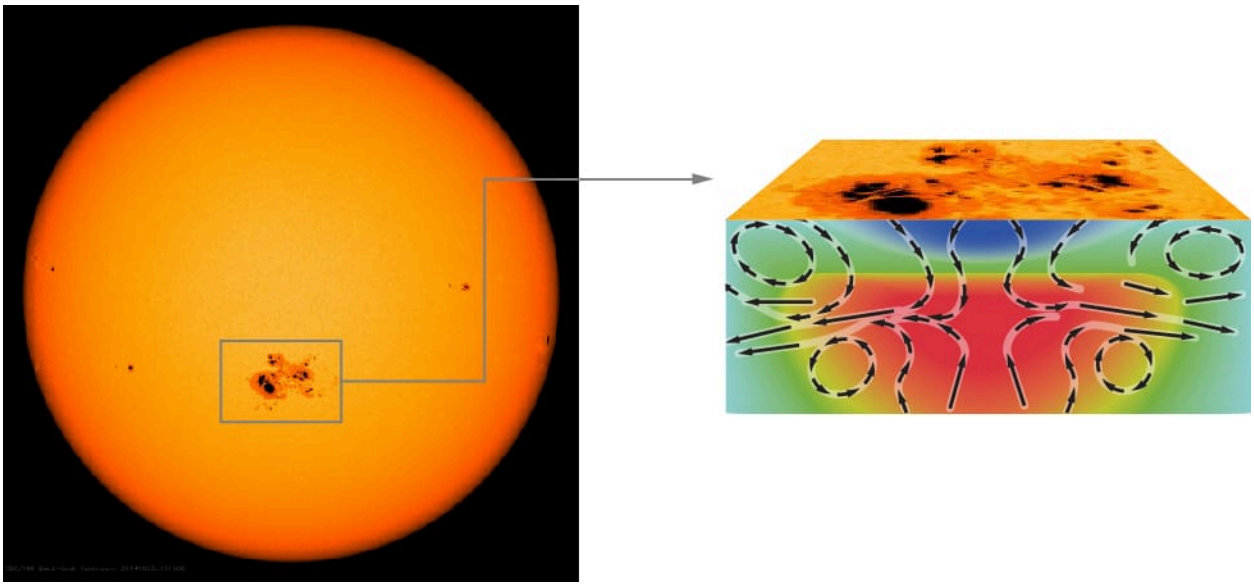


Figure 16.18 : Sunspot Structure. This drawing shows our new understanding, from helioseismology, of what lies beneath a sunspot. The black arrows show the direction of the flow of material. The intense magnetic field associated with the sunspot stops the upward flow of hot material and creates a kind of plug that blocks the hot gas. As the material above the plug cools (shown in blue), it becomes denser and plunges inward, drawing more gas and more magnetic field behind it into the spot. The concentrated magnetic field causes more cooling, thereby setting up a self-perpetuating cycle that allows a spot to survive for several weeks. Since the plug keeps hot material from flowing up into the sunspot, the region below the plug, represented by red in this picture, becomes hotter. This material flows sideways and then upward, eventually reaching the solar surface in the area surrounding the sunspot. (credit: modification of work by NASA, SDO)

The downward-flowing cool material acts as a kind of plug that block the upward flow of hot material, which is then diverted sideways and eventually reaches the solar surface in the region around the sunspot. This outward flow of hot material accounts for the paradox that we described in The Sun: A Garden-Variety Star—namely, that the Sun emits slightly more energy when more of its surface is covered by cool sunspots.

Helioseismology has become an important tool for predicting solar storms that might impact Earth. Active regions can appear and grow large in only a few days. The solar rotation period is about 28 days. Therefore, regions capable of producing solar flares and coronal mass ejections can develop on the far side of the Sun, where, for a long time, we couldn't see them directly.

Fortunately, we now have space telescopes monitoring the Sun from all angles, so we know if there are sunspots forming on the opposite side of the Sun. Moreover, sound waves travel slightly faster in regions of high magnetic field, and waves generated in active regions traverse the Sun about 6 seconds faster than waves generated in quiet regions. By detecting this subtle difference, scientists can provide warnings of a week or more to operators of electric utilities and satellites about when a potentially dangerous active region might rotate into view. With this warning, it is possible to plan for disruptions, put key instruments into safe mode, or reschedule spacewalks in order to protect astronauts.

Solar Neutrinos

The second technique for obtaining information about the Sun's interior involves the detection of a few of those elusive neutrinos created during nuclear fusion. Recall from our earlier discussion that neutrinos created in the center of the Sun make their way directly out of the Sun and travel to Earth at nearly the speed of light. As far as neutrinos are concerned, the Sun is transparent.

About 3% of the total energy generated by nuclear fusion in the Sun is carried away by neutrinos. So many protons react and form neutrinos inside the Sun's core that, scientists calculate, 35 million billion (3.5×10^{16}) solar neutrinos pass through each square meter of Earth's surface every second. If we can devise a way to detect even a few of these solar neutrinos, then we can obtain information directly about what is going on in the center of the Sun. Unfortunately for those trying to "catch" some neutrinos, Earth and everything on it are also nearly transparent to passing neutrinos, just like the Sun.

On very, very rare occasions, however, one of the billions and billions of solar neutrinos will interact with another atom. The first successful detection of solar neutrinos made use of cleaning fluid (C_2Cl_4), which is the least expensive way to get a lot of chlorine atoms together. The nucleus of a chlorine (Cl) atom in the cleaning fluid can be turned into a radioactive argon nucleus by an interaction with a neutrino. Because the argon is radioactive, its presence can be detected. However, since the interaction of a neutrino with chlorine happens so rarely, a huge amount of chlorine is needed.

Raymond Davis, Jr. (Figure 16.19) and his colleagues at Brookhaven National Laboratory, placed a tank containing nearly 400,000 liters of cleaning fluid 1.5 kilometers beneath Earth's surface in a gold mine at Lead, South Dakota. A mine was chosen so that the surrounding material of Earth would keep cosmic rays (high-energy particles from space) from reaching the cleaning fluid and creating false signals. (Cosmic-ray particles are stopped by thick layers of Earth, but neutrinos find them of no significance.) Calculations show that solar neutrinos should produce about one atom of radioactive argon in the tank each day.

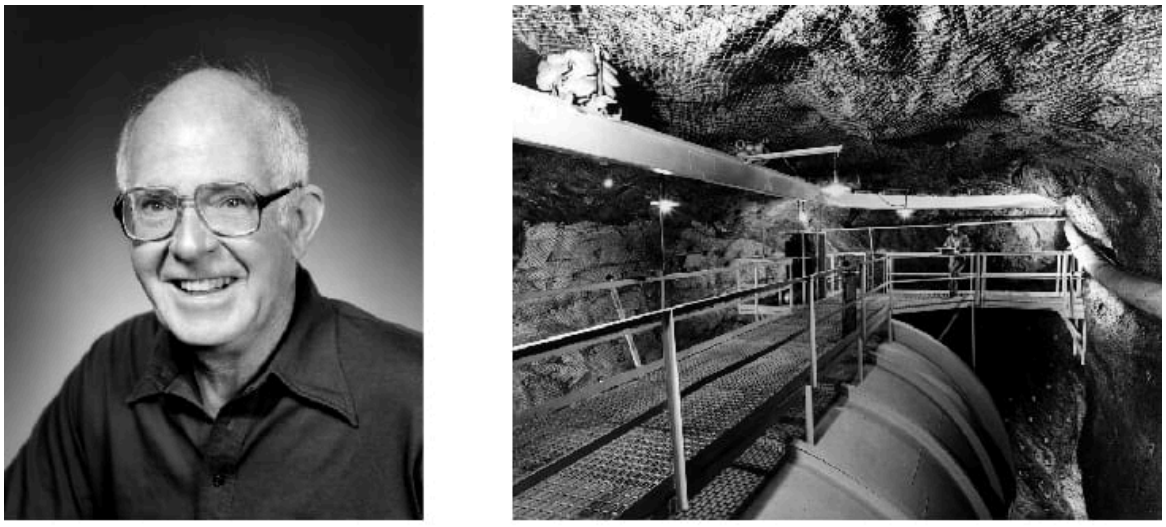


Figure 16.19 : Davis Experiment. (a) Raymond Davis received the Nobel Prize in physics in 2002. (b) Davis' experiment at the bottom of an abandoned gold mine first revealed problems with our understanding of neutrinos. (credit a: modification of work by Brookhaven National Laboratory; credit b: modification of work by the United States Department of Energy)

This was an amazing project: they counted argon atoms about once per month—and remember, they were looking for a tiny handful of argon atoms in a massive tank of chlorine atoms. When all was said and done, Davis' experiment, begun in 1970, detected only about one-third as many neutrinos as predicted by solar models! This was a shocking result because astronomers thought they had a pretty good understanding of both neutrinos and the Sun's interior. For many years, astronomers and physicists

wrestled with Davis' results, trying to find a way out of the dilemma of the "missing" neutrinos.

Eventually Davis' result was explained by the surprising discovery that there are actually three types of neutrinos. Solar fusion produces only one type of neutrino, the so-called electron neutrino, and the initial experiments to detect solar neutrinos were designed to detect this one type. Subsequent experiments showed that these neutrinos change to a different type during their journey from the center of the Sun through space to Earth in a process called *neutrino oscillation*.

An experiment, conducted at the Sudbury Neutrino Observatory in Canada, was the first one designed to capture all three types of neutrinos (Figure 16.20). The experiment was located in a mine 2 kilometers underground. The neutrino detector consisted of a 12-meter-diameter transparent acrylic plastic sphere, which contained 1000 metric tons of heavy water. Remember that an ordinary water nucleus contains two hydrogen atoms and one oxygen atom. Heavy water instead contains two deuterium atoms and one oxygen atom, and incoming neutrinos can occasionally break up the loosely bound proton and neutron that make up the deuterium nucleus. The sphere of heavy water was surrounded by a shield of 1700 metric tons of very pure water, which in turn was surrounded by 9600 photomultipliers, devices that detect flashes of light produced after neutrinos interact with the heavy water.

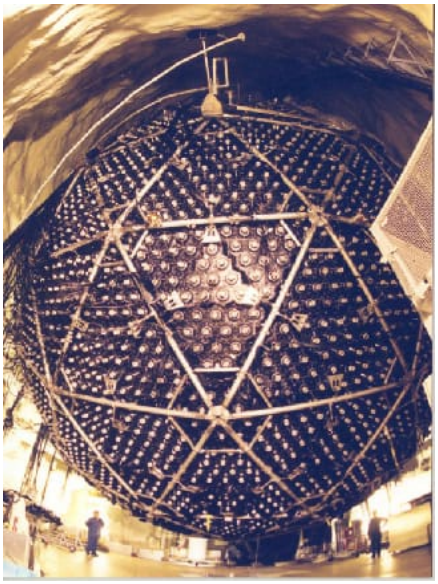


Figure 16.20 : Sudbury Neutrino Detector. The 12-meter sphere of the Sudbury Neutrino Detector lies more than 2 kilometers underground and holds 1000 metric tons of heavy water. (credit: A.B. McDonald (Queen's University) et al., The Sudbury Neutrino Observatory Institute)

To the enormous relief of astronomers who make models of the Sun, the Sudbury experiment detected about 1 neutrino per hour and has shown that the *total* number of neutrinos reaching the heavy water is just what solar models predict. Only one-third of these, however, are electron neutrinos. It appears that two-thirds of the electron neutrinos produced by the Sun transform themselves into one of the other types of neutrinos as they make their way from the core of the Sun to Earth. This is why the earlier experiments saw only one-third the number of neutrinos expected.

Although it is not intuitively obvious, such neutrino oscillations can happen only if the mass of the electron neutrino is not zero. Other experiments indicate that its mass is tiny (even compared to the electron). The 2015 Nobel Prize in physics was awarded to researchers Takaaki Kajita and Arthur B. McDonald for their work establishing the changeable nature of neutrinos. (Raymond Davis shared the 2002 Nobel Prize with Japan's Masatoshi Koshiba for the experiments that led to our understanding of the neutrino problem in the first place.) But the fact that the neutrino has mass at all has deep implications for both physics and astronomy. For example, we will look at the role that neutrinos play in the inventory of the mass of the universe in The Big Bang.

The Borexino experiment, an international experiment conducted in Italy, detected neutrinos coming from the Sun that were identified as coming from different reactions. Whereas the p-p chain is the reaction producing most of the Sun's energy, it is not the only nuclear reaction occurring in the Sun's core. There are side reactions involving nuclei of such elements as beryllium and boron. By probing the number of neutrinos that come from each reaction, the Borexino experiment has helped us confirm in detail our understanding of nuclear fusion in the Sun. In 2014, the Borexino experiment also identified neutrinos that were produced by the first step in the p-p chain, confirming the models of solar astronomers.

In 2020, scientists with the Borexino experiment were also able to identify neutrinos coming specifically from the CNO cycle of

nuclear fusion. As discussed toward the end of Telescopes Today, CNO fusion plays a very minor role in the Sun, but becomes important in stars that are more massive and hotter than the Sun.

It's amazing that a series of experiments that began with enough cleaning fluid to fill a swimming pool brought down the shafts of an old gold mine is now teaching us about the energy source of the Sun and the properties of matter! This is a good example of how experiments in astronomy and physics, coupled with the best theoretical models we can devise, continue to lead to fundamental changes in our understanding of nature.

This page titled 11.5: The Solar Interior - Observations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.5: The Solar Interior - Observations by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.6: The Solar Cycle

Learning Objectives

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

- Describe the sunspot cycle and, more generally, the solar cycle
- Explain how magnetism is the source of solar activity

Before the invention of the telescope, the Sun was thought to be an unchanging and perfect sphere. We now know that the Sun is in a perpetual state of change: its surface is a seething, bubbling cauldron of hot gas. Areas that are darker and cooler than the rest of the surface come and go. Vast plumes of gas erupt into the chromosphere and corona. Occasionally, there are even giant explosions on the Sun that send enormous streamers of charged particles and energy hurtling toward Earth. When they arrive, these can cause power outages and other serious effects on our planet.

Sunspots

The first evidence that the Sun changes came from studies of sunspots, which are large, dark features seen on the surface of the Sun caused by increased magnetic activity. They look darker because the spots are typically at a temperature of about 3800 K, whereas the bright regions that surround them are at about 5800 K (Figure 15.13). Occasionally, these spots are large enough to be visible to the unaided eye, and we have records going back over a thousand years from observers who noticed them when haze or mist reduced the Sun's intensity. (We emphasize what your parents have surely told you: looking at the Sun for even a brief time can cause permanent eye damage. This is the one area of astronomy where we don't encourage you to do your own observing without getting careful instructions or filters from your instructor.)

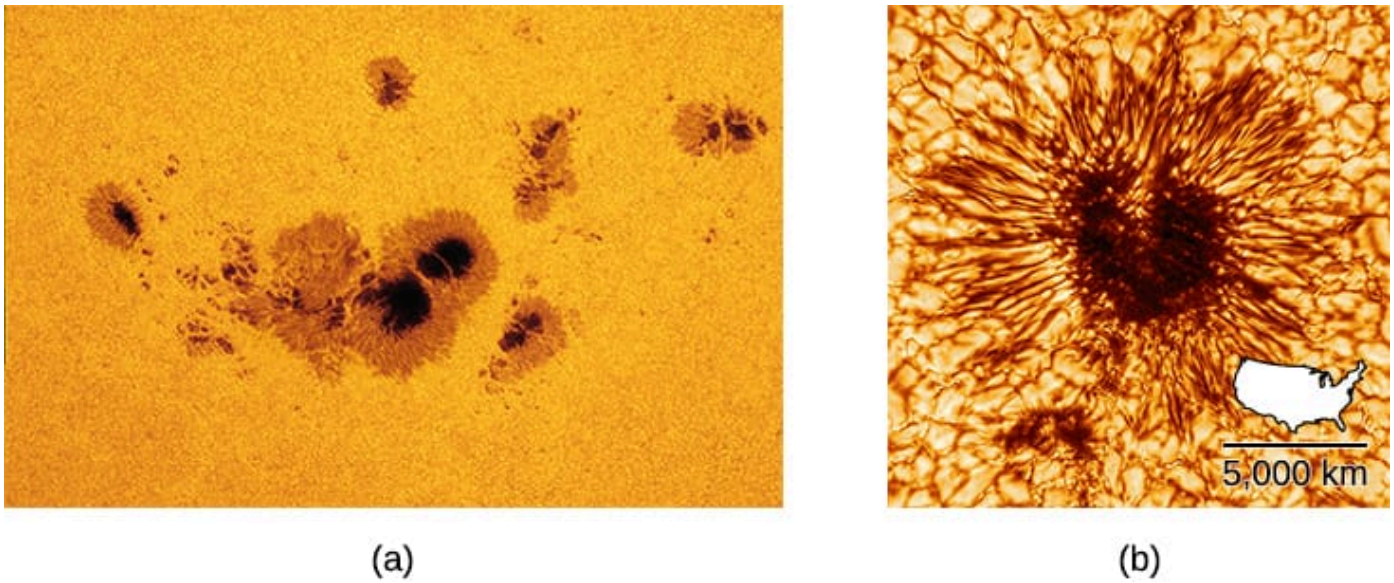


Figure 15.13 : Sunspots. a) Image of sunspots, cooler and thus darker regions on the Sun, taken in July 2012. You can see the dark, central region of each sunspot (called the umbra) surrounded by a less dark region (the penumbra). The largest spot shown here is about 11 Earths wide. b) The new Daniel Inouye Solar Telescope took the most detailed image of a sunspot ever recorded in January 2020. A map of the United States is shown to scale. Although sunspots appear dark when seen next to the hotter gases of the photosphere, an average sunspot, cut out of the solar surface and left standing in the night sky, would be about as bright as the full moon. The mottled appearance of the Sun's surface is granulation. (credit a: modification of work by NASA Goddard Space Flight Center, Alan Friedman; credit b: modification of work by NSO/AURA/NSF)

While we understand that sunspots look darker because they are cooler, they are nevertheless hotter than the surfaces of many stars. If they could be removed from the Sun, they would shine brightly. They appear dark only in contrast with the hotter, brighter

photosphere around them.

Individual sunspots come and go, with lifetimes that range from a few hours to a few months. If a spot lasts and develops, it usually consists of two parts: an inner darker core, the *umbra*, and a surrounding less dark region, the *penumbra*. Many spots become much larger than Earth, and a few, like the largest one shown in Figure 15.13, have reached diameters over 140,000 kilometers. Frequently, spots occur in groups of 2 to 20 or more. The largest groups are very complex and may have over 100 spots. Like storms on Earth, sunspots are not fixed in position, but they drift slowly compared with the Sun's rotation.

By recording the apparent motions of the sunspots as the turning Sun carried them across its disk (Figure 15.14), Galileo, in 1612, demonstrated that the Sun rotates on its axis with a rotation period of approximately 1 month. Our star turns in a west-to-east direction, like the orbital motions of the planets. The Sun, however, is a gas and does not have to rotate rigidly, the way a solid body like Earth does. Modern observations show that the speed of rotation of the Sun varies according to latitude, that is, it's different as you go north or south of the Sun's equator. The rotation period is about 25 days at the equator, 28 days at latitude 40°, and 36 days at latitude 80°. We call this behavior differential rotation.

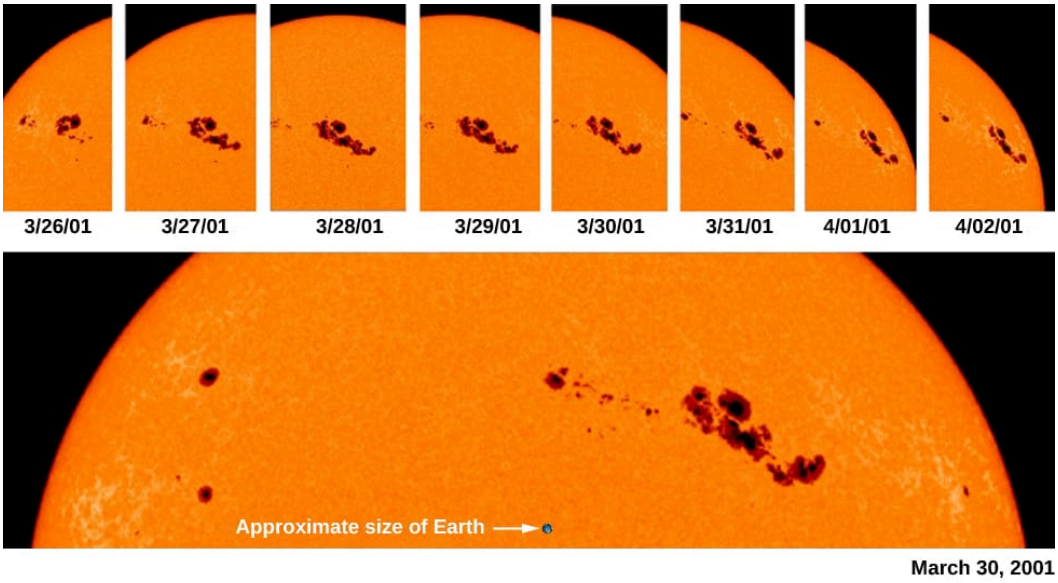


Figure 15.14 : Sunspots Rotate Across Sun's Surface. This sequence of photographs of the Sun's surface tracks the movement of sunspots across the visible hemisphere of the Sun. On March 30, 2001, this group of sunspots extended across an area about 13 times the diameter of Earth. This region produced many flares and coronal mass ejections. (credit: modification of work by SOHO/NASA/ESA)

The Sunspot Cycle

Between 1826 and 1850, Heinrich Schwabe, a German pharmacist and amateur astronomer, kept daily records of the number of sunspots. What he was really looking for was a planet inside the orbit of Mercury, which he hoped to find by observing its dark silhouette as it passed between the Sun and Earth. He failed to find the hoped-for planet, but his diligence paid off with an even-more important discovery: the sunspot cycle. He found that the number of sunspots varied systematically, in cycles about a decade long.

What Schwabe observed was that, although individual spots are short lived, the total number visible on the Sun at any one time was likely to be very much greater at certain times—the periods of *sunspot maximum*—than at other times—the periods of *sunspot minimum*. We now know that sunspot maxima occur at an *average* interval of 11 years, but the intervals between successive maxima have ranged from as short as 9 years to as long as 14 years. During sunspot maxima, more than 100 spots can often be seen at once. Even then, less than one-half of one percent of the Sun's surface is covered by spots (Figure 15.22). During sunspot minima, sometimes no spots are visible. The Sun's activity reached its most recent maximum in 2014.

Link to Learning

Watch this brief video from NASA's Goddard Space Flight Center that explains the sunspot cycle.

Explore the number of sunspots measured each year going back to 1707. Point your cursor at any year, and the diagram tells you the number of sunspots.

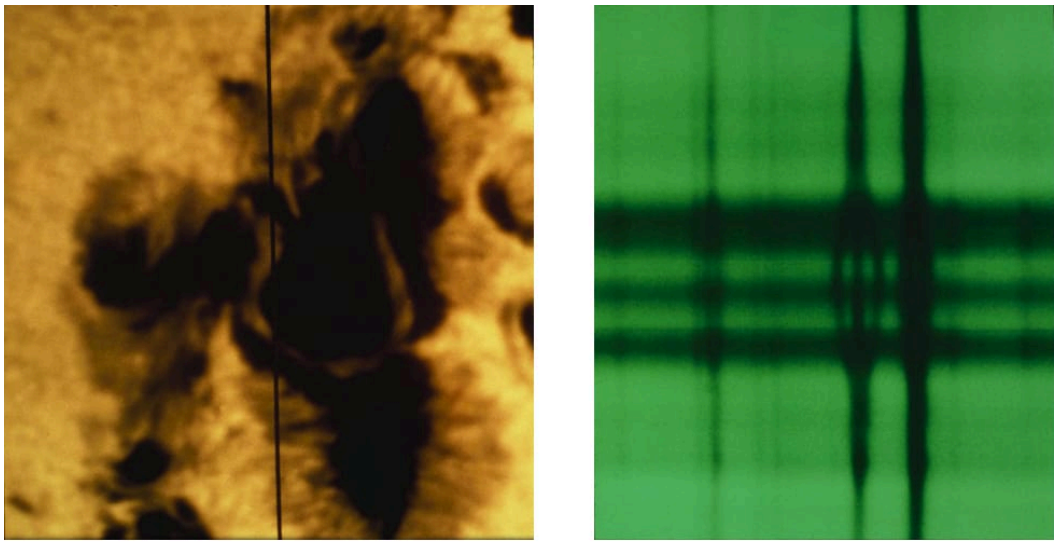
Magnetism and the Solar Cycle

Now that we have discussed the Sun's activity cycle, you might be asking, "Why does the Sun change in such a regular way?" Astronomers now understand that it is the Sun's changing magnetic field that drives solar activity.

The solar magnetic field is measured using a property of atoms called the *Zeeman effect*. Recall from Radiation and Spectra that an atom has many energy levels and that spectral lines are formed when electrons shift from one level to another. If each energy level is precisely defined, then the difference between them is also quite precise. As an electron changes levels, the result is a sharp, narrow spectral line (either an absorption or emission line, depending on whether the electron's energy increases or decreases in the transition).

In the presence of a strong magnetic field, however, each energy level is separated into several levels very close to one another. The separation of the levels is proportional to the strength of the field. As a result, spectral lines formed in the presence of a magnetic field are not single lines but a series of very closely spaced lines corresponding to the subdivisions of the atomic energy levels. This splitting of lines in the presence of a magnetic field is what we call the Zeeman effect (after the Dutch scientist who first discovered it in 1896).

Measurements of the Zeeman effect in the spectra of the light from sunspot regions show them to have strong magnetic fields (Figure 15.15). Bear in mind that magnets always have a north pole and a south pole. Whenever sunspots are observed in pairs, or in groups containing two principal spots, one of the spots usually has the magnetic polarity of a north-seeking magnetic pole and the other has the opposite polarity. Moreover, during a given cycle, the leading spots of pairs (or leading principle spots of groups) in the Northern Hemisphere all tend to have the same polarity, whereas those in the Southern Hemisphere all tend to have the opposite polarity.



(a) (b)

Figure 15.15 : Zeeman Effect. These photographs show how magnetic fields in sunspots are measured by means of the Zeeman effect. (left) The vertical black line indicates the position of the spectrograph slit through which light is passed to obtain the spectrum in (right). (credit: modification of work by NSO/AURA/NSF)

During the next sunspot cycle, however, the polarity of the leading spots is reversed in each hemisphere. For example, if during one cycle, the leading spots in the Northern Hemisphere all had the polarity of a north-seeking pole, then the leading spots in the Southern Hemisphere would have the polarity of a south-seeking pole. During the next cycle, the leading spots in the Northern Hemisphere would have south-seeking polarity, whereas those in the Southern Hemisphere would have north-seeking polarity. Therefore, strictly speaking, the sunspot cycle does not repeat itself in regard to magnetic polarity until two 11-year cycles have passed. A visual representation of the Sun's magnetic fields, called a *magnetogram*, can be used to see the relationship between sunspots and the Sun's magnetic field (Figure 15.16).

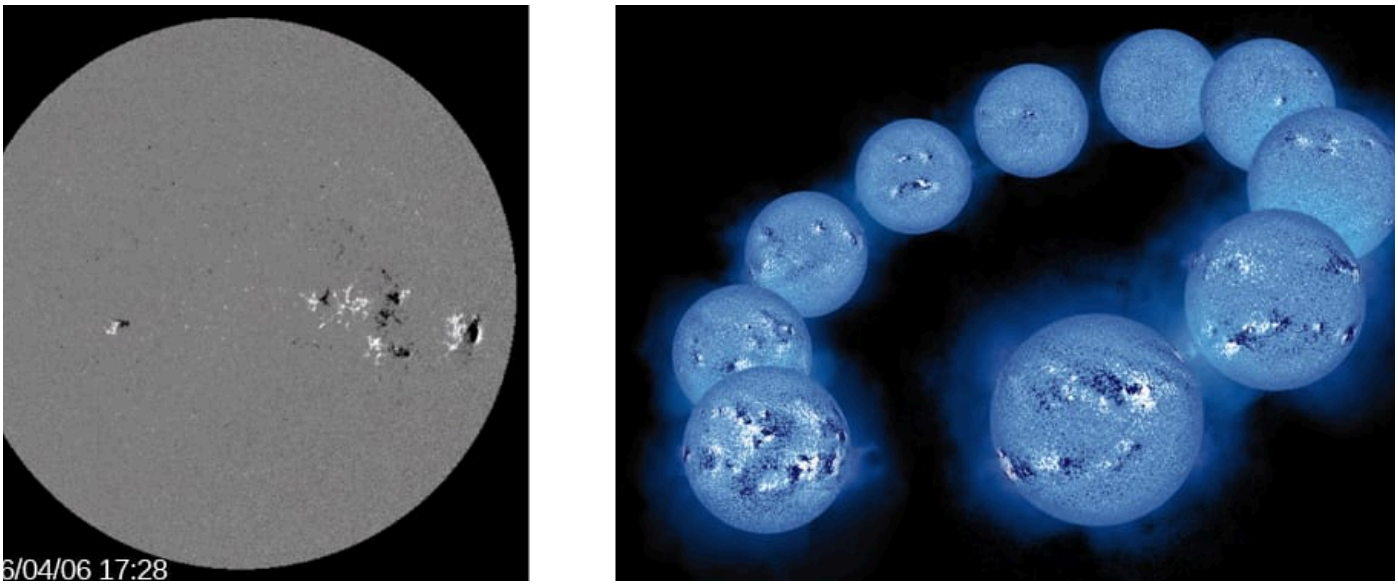


Figure 15.16 : Magnetogram and Solar Cycle. In the image on the left, called a magnetogram, we see the magnetic polarity of sunspots. The black areas are where the magnetism is pointing toward the Sun's core, whereas the white regions are where it is pointing away from the core, toward us. This dramatic sequence on the right shows the activity cycle of the Sun. The 10 maps of the magnetic field on the surface of the Sun span a period of 7.5 years. The two magnetic polarities (N and S) of the magnetic field are shown against a blue disk as dark blue to black (N) and as light blue to white (S). The earliest image, taken on January 8, 1992, is at the lower left and was taken just after solar maximum. Each image, from left to right around the arc, was taken one-half to one year after the preceding one. The last image was taken on July 25, 1999, as the Sun was approaching the next solar maximum. Note a few striking patterns in the magnetic maps: the direction from white to black polarity in the Southern Hemisphere is opposite from that in the Northern Hemisphere. (credit left: modification of work by NASA/SDO; credit right: modification of work by NASA/SOHO)

Why is the Sun such a strong and complicated magnet? Astronomers have found that it is the Sun's *dynamo* that generates the magnetic field. A dynamo is a machine that converts kinetic energy (i.e., the energy of motion) into electricity. On Earth, dynamos are found in power plants where, for example, the energy from wind or flowing water is used to cause turbines to rotate. In the Sun, the source of kinetic energy is the churning of turbulent layers of ionized gas within the Sun's interior that we mentioned earlier. These generate electric currents—moving electrons—which in turn generate magnetic fields.

Most solar researchers agree that the solar dynamo is located in the convection zone or in the interface layer between the convection zone and the radiative zone below it. As the magnetic fields from the Sun's dynamo interact, they break, reconnect, and rise through the Sun's surface.

We should say that, although we have good observations that show us *how* the Sun changes during each solar cycle, it is still very difficult to build physical models of something as complicated as the Sun that can account satisfactorily for *why* it changes. Researchers have not yet developed a generally accepted model that describes in detail the physical processes that control the solar cycle. Calculations do show that differential rotation (the idea that the Sun rotates at different rates at different latitudes) and convection just below the solar surface can twist and distort the magnetic fields. This causes them to grow and then decay, regenerating with opposite polarity approximately every 11 years. The calculations also show that as the fields grow stronger near solar maximum, they flow from the interior of the Sun toward its surface in the form of loops. When a large loop emerges from the solar surface, it creates regions of sunspot activity (Figure 15.17).

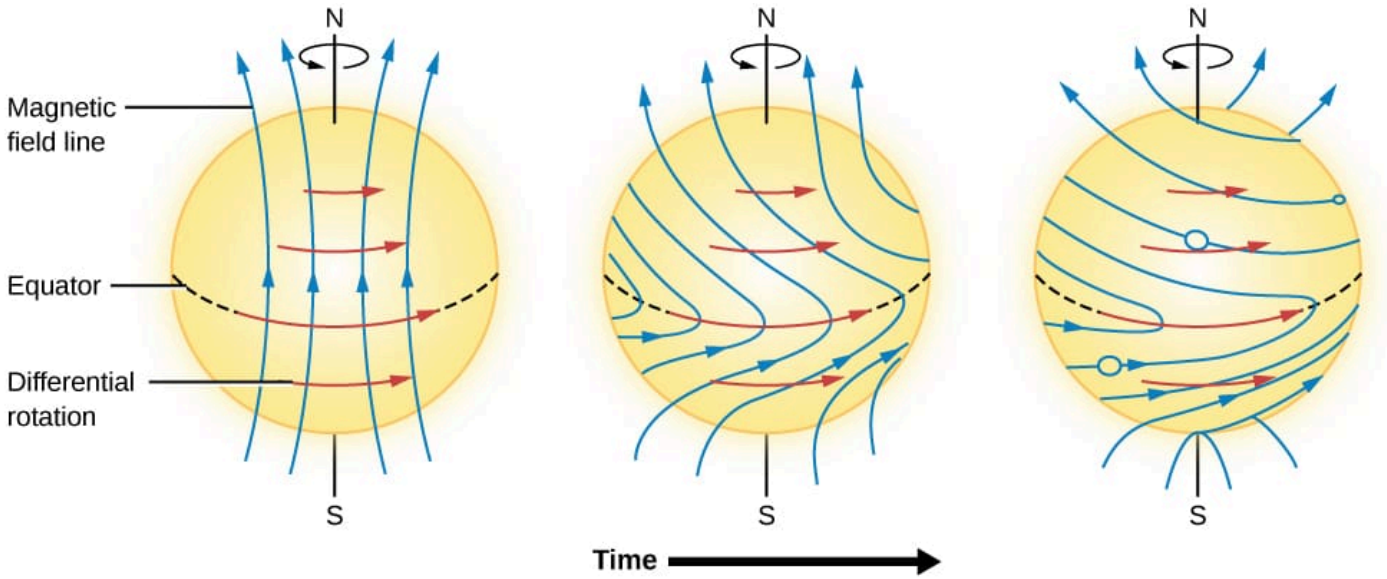


Figure 15.17 : Magnetic Field Lines Wind Up. Because the Sun spins faster at the equator than near the poles, the magnetic fields in the Sun tend to wind up as shown, and after a while make loops. This is an idealized diagram; the real situation is much more complex.

This idea of magnetic loops offers a natural explanation of why the leading and trailing sunspots in an active region have opposite polarity. The leading sunspot coincides with one end of the loop and the trailing spot with the other end. Magnetic fields also hold the key to explaining why sunspots are cooler and darker than the regions without strong magnetic fields. The forces produced by the magnetic field resist the motions of the bubbling columns of rising hot gases. Since these columns carry most of the heat from inside the Sun to the surface by means of convection, and strong magnetic fields inhibit this convection, the surface of the Sun is allowed to cool. As a result, these regions are seen as darker, cooler sunspots.

Beyond this general picture, researchers are still trying to determine why the magnetic fields are as large as they are, why the polarity of the field in each hemisphere flips from one cycle to the next, why the length of the solar cycle can vary from one cycle to the next, and why events like the Maunder Minimum occur.

Link to Learning

In this video solar scientist Holly Gilbert discusses the Sun's magnetic field.

This page titled 11.6: The Solar Cycle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 15.3: The Solar Cycle by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.7: Solar Activity above the Photosphere

Learning Objectives

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

- Describe the various ways in which the solar activity cycle manifests itself, including flares, coronal mass ejections, prominences, and plages

Sunspots are not the only features that vary during a solar cycle. There are dramatic changes in the chromosphere and corona as well. To see what happens in the chromosphere, we must observe the emission lines from elements such as hydrogen and calcium, which emit useful spectral lines at the temperatures in that layer. The hot corona, on the other hand, can be studied by observations of X-rays and of extreme ultraviolet and other wavelengths at high energies.

Plages and Prominences

As we saw, emission lines of hydrogen and calcium are produced in the hot gases of the chromosphere. Astronomers routinely photograph the Sun through filters that transmit light only at the wavelengths that correspond to these emission lines. Pictures taken through these special filters show bright "clouds" in the chromosphere around sunspots; these bright regions are known as plages (Figure 15.18). These are regions within the chromosphere that have higher temperature and density than their surroundings. The plages actually contain all of the elements in the Sun, not just hydrogen and calcium. It just happens that the spectral lines of hydrogen and calcium produced by these clouds are bright and easy to observe.

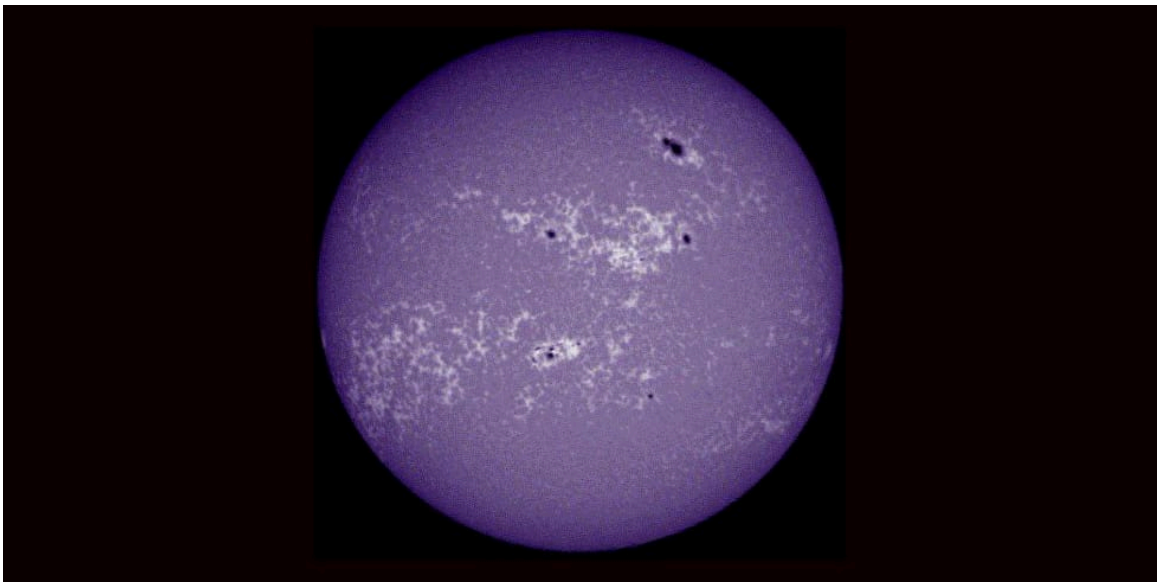


Figure 15.18 : Plages on the Sun. This image of the Sun was taken with a filter that transmits only the light of the spectral line produced by singly ionized calcium. The bright cloud-like regions are the plages. (credit: modification of work by NASA)

Moving higher into the Sun's atmosphere, we come to the spectacular phenomena called prominences (Figure 15.19), which usually originate near sunspots. Eclipse observers often see prominences as red features rising above the eclipsed Sun and reaching high into the corona. Some, the *quiescent* prominences, are graceful loops of plasma (ionized gas) that can remain nearly stable for many hours or even days. The relatively rare *eruptive* prominences appear to send matter upward into the corona at high speeds, and the most active *surge* prominences may move as fast as 1300 kilometers per second (almost 3 million miles per hour). Some eruptive prominences have reached heights of more than 1 million kilometers above the photosphere; Earth would be completely lost inside one of those awesome displays (Figure 15.19).

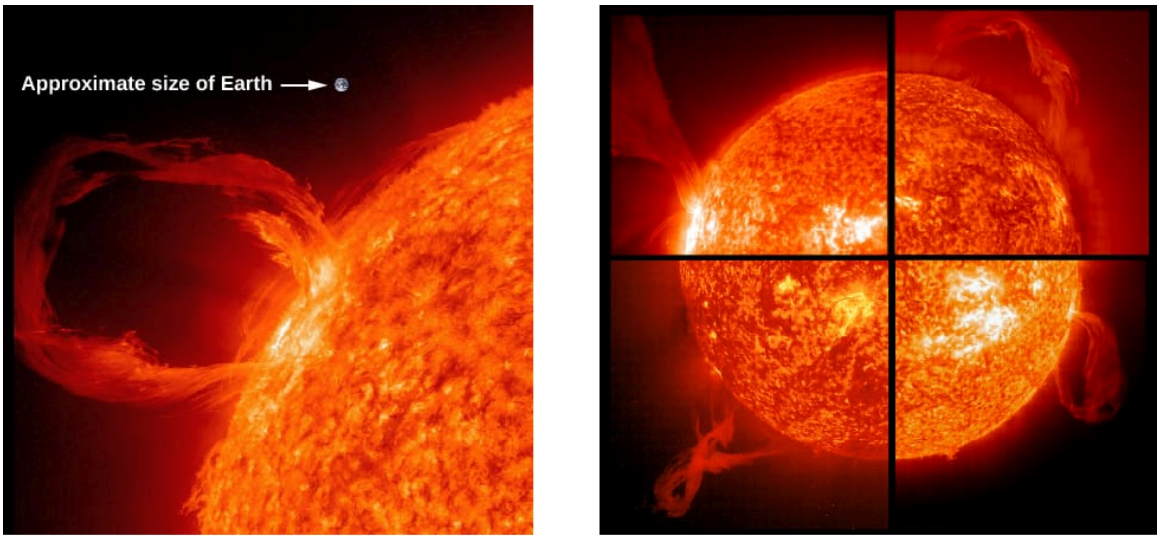


Figure 15.19 : Prominences. (a) This image of an eruptive prominence was taken in the light of singly ionized helium in the extreme ultraviolet part of the spectrum. The prominence is a particularly large one. An image of Earth is shown at the same scale for comparison. (b) A prominence is a huge cloud of relatively cool (about 60,000 K in this case), fairly dense gas suspended in the much hotter corona. These pictures, taken in ultraviolet, are color coded so that white corresponds to the hottest temperatures and dark red to cooler ones. The four images were taken, moving clockwise from the upper left, on May 15, 2001; March 28, 2000; January 18, 2000; and February 2, 2001. (credit a: modification of work by NASA/SOHO; credit b: modification of work by NASA/SDO)

Flares and Coronal Mass Ejections

The most violent event on the surface of the Sun is a rapid eruption called a solar flare (Figure 15.20). A typical flare lasts for 5 to 10 minutes and releases a total amount of energy equivalent to that of perhaps a million hydrogen bombs. The largest flares last for several hours and emit enough energy to power the entire United States at its current rate of electrical consumption for 100,000 years. Near sunspot maximum, small flares occur several times per day, and major ones may occur every few weeks.

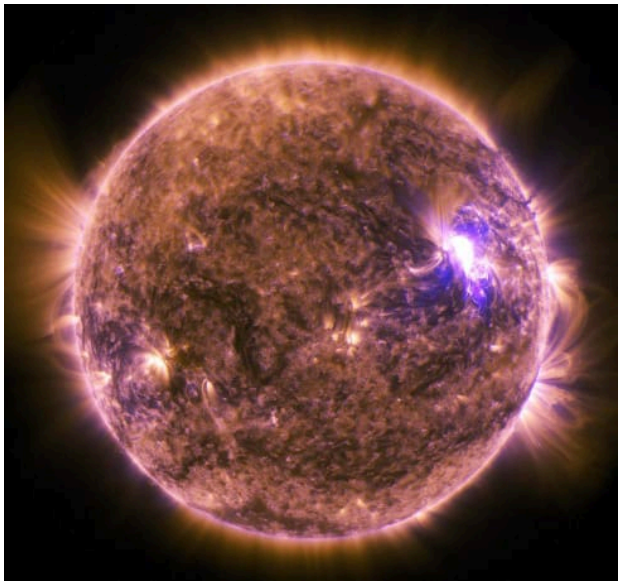


Figure 15.20 : Solar Flare. The bright white area seen on the right side of the Sun in this image from the Solar Dynamics Observer spacecraft is a solar flare that was observed on June 25, 2015. (credit: NASA/SDO)

Flares, like the one shown in Figure 15.21, are often observed in the red light of hydrogen, but the visible emission is only a tiny fraction of the energy released when a solar flare explodes. At the moment of the explosion, the matter associated with the flare is heated to temperatures as high as 10 million K. At such high temperatures, a flood of X-ray and ultraviolet radiation is emitted.

Flares seem to occur when magnetic fields pointing in opposite directions release energy by interacting with and destroying each other—much as a stretched rubber band releases energy when it breaks.

What is different about flares is that their magnetic interactions cover a large volume in the solar corona and release a tremendous amount of electromagnetic radiation. In some cases, immense quantities of coronal material—mainly protons and electrons—may also be ejected at high speeds (500–1000 kilometers per second) into interplanetary space. Such a coronal mass ejection (CME) can affect Earth in a number of ways (which we will discuss in the section on space weather).

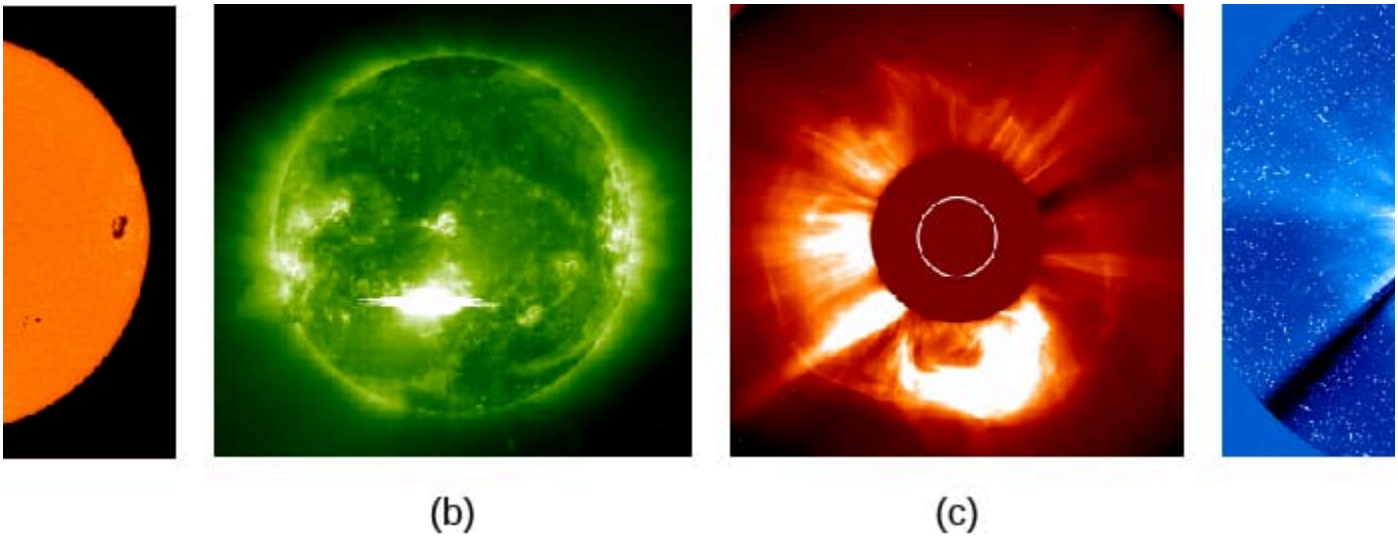


Figure 15.21 : Flare and Coronal Mass Ejection. This sequence of four images shows the evolution over time of a giant eruption on the Sun. (a) The event began at the location of a sunspot group, and (b) a flare is seen in far-ultraviolet light. (c) Fourteen hours later, a CME is seen blasting out into space. (d) Three hours later, this CME has expanded to form a giant cloud of particles escaping from the Sun and is beginning the journey out into the solar system. The white circle in (c) and (d) shows the diameter of the solar photosphere. The larger dark area shows where light from the Sun has been blocked out by a specially designed instrument to make it possible to see the faint emission from the corona. (credit a, b, c, d: modification of work by SOHO/EIT, SOHO/LASCO, SOHO/MDI (ESA & NASA))

Link to Learning

See a coronal mass ejection recorded by the Solar Dynamics Observatory.

Active Regions

To bring the discussion of the last two sections together, astronomers now realize that sunspots, flares, and bright regions in the chromosphere and corona tend to occur together on the Sun in time and space. That is, they all tend to have similar longitudes and latitudes, but they are located at different heights in the atmosphere. Because they all occur together, they vary with the sunspot cycle.

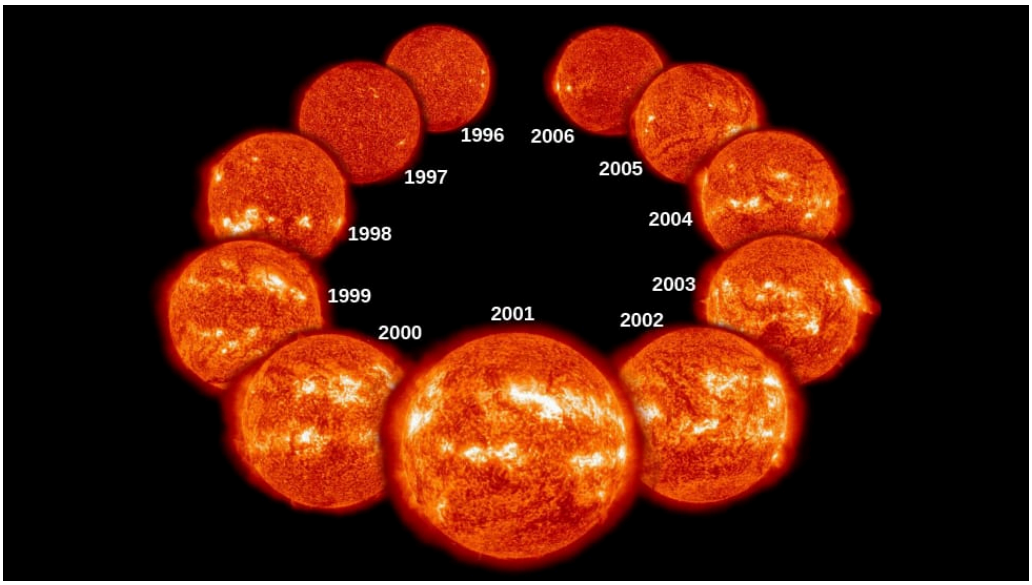


Figure 15.22 : Solar Cycle. This dramatic sequence of images taken from the SOHO satellite over a period of 11 years shows how active regions change during the solar cycle. The images were taken in the ultraviolet region of the spectrum and show that active regions on the Sun increase and decrease during the cycle. Sunspots are located in the cooler photosphere, beneath the hot gases shown in this image, and vary in phase with the emission from these hot gases—more sunspots and more emission from hot gases occur together. (credit: modification of work by ESA/NASA/SOHO)

For example, flares are more likely to occur near sunspot maximum, and the corona is much more conspicuous at that time (see Figure 15.22). A place on the Sun where a number of these phenomena are seen is called an active region (Figure 15.23). As you might deduce from our earlier discussion, active regions are always associated with strong magnetic fields.

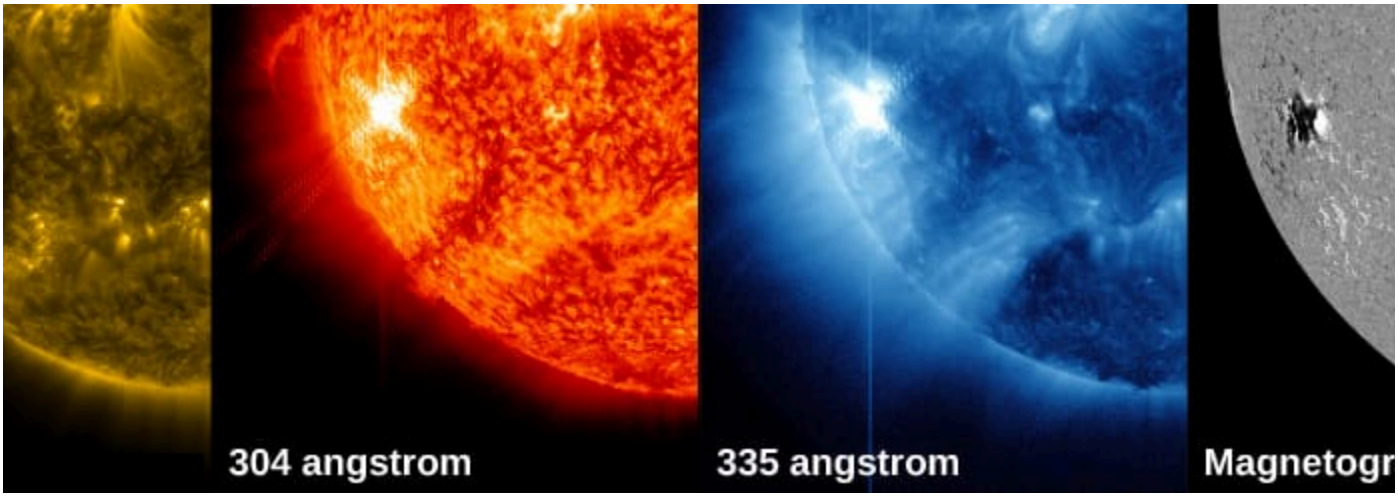


Figure 15.23 : Solar Active Region Observed at Different Heights in the Sun's Atmosphere. These four images of a solar flare on October 22, 2012, show from the left: light from the Sun at a wavelength of 171 angstroms, which shows the structure of loops of solar material in the corona; ultraviolet at 304 angstroms, which shows light from the region of the Sun's atmosphere where flares originate; light at 335 angstroms, which highlights radiation from active regions in the corona; a magnetogram, which shows magnetically active regions on the Sun. Note how these different types of activity all occur above a sunspot region with a strong magnetic field. (credit: modification of work by NASA/SDO/ Goddard)

This page titled 11.7: Solar Activity above the Photosphere is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 15.4: Solar Activity above the Photosphere by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

11.8: Space Weather

Learning Objectives

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

- Explain what space weather is and how it affects Earth

In the previous sections, we have seen that some of the particles coming off the Sun—either steadily as in the solar wind or in great bursts like CMEs—will reach Earth and its *magnetosphere* (the zone of magnetic influence that surrounds our planet). As if scientists did not have enough trouble trying to predict weather on Earth, this means that they are now facing the challenge of predicting the effects of solar storms on Earth. This field of research is called *space weather*; when that weather turns stormy, our technology turns out to be at risk.

With thousands of satellites in orbit, astronauts taking up long-term residence in the International Space Station, millions of people using cell phones, GPS, and wireless communication, and nearly everyone relying on the availability of dependable electrical power, governments are now making major investments in trying to learn how to predict when solar storms will occur and how strongly they will affect Earth.

Some History

What we now study as space weather was first recognized (though not yet understood) in 1859, in what is now known as the Carrington Event. In early September of that year, two amateur astronomers, including Richard Carrington in England, independently observed a solar flare. This was followed a day or two later by a significant solar storm reaching the region of Earth's magnetic field, which was soon overloaded with charged particles (see Earth as a Planet).

As a result, aurora activity was intense and the northern lights were visible well beyond their normal locations near the poles—as far south as Hawaii and the Caribbean. The glowing lights in the sky were so intense that some people reported getting up in the middle of the night, thinking it must be daylight.

The 1859 solar storm happened at a time when a new technology was beginning to tie people in the United States and some other countries together: the telegraph system. This was a machine and network for sending messages in code through overhead electrical wires (a bit like a very early version of the internet). The charged particles that overwhelmed Earth's magnetic field descended toward our planet's surface and affected the wires of the telegraph system. Sparks were seen coming out of exposed wires and out of the telegraph machines in the system's offices.

The observation of the bright flare that preceded these effects on Earth led to scientific speculation that a connection existed between solar activity and impacts on Earth—this was the beginning of our understanding of what today we call space weather.

Link to Learning

Watch NASA scientists answer some questions about space weather, and discuss some effects it can have in space and on Earth.

Sources of Space Weather

Three solar phenomena—coronal holes, solar flares, and CMEs—account for most of the space weather we experience. Coronal holes allow the solar wind to flow freely away from the Sun, unhindered by solar magnetic fields. When the solar wind reaches Earth, as we saw, it causes Earth's magnetosphere to contract and then expand after the solar wind passes by. These changes can cause (usually mild) electromagnetic disturbances on Earth.

More serious are solar flares, which shower the upper atmosphere of Earth with X-rays, energetic particles, and intense ultraviolet

radiation. The X-rays and ultraviolet radiation can ionize atoms in Earth's upper atmosphere, and the freed electrons can build up a charge on the surface of a spacecraft. When this static charge discharges, it can damage the electronics in the spacecraft—just as you can receive a shock when you walk across a carpet in your stocking feet in a dry climate and then touch a light switch or some other metal object.

Most disruptive are coronal mass ejections. A CME is an erupting bubble of tens of millions of tons of gas blown away from the Sun into space. When this bubble reaches Earth a few days after leaving the Sun, it heats the ionosphere, which expands and reaches farther into space. As a consequence, friction between the atmosphere and spacecraft increases, dragging satellites to lower altitudes.

At the time of a particularly strong flare and CME in March 1989, the system responsible for tracking some 19,000 objects orbiting Earth temporarily lost track of 11,000 of them because their orbits were changed by the expansion of Earth's atmosphere. During solar maximum, a number of satellites are brought to such a low altitude that they are destroyed by friction with the atmosphere. Both the Hubble Space Telescope and the International Space Station (Figure 15.24) require reboosts to higher altitude so that they can remain in orbit.

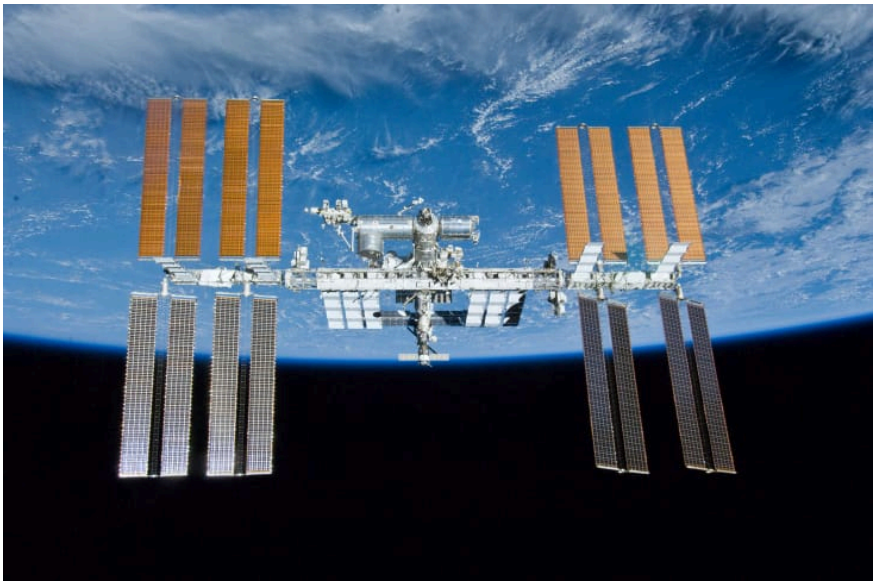


Figure 15.24 : International Space Station. The International Space Station is seen above Earth, as photographed in 2010 by the departing crew of the Space Shuttle Atlantis. (credit: NASA)

Solar Storm Damage on Earth

When a CME reaches Earth, it distorts Earth's magnetic field. Since a changing magnetic field induces electrical current, the CME accelerates electrons, sometimes to very high speeds. These "killer electrons" can penetrate deep into satellites, sometimes destroying their electronics and permanently disabling operation. This has happened with some communications satellites.

Disturbances in Earth's magnetic field can cause disruptions in communications, especially cell phone and wireless systems. In fact, disruptions can be expected to occur several times a year during solar maximum. Changes in Earth's magnetic field due to CMEs can also cause surges in power lines large enough to burn out transformers and cause major power outages. For example, in 1989, parts of Montreal and Quebec Province in Canada were without power for up to 9 hours as a result of a major solar storm. Electrical outages due to CMEs are more likely to occur in North America than in Europe because North America is closer to Earth's magnetic pole, where the currents induced by CMEs are strongest.

Besides changing the orbits of satellites, CMEs can also distort the signals sent by them. These effects can be large enough to reduce the accuracy of GPS-derived positions so that they cannot meet the limits required for airplane systems, which must know their positions to within 160 feet. Such disruptions caused by CMEs have occasionally forced the Federal Aviation Administration to restrict flights for minutes or, in a few cases, even days.

Solar storms also expose astronauts, passengers in high-flying airplanes, and even people on the surface of Earth to increased amounts of radiation. Astronauts, for example, are limited in the total amount of radiation to which they can be exposed during

their careers. A single ill-timed solar outburst could end an astronaut's career. This problem becomes increasingly serious as astronauts spend more time in space. For example, the typical *daily* dose of radiation aboard the Russian Mir space station was equivalent to about eight chest X-rays. One of the major challenges in planning the human exploration of Mars is devising a way to protect astronauts from high-energy solar radiation.

Advance warning of solar storms would help us minimize their disruptive effects. Power networks could be run at less than their full capacity so that they could absorb the effects of power surges. Communications networks could be prepared for malfunctions and have backup plans in place. Spacewalks could be timed to avoid major solar outbursts. Scientists are now trying to find ways to predict where and when flares and CMEs will occur, and whether they will be big, fast events or small, slow ones with little consequence for Earth.

The strategy is to relate changes in the appearance of small, active regions and changes in local magnetic fields on the Sun to subsequent eruptions. However, right now, our predictive capability is still poor, and so the only real warning we have is from actually seeing CMEs and flares occur. Since a CME travels outward at about 500 kilometers per second, an observation of an eruption provides several days warning at the distance of Earth. However, the severity of the impact on Earth depends on how the magnetic field associated with the CME is oriented relative to Earth's magnetic field. The orientation can be measured only when the CME flows past a satellite we have put up for this purpose. However, it is located only about an hour upstream from Earth.

Space weather predictions are now available online to scientists and the public. Outlooks are given a week ahead, bulletins are issued when there is an event that is likely to be of interest to the public, and warnings and alerts are posted when an event is imminent or already under way (Figure 15.25).

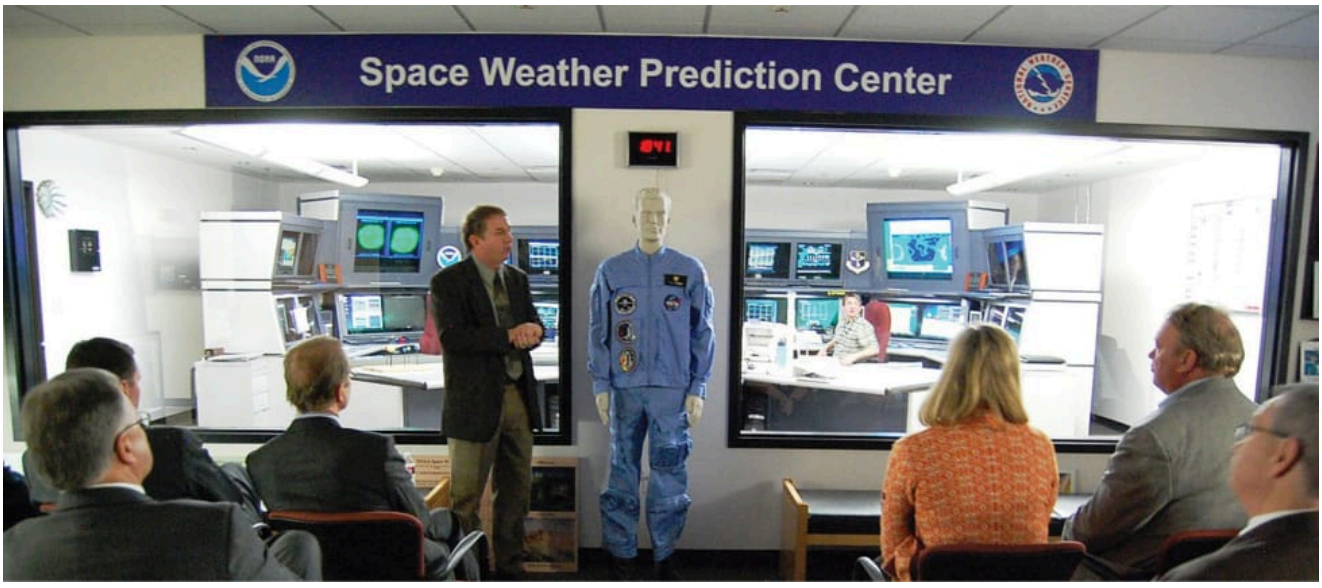


Figure 15.25 : NOAA Space Weather Prediction Operations Center. Bill Murtagh, a space weather forecaster, leads a workshop on preparedness for events like geomagnetic storms. (credit: modification of work by FEMA/Jerry DeFelice)

Link to Learning

To find public information and alerts about space weather, you can turn to the National Space Weather Prediction Center or SpaceWeather for consolidated information from many sources.

More satellites are being launched that will allow us to determine whether CMEs are headed toward Earth and how big they are. Models are being developed that will then allow scientists to use early information about the CME to predict its likely impact on Earth.

The hope is that soon, solar weather forecasting will have some of the predictive capability that meteorologists have achieved for terrestrial weather at Earth's surface. However, the most difficult events to predict are the largest and most damaging storms—hurricanes on Earth and extreme, rare storm events on the Sun. Thus, it is inevitable that the Sun will continue to surprise us.

Example 15.1: The Timing of Solar Events

A basic equation is useful in figuring out when events on the Sun will impact Earth:

$$\text{distance} = \text{velocity} \times \text{time}, \text{or } D = v \times t$$

Dividing both sides by *v*, we get

$$T = D/v$$

Suppose you observe a major solar flare while astronauts are orbiting Earth. If the average speed of solar wind is 400 km/s and the distance to the Sun is $1.496 \times 10^8 \text{km}$, how long will it be before the charged particles ejected from the Sun during the flare reach the space station?

Solution

The time required for solar wind particles to reach Earth is *T* = *D*/*v*.

$$\frac{1.496 \times 10^8 \text{ km}}{400 \text{ km/s}} = 3.74 \times 10^5 \text{ s}, \text{ or } \frac{3.74 \times 10^5 \text{ s}}{60 \text{ s/min} \times 60 \text{ min/h} \times 24 \text{ h/d}} = 4.3 \text{ d}$$

Exercise 11.8.1

How many days would it take for the particles to reach Earth if the solar wind speed increased to 500 km/s?

Answer available at the end of this section.

Earth's Climate and the Sunspot Cycle: Is There a Connection?

While the Sun rises faithfully every day at a time that can be calculated precisely, scientists have determined that the Sun's energy output is not truly constant but varies over the centuries by a small amount—probably less than 1%. We've seen that the number of sunspots varies, with the time between sunspot maxima of about 11 years, and that the number of sunspots at maximum is not always the same. Considerable evidence shows that between the years 1645 and 1715, the number of sunspots, even at sunspot maximum, was much lower than it is now. This interval of significantly low sunspot numbers was first noted by Gustav Spörer in 1887 and then by E. W. Maunder in 1890; it is now called the Maunder Minimum. The variation in the number of sunspots over the past three centuries is shown in Figure 15.26. Besides the Maunder Minimum in the seventeenth century, sunspot numbers were somewhat lower during the first part of the nineteenth century than they are now; this period is called the Little Maunder Minimum.

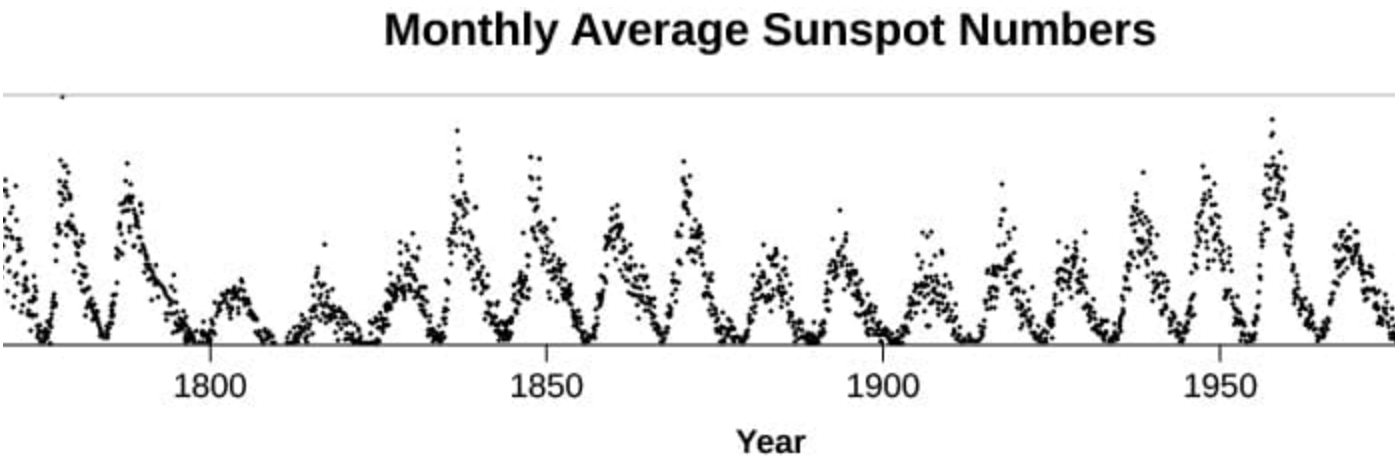


Figure 15.26 : Numbers of Sunspots over Time. This diagram shows how the number of sunspots has changed with time since counts of the numbers of spots began to be recorded on a consistent scale. Note the low number of spots during the early years of the nineteenth century, the Little Maunder Minimum. (credit: modification of work by NASA/ARC)

When the number of sunspots is high, the Sun is active in various other ways as well, and, as we will see in several sections below, some of this activity affects Earth directly. For example, there are more auroral displays when the sunspot number is high. Auroras are caused when energetically charged particles from the Sun interact with Earth's magnetosphere, and the Sun is more likely to spew out particles when it is active and the sunspot number is high. Historical accounts also indicate that auroral activity was abnormally low throughout the several decades of the Maunder Minimum.

The Maunder Minimum was a time of exceptionally low temperatures in Europe—so low that this period is described as the Little Ice Age. This coincidence in time caused scientists to try to understand whether small changes in the Sun could affect the climate on Earth. There is clear evidence that it was unusually cold in Europe during part of the seventeenth century. The River Thames in London froze at least 11 times, ice appeared in the oceans off the coasts of southeast England, and low summer temperatures led to short growing seasons and poor harvests. However, whether and how changes on the Sun on this timescale influence Earth's climate is still a matter of debate among scientists.

Other small changes in climate like the Little Ice Age have occurred and have had their impacts on human history. For example, explorers from Norway first colonized Iceland and then reached Greenland by 986. From there, they were able to make repeated visits to the northeastern coasts of North America, including Newfoundland, between about 1000 and 1350. (The ships of the time did not allow the Norse explorers to travel all the way to North America directly, but only from Greenland, which served as a station for further exploration.)

Most of Greenland is covered by ice, and the Greenland station was never self-sufficient; rather, it depended on imports of food and other goods from Norway for its survival. When a little ice age began in the thirteenth century, voyaging became very difficult, and support of the Greenland colony was no longer possible. The last-known contact with it was made by a ship from Iceland blown off course in 1410. When European ships again began to visit Greenland in 1577, the entire colony there had disappeared.

The estimated dates for these patterns of migration follow what we know about solar activity. Solar activity was unusually high between 1100 and 1250, which includes the time when the first European contacts were made with North America. Activity was low from 1280 to 1340 and there was a little ice age, which was about the time regular contact with North America and between Greenland and Europe stopped.

One must be cautious, however, about assuming that low sunspot numbers or variations in the Sun's output of energy *caused* the Little Ice Age. There is no satisfactory model that can explain how a reduction in solar activity might cause cooler temperatures on Earth. An alternative possibility is that the cold weather during the Little Ice Age was related to volcanic activity. Volcanoes can eject aerosols (tiny droplets or particles) into the atmosphere that efficiently reflect sunlight. Observations show, for example, that the Pinatubo eruption in 1991 ejected SO_2 aerosols into the atmosphere, which reduced the amount of sunlight reaching Earth's surface enough to lower global temperatures by 0.4 °C.

Satellite data show that the energy output from the Sun during a solar cycle varies by only about 0.1%. We know of no physical process that would explain how such a small variation could cause global temperature changes. The level of solar activity may, however, have other effects. For example, although the Sun's total energy output varies by only 0.1% during a solar cycle, its extreme ultraviolet radiation is 10 times higher at times of solar maximum than at solar minimum. This large variation can affect

the chemistry and temperature structure of the upper atmosphere. One effect might be a reduction in the ozone layer and a cooling of the stratosphere near Earth's poles. This, in turn, could change the circulation patterns of winds aloft and, hence, the tracks of storms. There is some recent evidence that variations in regional rainfall correlate better with solar activity than does the global temperature of Earth. But, as you can see, the relationship between what happens on the Sun and what happens to Earth's climate over the short term is still an area that scientists are investigating and debating.

Whatever the effects of solar activity may be on local rainfall or temperature patterns, we want to emphasize one important idea: Our climate change data and the models developed to account for the data consistently show that solar variability is *not* the cause of the global warming that has occurred during the past 50 years.

Answers

Exercise 11.8.1

$\frac{1.496 \times 10^8\,\text{km}}{500\,\text{km/s}} = 2.99 \times 10^5\,\text{s}$, or $\frac{2.99 \times 10^5\,\text{s}}{60\,\text{s/min} \times 60\,\text{min/h} \times 24\,\text{h/d}} = 3.46\,\text{d}$

This page titled 11.8: Space Weather is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 15.5: Space Weather by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12: THE STELLAR CENSUS

12.1: The Brightness of Stars
12.2: Surveying the Stars
12.3: Colors of Stars
12.4: The Spectra of Stars (and Brown Dwarfs)
12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion
12.6: A Stellar Census
12.7: Measuring Stellar Masses
12.8: Diameters of Stars
12.9: The H-R Diagram
12.10: Star Clusters
12.11: Checking Out the Theory

This page titled 12: The Stellar Census is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.1: The Brightness of Stars

Learning Objectives

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

- Explain the difference between luminosity and apparent brightness
- Understand how astronomers specify brightness with magnitudes

Luminosity

Perhaps the most important characteristic of a star is its luminosity—the total amount of energy at all wavelengths that it emits per second. Earlier, we saw that the Sun puts out a tremendous amount of energy every second. (And there are stars far more luminous than the Sun out there.) To make the comparison among stars easy, astronomers express the luminosity of other stars in terms of the Sun's luminosity. For example, the luminosity of Sirius is about 25 times that of the Sun. We use the symbol L_{Sun} to denote the Sun's luminosity; hence, that of Sirius can be written as 25 L_{Sun}. In a later chapter, we will see that if we can measure how much energy a star emits and we also know its mass, then we can calculate how long it can continue to shine before it exhausts its nuclear energy and begins to die.

Apparent Brightness

Astronomers are careful to distinguish between the luminosity of the star (the total energy output) and the amount of energy that happens to reach our eyes or a telescope on Earth. Stars are democratic in how they produce radiation; they emit the same amount of energy in every direction in space. Consequently, only a minuscule fraction of the energy given off by a star actually reaches an observer on Earth. We call the amount of a star's energy that reaches a given area (say, one square meter) each second here on Earth its apparent brightness. If you look at the night sky, you see a wide range of apparent brightnesses among the stars. Most stars, in fact, are so dim that you need a telescope to detect them.

If all stars were the same luminosity—if they were like standard bulbs with the same light output—we could use the difference in their apparent brightnesses to tell us something we very much want to know: how far away they are. Imagine you are in a big concert hall or ballroom that is dark except for a few dozen 25-watt bulbs placed in fixtures around the walls. Since they are all 25-watt bulbs, their luminosity (energy output) is the same. But from where you are standing in one corner, they do *not* have the same apparent brightness. Those close to you appear brighter (more of their light reaches your eye), whereas those far away appear dimmer (their light has spread out more before reaching you). In this way, you can tell which bulbs are closest to you. In the same way, if all the stars had the same luminosity, we could immediately infer that the brightest-appearing stars were close by and the dimmest-appearing ones were far away.

To pin down this idea more precisely, recall from the Radiation and Spectra chapter that we know exactly how light fades with increasing distance. The energy we receive is inversely proportional to the square of the distance. If, for example, we have two stars of the same luminosity and one is twice as far away as the other, it will look four times dimmer than the closer one. If it is three times farther away, it will look nine (three squared) times dimmer, and so forth.

Alas, the stars do not all have the same luminosity. (Actually, we are pretty glad about that because having many different types of stars makes the universe a much more interesting place.) But this means that if a star looks dim in the sky, we cannot tell whether it appears dim because it has a low luminosity but is relatively nearby, or because it has a high luminosity but is very far away. To measure the luminosities of stars, we must first compensate for the dimming effects of distance on light, and to do that, we must know how far away they are. Distance is among the most difficult of all astronomical measurements. We will return to how it is determined after we have learned more about the stars. For now, we will describe how astronomers specify the apparent brightness of stars.

The Magnitude Scale

The process of measuring the apparent brightness of stars is called *photometry* (from the Greek *photo* meaning “light” and *–metry* meaning “to measure”). As we saw Observing the Sky: The Birth of Astronomy, astronomical photometry began with Hipparchus. Around 150 B.C.E., he erected an observatory on the island of Rhodes in the Mediterranean. There he prepared a catalog of nearly 1000 stars that included not only their positions but also estimates of their apparent brightnesses.

Hipparchus did not have a telescope or any instrument that could measure apparent brightness accurately, so he simply made estimates with his eyes. He sorted the stars into six brightness categories, each of which he called a magnitude. He referred to the brightest stars in his catalog as first-magnitudes stars, whereas those so faint he could barely see them were sixth-magnitude stars. During the nineteenth century, astronomers attempted to make the scale more precise by establishing exactly how much the apparent brightness of a sixth-magnitude star differs from that of a first-magnitude star. Measurements showed that we receive about 100 times more light from a first-magnitude star than from a sixth-magnitude star. Based on this measurement, astronomers then defined an accurate magnitude system in which a difference of five magnitudes corresponds exactly to a brightness ratio of 100:1. In addition, the magnitudes of stars are decimalized; for example, a star isn’t just a “second-magnitude star,” it has a magnitude of 2.0 (or 2.1, 2.3, and so forth). So what number is it that, when multiplied together five times, gives you this factor of 100? Play on your calculator and see if you can get it. The answer turns out to be about 2.5, which is the fifth root of 100. This means that a magnitude 1.0 star and a magnitude 2.0 star differ in brightness by a factor of about 2.5. Likewise, we receive about 2.5 times as much light from a magnitude 2.0 star as from a magnitude 3.0 star. What about the difference between a magnitude 1.0 star and a magnitude 3.0 star? Since the difference is 2.5 times for each “step” of magnitude, the total difference in brightness is 2.5 × 2.5 = 6.25 times.

Here are a few rules of thumb that might help those new to this system. If two stars differ by 0.75 magnitudes, they differ by a factor of about 2 in brightness. If they are 2.5 magnitudes apart, they differ in brightness by a factor of 10, and a 4-magnitude difference corresponds to a difference in brightness of a factor of 40.You might be saying to yourself at this point, “Why do astronomers continue to use this complicated system from more than 2000 years ago?” That’s an excellent question and, as we shall discuss, astronomers today can use other ways of expressing how bright a star looks. But because this system is still used in many books, star charts, and computer apps, we felt we had to introduce students to it (even though we were very tempted to leave it out.)

The brightest stars, those that were traditionally referred to as first-magnitude stars, actually turned out (when measured accurately) not to be identical in brightness. For example, the brightest star in the sky, Sirius, sends us about 10 times as much light as the average first-magnitude star. On the modern magnitude scale, Sirius, the star with the brightest apparent magnitude, has been assigned a magnitude of –1.5. Other objects in the sky can appear even brighter. Venus at its brightest is of magnitude –4.4, while the Sun has a magnitude of –26.8. Figure 17.2 shows the range of observed magnitudes from the brightest to the faintest, along with the actual magnitudes of several well-known objects. The important fact to remember when using magnitude is that the system goes backward: the *larger* the magnitude, the *fainter* the object you are observing.

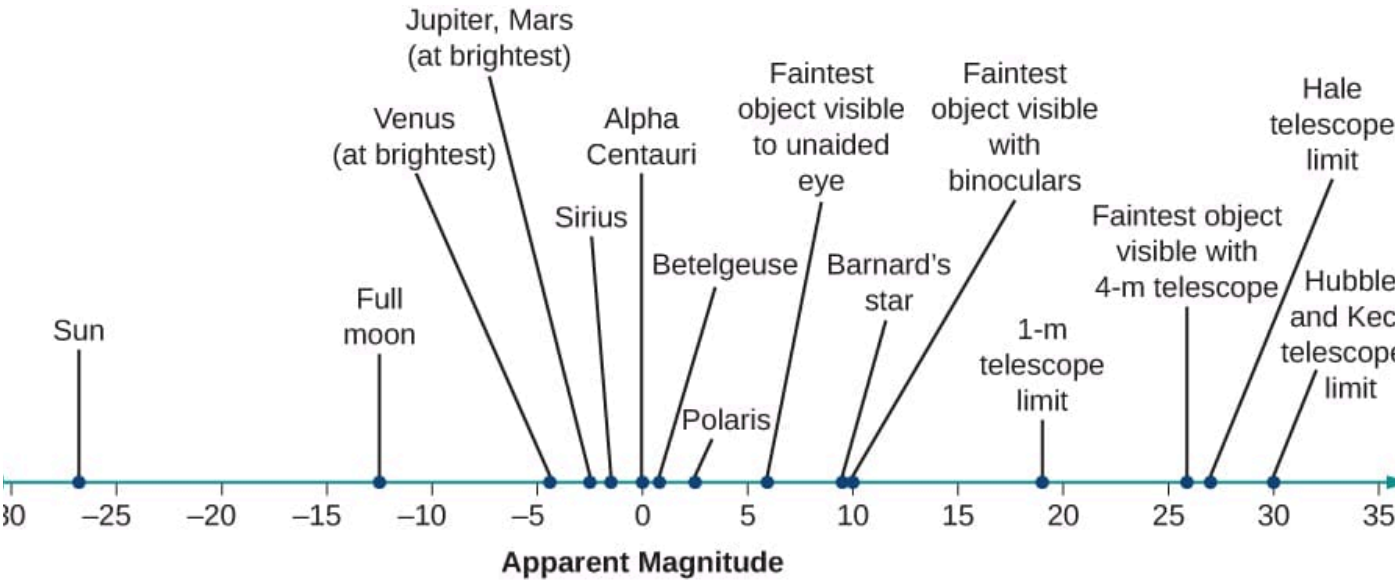


Figure 17.2 : Apparent Magnitudes of Well-Known Objects. The faintest magnitudes that can be detected by the unaided eye, binoculars, and large telescopes are also shown.

Example 17.1: The Magnitude Equation

Even scientists can't calculate fifth roots in their heads, so astronomers have summarized the above discussion in an equation to help calculate the difference in brightness for stars with different magnitudes. If m_1 and m_2 are the magnitudes of two stars, then we can calculate the ratio of their brightness $\left(\frac{b_2}{b_1}\right)$ using this equation:

$$m_1 - m_2 = 2.5 \log\left(\frac{b_2}{b_1}\right) \qquad \text{or} \qquad \frac{b_2}{b_1} = 2.5^{m_1-}$$

Here is another way to write this equation:

$$\frac{b_2}{b_1} = \left(100^{0.2}\right)^{m_1-}$$

Let's do a real example, just to show how this works. Imagine that an astronomer has discovered something special about a dim star (magnitude 8.5), and she wants to tell her students how much dimmer the star is than Sirius. Star 1 in the equation will be our dim star and star 2 will be Sirius.

Solution

Remember, Sirius has a magnitude of -1.5. In that case:

$$\begin{aligned} \frac{b_2}{b_1} &= \left(100^{0.2}\right)^{8.5-(-1.5)} = \left(100^{0.2}\right)^{10} \\ &= (100)^2 = 100 \times 100 = 10, \end{aligned}$$

Exercise 12.1.1

It is a common misconception that Polaris (magnitude 2.0) is the brightest star in the sky, but, as we saw, that distinction actually belongs to Sirius (magnitude –1.5). How does Sirius' apparent brightness compare to that of Polaris?

Answer available at the end of this section.

Other Units of Brightness

Although the magnitude scale is still used for visual astronomy, it is not used at all in newer branches of the field. In radio astronomy, for example, no equivalent of the magnitude system has been defined. Rather, radio astronomers measure the amount of energy being collected each second by each square meter of a radio telescope and express the brightness of each source in terms of, for example, watts per square meter.

Similarly, most researchers in the fields of infrared, X-ray, and gamma-ray astronomy use energy per area per second rather than magnitudes to express the results of their measurements. Nevertheless, astronomers in all fields are careful to distinguish between the *luminosity* of the source (even when that luminosity is all in X-rays) and the amount of energy that happens to reach us on Earth. After all, the luminosity is a really important characteristic that tells us a lot about the object in question, whereas the energy that reaches Earth is an accident of cosmic geography.

To make the comparison among stars easy, in this text, we avoid the use of magnitudes as much as possible and will express the luminosity of other stars in terms of the Sun's luminosity. For example, the luminosity of Sirius is 25 times that of the Sun. We use the symbol L_{Sun} to denote the Sun's luminosity; hence, that of Sirius can be written as 25 L_{Sun}.

Answers

Exercise 12.1.1

$\frac{b_{\text{Sirius}}}{b_{\text{Polaris}}} = \left(100^{0.2}\right)^{2.0-(-1.5)} = \left(100^{0.2}\right)^{3.5} = 100^{0.7} =$

(Hint: If you only have a basic calculator, you may wonder how to take 100 to the 0.7th power. But this is something you can ask Google to do. Google now accepts mathematical questions and will answer them. So try it for yourself. Ask Google, "What is 100 to the 0.7th power?")

Our calculation shows that Sirius' apparent brightness is 25 times greater than Polaris' apparent brightness.

This page titled 12.1: The Brightness of Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 17.2: The Brightness of Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.2: Surveying the Stars

Learning Objectives

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

- Understand the concept of triangulating distances to distant objects, including stars
- Explain why space-based satellites deliver more precise distances than ground-based methods
- Discuss astronomers' efforts to determine the distance of the stars closest to the Sun

It is an enormous step to go from the planets to the stars. For example, our Voyager 1 probe, which was launched in 1977, has now traveled farther from Earth than any other spacecraft. As of 2021, Voyager 1 is 152 AU from the Sun.[1] The nearest star, however, is hundreds of thousands of AU from Earth. Even so, we can, in principle, survey distances to the stars using the same technique that a civil engineer employs to survey the distance to an inaccessible mountain or tree—the method of *triangulation*.

Triangulation in Space

A practical example of triangulation is your own depth perception. As you are pleased to discover every morning when you look in the mirror, your two eyes are located some distance apart. You therefore view the world from two different vantage points, and it is this dual perspective that allows you to get a general sense of how far away objects are.

To see what we mean, take a pen and hold it a few inches in front of your face. Look at it first with one eye (closing the other) and then switch eyes. Note how the pen seems to shift relative to objects across the room. Now hold the pen at arm's length: the shift is less. If you play with moving the pen for a while, you will notice that the farther away you hold it, the less it seems to shift. Your brain automatically performs such comparisons and gives you a pretty good sense of how far away things in your immediate neighborhood are.

If your arms were made of rubber, you could stretch the pen far enough away from your eyes that the shift would become imperceptible. This is because our depth perception fails for objects more than a few tens of meters away. In order to see the shift of an object a city block or more from you, your eyes would need to be spread apart a lot farther.

Let's see how surveyors take advantage of the same idea. Suppose you are trying to measure the distance to a tree across a deep river (Figure 19.4). You set up two observing stations some distance apart. That distance (line AB in Figure 19.4) is called the *baseline*. Now the direction to the tree (C in the figure) in relation to the baseline is observed from each station. Note that C appears in different directions from the two stations. This apparent change in direction of the remote object due to a change in vantage point of the observer is called parallax.

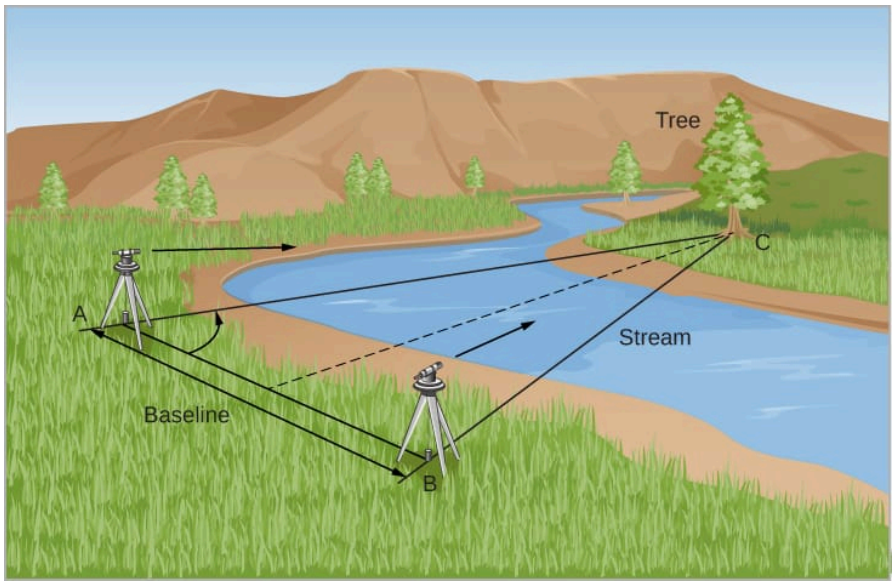


Figure 19.4 : Triangulation. Triangulation allows us to measure distances to inaccessible objects. By getting the angle to a tree from two different vantage points, we can calculate the properties of the triangle they make and thus the distance to the tree.

The parallax is also the angle that lines AC and BC make—in mathematical terms, the angle subtended by the baseline. A knowledge of the angles at A and B and the length of the baseline, AB, allows the triangle ABC to be solved for any of its dimensions—say, the distance AC or BC. The solution could be reached by constructing a scale drawing or by using trigonometry to make a numerical calculation. If the tree were farther away, the whole triangle would be longer and skinnier, and the parallax angle would be smaller. Thus, we have the general rule that the smaller the parallax, the more distant the object we are measuring must be.

In practice, the kinds of baselines surveyors use for measuring distances on Earth are completely useless when we try to gauge distances in space. The farther away an astronomical object lies, the longer the baseline has to be to give us a reasonable chance of making a measurement. Unfortunately, nearly all astronomical objects are very far away. To measure their distances requires a very large baseline and highly precise angular measurements. The Moon is the only object near enough that its distance can be found fairly accurately with measurements made without a telescope. Ptolemy determined the distance to the Moon correctly to within a few percent. He used the turning Earth itself as a baseline, measuring the position of the Moon relative to the stars at two different times of night.

With the aid of telescopes, later astronomers were able to measure the distances to the nearer planets and asteroids using Earth's diameter as a baseline. This is how the AU was first established. To reach for the stars, however, requires a much longer baseline for triangulation and extremely sensitive measurements. Such a baseline is provided by Earth's annual trip around the Sun.

Distances to Stars

As Earth travels from one side of its orbit to the other, it graciously provides us with a baseline of 2 AU, or about 300 million kilometers. Although this is a much bigger baseline than the diameter of Earth, the stars are *so far away* that the resulting parallax shift is *still* not visible to the naked eye—not even for the closest stars.

In the chapter on Observing the Sky: The Birth of Astronomy, we discussed how this dilemma perplexed the ancient Greeks, some of whom had actually suggested that the Sun might be the center of the solar system, with Earth in motion around it. Aristotle and others argued, however, that Earth could not be revolving about the Sun. If it were, they said, we would surely observe the parallax of the nearer stars against the background of more distant objects as we viewed the sky from different parts of Earth's orbit (Figure 19.6). Tycho Brahe (1546–1601) advanced the same faulty argument nearly 2000 years later, when his careful measurements of stellar positions with the unaided eye revealed no such shift.

These early observers did not realize how truly distant the stars were and how small the change in their positions therefore was, even with the entire orbit of Earth as a baseline. The problem was that they did not have tools to measure parallax shifts too small to be seen with the human eye. By the eighteenth century, when there was no longer serious doubt about Earth's revolution, it became clear that the stars must be extremely distant. Astronomers equipped with telescopes began to devise instruments

capable of measuring the tiny shifts of nearby stars relative to the background of more distant (and thus unshifting) celestial objects.

This was a significant technical challenge, since, even for the nearest stars, parallax angles are usually only a fraction of a second of arc. Recall that one second of arc (arcsec) is an angle of only 1/3600 of a degree. A coin the size of a US quarter would appear to have a diameter of 1 arcsecond if you were viewing it from a distance of about 5 kilometers (3 miles). Think about how small an angle that is. No wonder it took astronomers a long time before they could measure such tiny shifts.

The first successful detections of stellar parallax were in the year 1838, when Friedrich Bessel in Germany (Figure 19.5), Thomas Henderson, a Scottish astronomer working at the Cape of Good Hope, and Friedrich Struve in Russia independently measured the parallaxes of the stars 61 Cygni, Alpha Centauri, and Vega, respectively. Even the closest star, Alpha Centauri, showed a total displacement of only about 1.5 arcseconds during the course of a year.

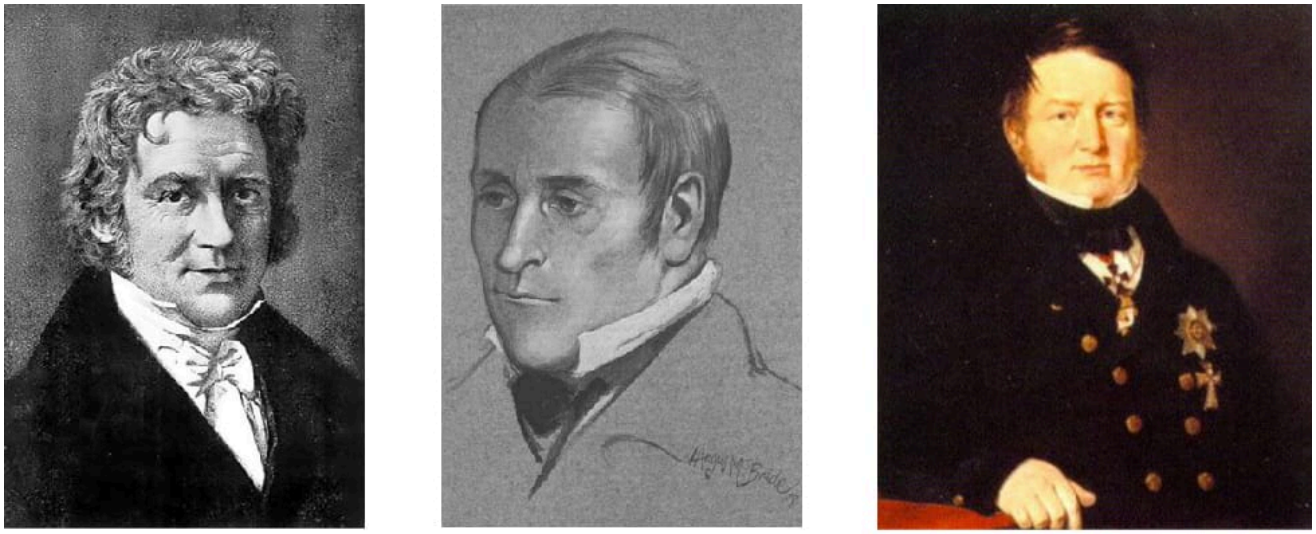


Figure 19.5 : Friedrich Wilhelm Bessel (1784–1846), Thomas J. Henderson (1798–1844), and Friedrich Struve (1793–1864). (a) Bessel made the first authenticated measurement of the distance to a star (61 Cygni) in 1838, a feat that had eluded many dedicated astronomers for almost a century. But two others, (b) Scottish astronomer Thomas J. Henderson and (c) Friedrich Struve, in Russia, were close on his heels.

Figure 19.6 shows how such measurements work. Seen from opposite sides of Earth's orbit, a nearby star shifts position when compared to a pattern of more distant stars. Astronomers actually define parallax to be *one-half* the angle that a star shifts when seen from opposite sides of Earth's orbit (the angle labeled *P* in Figure 19.6). The reason for this definition is just that they prefer to deal with a baseline of 1 AU instead of 2 AU.

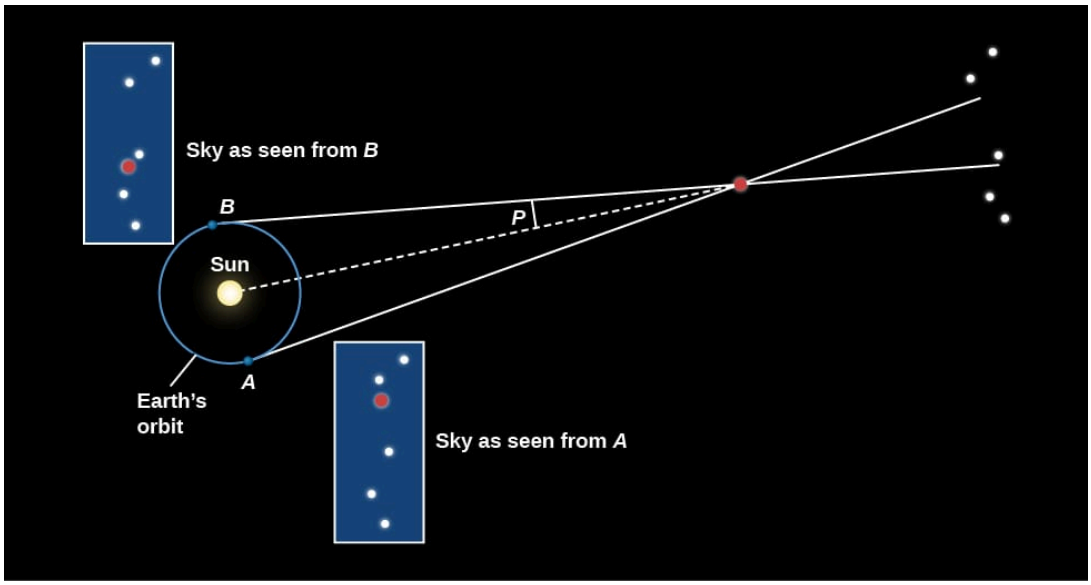


Figure 19.6 : Parallax. As Earth revolves around the Sun, the direction in which we see a nearby star varies with respect to distant stars. We define the parallax of the nearby star to be one half of the total change in direction, and we usually measure it in arcseconds.

Link to Learning

Use the Astronomical Parallax model to explore how the Earth's motion around the Sun causes nearby stars to appear to "wobble" back and forth compared to background stars. Click on the Orbit box, and use the view along the bottom to see the apparent motion of the nearby object.

Units of Stellar Distance

With a baseline of one AU, how far away would a star have to be to have a parallax of 1 arcsecond? The answer turns out to be 206,265 AU, or 3.26 light-years. This is equal to 3.1×10^{13} kilometers (in other words, 31 trillion kilometers). We give this unit a special name, the parsec (pc)—derived from "the distance at which we have a *par*allax of one *sec*ond." The distance (D) of a star in parsecs is just the reciprocal of its parallax (p) in arcseconds; that is,

$$D = \frac{1}{p}$$

Thus, a star with a parallax of 0.1 arcsecond would be found at a distance of 10 parsecs, and one with a parallax of 0.05 arcsecond would be 20 parsecs away.

Back in the days when most of our distances came from parallax measurements, a parsec was a useful unit of distance, but it is not as intuitive as the light-year. One advantage of the light-year as a unit is that it emphasizes the fact that, as we look out into space, we are also looking back into time. The light that we see from a star 100 light-years away left that star 100 years ago. What we study is not the star as it is now, but rather as it was in the past. The light that reaches our telescopes today from distant galaxies left them before Earth even existed.

In this text, we will use light-years as our unit of distance, but many astronomers still use parsecs when they write technical papers or talk with each other at meetings. To convert between the two distance units, just bear in mind: 1 parsec = 3.26 light-year, and 1 light-year = 0.31 parsec.

Example 19.1: How Far Is a Light-Year?

A light-year is the distance light travels in 1 year. Given that light travels at a speed of 300,000 km/s, how many kilometers are there in a light-year?

Solution

We learned earlier that speed = distance/time. We can rearrange this equation so that distance = velocity × time. Now, we need to determine the number of seconds in a year.

There are approximately 365 days in 1 year. To determine the number of seconds, we must estimate the number of seconds in 1 day.

We can change units as follows (notice how the units of time cancel out):

$$1\,\text{day} \times 24\,\text{hr/day} \times 60\,\text{min/hr} \times 60\,\text{s/min} = 86{,}400\,\text{s/day}$$

Next, to get the number of seconds per year:

$$365\,\text{days/year} \times 86{,}400\,\text{s/day} = 31{,}536{,}000\,\text{s/year}$$

Now we can multiply the speed of light by the number of seconds per year to get the distance traveled by light in 1 year:

$$\begin{aligned}\text{distance} &= \text{velocity} \times \text{time}\\ &= 300{,}000\,\text{km/s} \times 31{,}536{,}000\,\text{s}\\ &= 9.46 \times 10^{12}\,\text{km}\end{aligned}$$

That's almost 10,000,000,000,000 km that light covers in a year. To help you imagine how long this distance is, we'll mention that a string 1 light-year long could fit around the circumference of Earth 236 million times.

Exercise 12.2.1

The number above is really large. What happens if we put it in terms that might be a little more understandable, like the diameter of Earth? Earth's diameter is about 12,700 km.

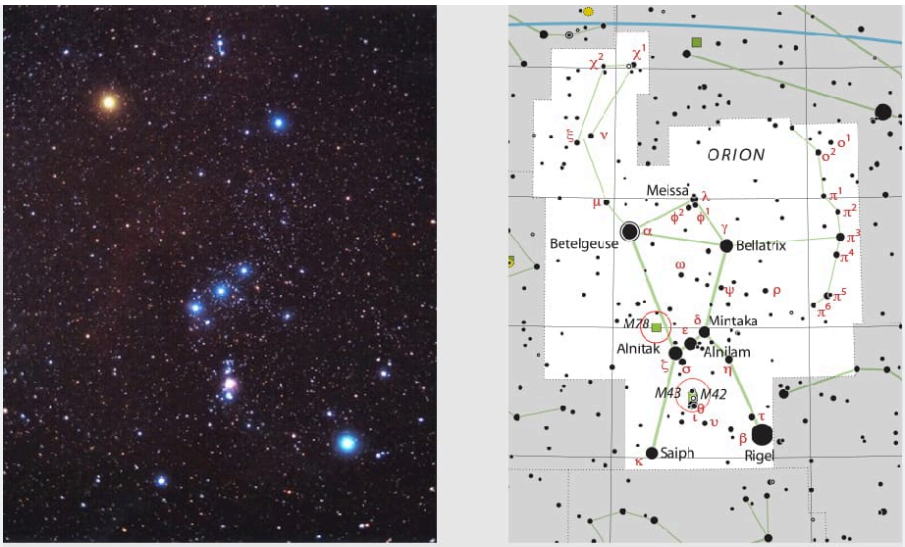
Answer available at the end of this section.

Astronomy Basics: Naming Stars

You may be wondering why stars have such a confusing assortment of names. Just look at the first three stars to have their parallaxes measured: 61 Cygni, Alpha Centauri, and Vega. Each of these names comes from a different tradition of designating stars.

The brightest stars have names that derive from the ancients. Some are from the Greek, such as Sirius, which means "the scorched one"—a reference to its brilliance. A few are from Latin, but many of the best-known names are from Arabic because, as discussed in Observing the Sky: The Birth of Astronomy, much of Greek and Roman astronomy was "rediscovered" in Europe after the Dark Ages by means of Arabic translations. Vega, for example, means "swooping Eagle," and Betelgeuse (pronounced "Beetle-juice") means "right hand of the central one."

In 1603, German astronomer Johann Bayer (1572–1625) introduced a more systematic approach to naming stars. For each constellation, he assigned a Greek letter to the brightest stars, roughly in order of brightness. In the constellation of Orion, for example, Betelgeuse is the brightest star, so it got the first letter in the Greek alphabet—alpha—and is known as Alpha Orionis. ("Orionis" is the possessive form of Orion, so Alpha Orionis means "the first of Orion.") A star called Rigel, being the second brightest in that constellation, is called Beta Orionis (Figure 19.7). Since there are 24 letters in the Greek alphabet, this system allows the labeling of 24 stars in each constellation, but constellations have many more stars than that.



(a) (b)

Figure 19.7 : Objects in Orion. (a) This image shows the brightest objects in or near the star pattern of Orion, the hunter (of Greek mythology), in the constellation of Orion. (b) Note the Greek letters of Bayer's system in this diagram of the Orion constellation. The objects denoted M42, M43, and M78 are not stars but nebulae—clouds of gas and dust; these numbers come from a list of "fuzzy objects" made by Charles Messier in 1781. (credit a: modification of work by Matthew Spinelli; credit b: modification of work by ESO, IAU and *Sky & Telescope*)

In 1725, the English Astronomer Royal John Flamsteed introduced yet another system, in which the brighter stars eventually got a number in each constellation in order of their location in the sky or, more precisely, their right ascension. (The system of sky coordinates that includes right ascension was discussed in Earth, Moon, and Sky.) In this system, Betelgeuse is called 58 Orionis and 61 Cygni is the 61st star in the constellation of Cygnus, the swan.

It gets worse. As astronomers began to understand more and more about stars, they drew up a series of specialized star catalogs, and fans of those catalogs began calling stars by their catalog numbers. If you look at Appendix I—our list of the nearest stars (many of which are much too faint to get an ancient name, Bayer letter, or Flamsteed number)—you will see references to some of these catalogs. An example is a set of stars labeled with a BD number, for "Bonner Durchmusterung." This was a mammoth catalog of over 324,000 stars in a series of zones in the sky, organized at the Bonn Observatory in the 1850s and 1860s. Keep in mind that this catalog was made before photography or computers came into use, so the position of each star had to be measured (at least twice) by eye, a daunting undertaking.

There is also a completely different system for keeping track of stars whose luminosity varies, and another for stars that brighten explosively at unpredictable times. Astronomers have gotten used to the many different star-naming systems, but students often find them bewildering and wish astronomers would settle on one. Don't hold your breath: in astronomy, as in many fields of human thought, tradition holds a powerful attraction. Still, with high-speed computer databases to aid human memory, names may become less and less necessary. Today's astronomers often refer to stars by their precise locations in the sky rather than by their names or various catalog numbers.

The Nearest Stars

No known star (other than the Sun) is within 1 light-year or even 1 parsec of Earth. The stellar neighbors nearest the Sun are three stars in the constellation of Centaurus. To the unaided eye, the brightest of these three stars is Alpha Centauri, which is only 30° from the south celestial pole and hence not visible from the mainland United States. Alpha Centauri itself is a binary star—two stars in mutual revolution—too close together to be distinguished without a telescope. These two stars are 4.4 light-years from us. Nearby is a third faint star, known as Proxima Centauri. Proxima, with a distance of 4.3 light-years, is slightly closer to us than the other two stars. If Proxima Centauri is part of a triple star system with the binary Alpha Centauri, as seems likely, then its orbital period may be longer than 500,000 years.

Proxima Centauri is an example of the most common type of star, and our most common type of stellar neighbor (as we saw in Stars: A Celestial Census.) Low-mass red M dwarfs make up about 70% of all stars and dominate the census of stars within 10

parsecs (33 light-years) of the Sun. For example, a recent survey of the solar neighborhood counted 357 stars and brown dwarfs within 10 parsecs, and 248 of these are red dwarfs. Yet, if you wanted to see an M dwarf with your naked eye, you would be out of luck. These stars only produce a fraction of the Sun's light, and nearly all of them require a telescope to be detected.

The nearest star visible without a telescope from most of the United States is the brightest appearing of all the stars, Sirius, which has a distance of a little more than 8 light-years. It too is a binary system, composed of a faint white dwarf orbiting a bluish-white, main-sequence star. It is an interesting coincidence of numbers that light reaches us from the Sun in about 8 minutes and from the next brightest star in the sky in about 8 years.

Example 19.2: Calculating the Diameter of the Sun

For nearby stars, we can measure the apparent shift in their positions as Earth orbits the Sun. We wrote earlier that an object must be 206,265 AU distant to have a parallax of one second of arc. This must seem like a very strange number, but you can figure out why this is the right value. We will start by estimating the diameter of the Sun and then apply the same idea to a star with a parallax of 1 arcsecond. Make a sketch that has a round circle to represent the Sun, place Earth some distance away, and put an observer on it. Draw two lines from the point where the observer is standing, one to each side of the Sun. Sketch a circle centered at Earth with its circumference passing through the center of the Sun. Now think about proportions.

The Sun spans about half a degree on the sky. A full circle has 360°. The circumference of the circle centered on Earth and passing through the Sun is given by:

$$\text{circumference} = 2\pi \times 93{,}000{,}000 \text{ miles}$$

Then, the following two ratios are equal:

$$\frac{0.5°}{360°} = \frac{\text{diameter of Sun}}{2\pi \times 93{,}000{,}000}$$

Calculate the diameter of the Sun. How does your answer compare to the actual diameter?

Solution

To solve for the diameter of the Sun, we can evaluate the expression above.

$$\begin{aligned} \text{diameter of the sun} &= \tfrac{0.5°}{360°} \times 2\pi \times 93{,}000{,}000 \text{ miles} \\ &= 811{,}577 \text{ miles} \end{aligned}$$

This is very close to the true value of about 848,000 miles.

Exercise 12.2.1

Now apply this idea to calculating the distance to a star that has a parallax of 1 arcsec. Draw a picture similar to the one we suggested above and calculate the distance in AU. (Hint: Remember that the parallax angle is defined by 1 AU, not 2 AU, and that 3600 arcseconds = 1 degree.)

Answer available at the end of this section.

Measuring Parallaxes in Space

The measurements of stellar parallax were revolutionized by the launch of the spacecraft Hipparcos in 1989, which measured distances for thousands of stars out to about 300 light-years with an accuracy of 10 to 20% (see Figure 19.8 and the feature on Parallax and Space Astronomy). However, even 300 light-years are less than 1% the size of our Galaxy's main disk.

In December 2013, the successor to Hipparcos, named *Gaia*, was launched by the European Space Agency. *Gaia* is measuring the position and distances to almost one billion stars with an accuracy of a few millionths of an arcsecond. *Gaia's* distance limit extends well beyond Hipparcos, studying stars out to 30,000 light-years (100 times farther than Hipparcos, covering nearly 1/3 of

the galactic disk). *Gaia* is also able to measure proper motions[2] for thousands of stars in the halo of the Milky Way—something that can only be done for the brightest stars right now. At the end of *Gaia's* mission, we will not only have a three-dimensional map of a large fraction of our own Milky Way Galaxy, but we will also have a strong link in the chain of cosmic distances that we are discussing in this chapter. Yet, to extend this chain beyond *Gaia's* reach and explore distances to nearby galaxies, we need some completely new techniques.

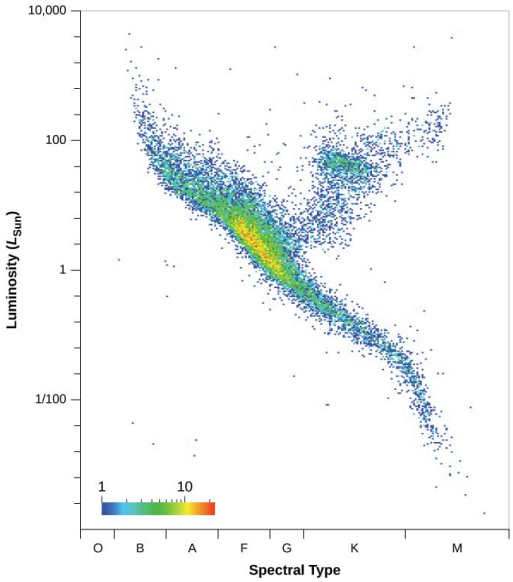


Figure 19.8 : H–R Diagram of Stars Measured by *Gaia* and Hipparcos. This plot includes 16,631 stars for which the parallaxes have an accuracy of 10% or better. The colors indicate the numbers of stars at each point of the diagram, with red corresponding to the largest number and blue to the lowest. Luminosity is plotted along the vertical axis, with luminosity increasing upward. An infrared color is plotted as a proxy for temperature, with temperature decreasing to the right. Most of the data points are distributed along the diagonal running from the top left corner (high luminosity, high temperature) to the bottom right (low temperature, low luminosity). These are main sequence stars. The large clump of data points above the main sequence on the right side of the diagram is composed of red giant stars. (credit: modification of work by the European Space Agency)

Making Connections: Parallax and Space Astronomy

One of the most difficult things about precisely measuring the tiny angles of parallax shifts from Earth is that you have to observe the stars through our planet's atmosphere. As we saw in Astronomical Instruments, the effect of the atmosphere is to spread out the points of starlight into fuzzy disks, making exact measurements of their positions more difficult. Astronomers had long dreamed of being able to measure parallaxes from space, and two orbiting observatories have now turned this dream into reality.

The name of the Hipparcos satellite, launched in 1989 by the European Space Agency, is both an abbreviation for High Precision Parallax Collecting Satellite and a tribute to Hipparchus, the pioneering Greek astronomer whose work we discussed in the Observing the Sky: The Birth of Astronomy. The satellite was designed to make the most accurate parallax measurements in history, from 36,000 kilometers above Earth. However, its onboard rocket motor failed to fire, which meant it did not get the needed boost to reach the desired altitude. Hipparcos ended up spending its 4-year life in an elliptical orbit that varied from 500 to 36,000 kilometers high. In this orbit, the satellite plunged into Earth's radiation belts every 5 hours or so, which finally took its toll on the solar panels that provided energy to power the instruments.

Nevertheless, the mission was successful, resulting in two catalogs. One gives positions of 120,000 stars to an accuracy of one-thousandth of an arcsecond—about the diameter of a golf ball in New York as viewed from Europe. The second catalog contains information for more than a million stars, whose positions have been measured to thirty-thousandths of an arcsecond. We now have accurate parallax measurements of stars out to distances of about 300 light-years. (With ground-based telescopes, accurate measurements were feasible out to only about 60 light-years.)

In order to build on the success of Hipparcos, in 2013, the European Space Agency launched a new satellite called *Gaia*. Because *Gaia* carries larger telescopes than Hipparcos, it can observe fainter stars and measure their positions 200 times more accurately. The main goal of the Gaia mission is to make an accurate three-dimensional map of that portion of the Galaxy within about 30,000 light-years by observing 1 billion stars 70 times each, measuring their positions and hence their parallaxes as well as their brightnesses.

For a long time, the measurement of parallaxes and accurate stellar positions was a backwater of astronomical research—mainly because the accuracy of measurements did not improve much for about 100 years. However, the ability to make measurements from space has revolutionized this field of astronomy and will continue to provide a critical link in our chain of cosmic distances.

Link to Learning

The European Space Agency (ESA) maintains a Gaia mission website where you can learn more about the Gaia mission and to get the latest news on *Gaia* observations.

To learn more about Hipparcos, explore this European Space Agency webpage with an ESA vodcast *Charting the Galaxy—from Hipparcos to Gaia*.

Footnotes

- 1To have some basis for comparison, the dwarf planet Pluto orbits at an average distance of 40 AU from the Sun, and the dwarf planet Eris is currently roughly 96 AU from the Sun.
- 2Proper motion (as discussed in Analyzing Starlight, is the motion of a star across the sky (perpendicular to our line of sight.)

Answers

Exercise 12.2.1

$$\begin{aligned} 1\,\text{light-year} &= 9.46 \times 10^{12}\,\text{km} \\ &= 9.46 \times 10^{12}\,\text{km} \times \tfrac{1\,\text{Earth diameter}}{12{,}700\,\text{km}} \\ &= 7.45 \times 10^{8}\,\text{Earth diameters} \end{aligned}$$

That means that 1 light-year is about 745 million times the diameter of Earth.

Exercise 12.2.1

206,265 AU

This page titled 12.2: Surveying the Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 19.3: Surveying the Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.3: Colors of Stars

Learning Objectives

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

- Compare the relative temperatures of stars based on their colors
- Understand how astronomers use color indexes to measure the temperatures of stars

Look at the beautiful picture of the stars in the Sagittarius Star Cloud shown in Figure 17.3. The stars show a multitude of colors, including red, orange, yellow, white, and blue. As we have seen, stars are not all the same color because they do not all have identical temperatures. To define *color* precisely, astronomers have devised quantitative methods for characterizing the color of a star and then using those colors to determine stellar temperatures. In the chapters that follow, we will provide the temperature of the stars we are describing, and this section tells you how those temperatures are determined from the colors of light the stars give off.

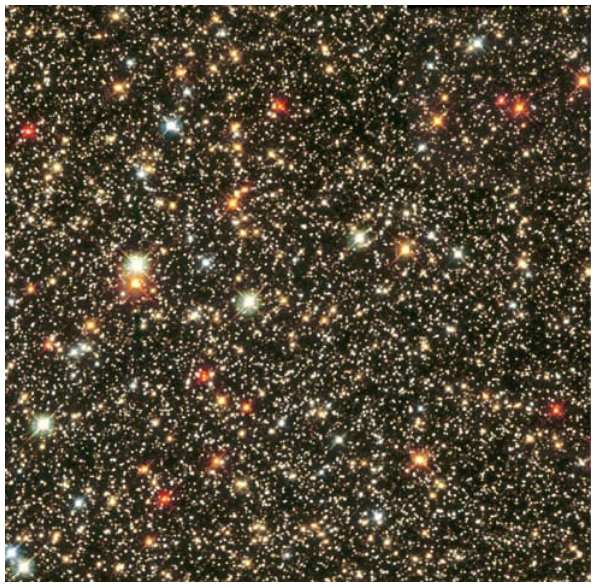


Figure 17.3 : Sagittarius Star Cloud. This image, which was taken by the Hubble Space Telescope, shows stars in the direction toward the center of the Milky Way Galaxy. The bright stars glitter like colored jewels on a black velvet background. The color of a star indicates its temperature. Blue-white stars are much hotter than the Sun, whereas red stars are cooler. On average, the stars in this field are at a distance of about 25,000 light-years (which means it takes light 25,000 years to traverse the distance from them to us) and the width of the field is about 13.3 light-years. (credit: Hubble Heritage Team (AURA/STScI/NASA))

Color and Temperature

As we learned in The Electromagnetic Spectrum section, Wien's law relates stellar color to stellar temperature. Blue colors dominate the visible light output of very hot stars (with much additional radiation in the ultraviolet). On the other hand, cool stars emit most of their visible light energy at red wavelengths (with more radiation coming off in the infrared) (Table 17.1). The color of a star therefore provides a measure of its intrinsic or true surface temperature (apart from the effects of reddening by interstellar dust, which will be discussed in Between the Stars: Gas and Dust in Space). Color does not depend on the distance to the object. This should be familiar to you from everyday experience. The color of a traffic signal, for example, appears the same no matter how far away it is. If we could somehow take a star, observe it, and then move it much farther away, its apparent brightness (magnitude) would change. But this change in brightness is the same for all wavelengths, and so its color would remain the same.

Table 17.1: Example Star Colors and Corresponding Approximate Temperatures

Star Color	Approximate Temperature	Example
Blue	25,000 K	Spica
White	10,000 K	Vega
Yellow	6000 K	Sun
Orange	4000 K	Aldebaran
Red	3000 K	Betelgeuse

Link to Learning

Go to this interactive simulation from the University of Colorado to see the color of a star changing as the temperature is changed.

The hottest stars have temperatures of over 40,000 K, and the coolest stars have temperatures of about 2000 K. Our Sun's surface temperature is about 6000 K; its peak wavelength color is a slightly greenish-yellow. In space, the Sun would look white, shining with about equal amounts of reddish and bluish wavelengths of light. It looks somewhat yellow as seen from Earth's surface because our planet's nitrogen molecules scatter some of the shorter (i.e., blue) wavelengths out of the beams of sunlight that reach us, leaving more long wavelength light behind. This also explains why the sky is blue: the blue sky is sunlight scattered by Earth's atmosphere.

Color Indices

In order to specify the exact color of a star, astronomers normally measure a star's apparent brightness through filters, each of which transmits only the light from a particular narrow band of wavelengths (colors). A crude example of a filter in everyday life is a green-colored, plastic, soft drink bottle, which, when held in front of your eyes, lets only the green colors of light through.

One commonly used set of filters in astronomy measures stellar brightness at three wavelengths corresponding to ultraviolet, blue, and yellow light. The filters are named: U (ultraviolet), B (blue), and V (visual, for yellow). These filters transmit light near the wavelengths of 360 nanometers (nm), 420 nm, and 540 nm, respectively. The brightness measured through each filter is usually expressed in magnitudes. The difference between any two of these magnitudes—say, between the blue and the visual magnitudes (B–V)—is called a color index.

By agreement among astronomers, the ultraviolet, blue, and visual magnitudes of the UBV system are adjusted to give a color index of 0 to a star with a surface temperature of about 10,000 K, such as Vega. The B–V color indexes of stars range from –0.4 for the bluest stars, with temperatures of about 40,000 K, to +2.0 for the reddest stars, with temperatures of about 2000 K. The B–V index for the Sun is about +0.65. Note that, by convention, the B–V index is always the "bluer" minus the "redder" color.

Why use a color index if it ultimately implies temperature? Because the brightness of a star through a filter is what astronomers actually measure, and we are always more comfortable when our statements have to do with measurable quantities.

This page titled 12.3: Colors of Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 17.3: Colors of Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.4: The Spectra of Stars (and Brown Dwarfs)

Learning Objectives

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

- Describe how astronomers use spectral classes to characterize stars
- Explain the difference between a star and a brown dwarf

Measuring colors is only one way of analyzing starlight. Another way is to use a spectrograph to spread out the light into a spectrum (see the Radiation and Spectra and the Astronomical Instruments chapters). In 1814, the German physicist Joseph Fraunhofer observed that the spectrum of the Sun shows dark lines crossing a continuous band of colors. In the 1860s, English astronomers Sir William Huggins and Lady Margaret Huggins (Figure 17.4) succeeded in identifying some of the lines in stellar spectra as those of known elements on Earth, showing that the same chemical elements found in the Sun and planets exist in the stars. Since then, astronomers have worked hard to perfect experimental techniques for obtaining and measuring spectra, and they have developed a theoretical understanding of what can be learned from spectra. Today, spectroscopic analysis is one of the cornerstones of astronomical research.

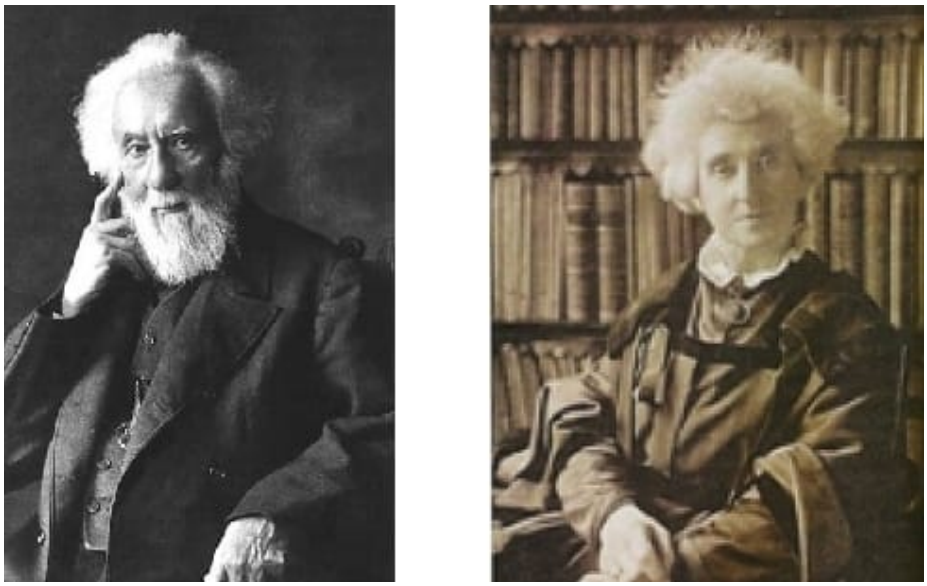


Figure 17.4 : William Huggins (1824–1910) and Margaret Huggins (1848–1915). William and Margaret Huggins were the first to identify the lines in the spectrum of a star other than the Sun; they also took the first spectrogram, or photograph of a stellar spectrum.

Formation of Stellar Spectra

When the spectra of different stars were first observed, astronomers found that they were not all identical. Since the dark lines are produced by the chemical elements present in the stars, astronomers first thought that the spectra differ from one another because stars are not all made of the same chemical elements. This hypothesis turned out to be wrong. *The primary reason that stellar spectra look different is because the stars have different temperatures*. Most stars have nearly the same composition as the Sun, with only a few exceptions.

Hydrogen, for example, is by far the most abundant element in most stars. However, lines of hydrogen are not seen in the spectra of the hottest and the coolest stars. In the atmospheres of the hottest stars, hydrogen atoms are completely ionized. Because the electron and the proton are separated, ionized hydrogen cannot produce absorption lines. (Recall from the Formation of Spectral Lines section, the lines are the result of electrons in orbit around a nucleus changing energy levels.)

In the atmospheres of the coolest stars, hydrogen atoms have their electrons attached and can switch energy levels to produce

lines. However, practically all of the hydrogen atoms are in the lowest energy state (unexcited) in these stars and thus can absorb only those photons able to lift an electron from that first energy level to a higher level. Photons with enough energy to do this lie in the ultraviolet part of the electromagnetic spectrum, and there are very few ultraviolet photons in the radiation from a cool star. What this means is that if you observe the spectrum of a very hot or very cool star with a typical telescope on the surface of Earth, the most common element in that star, hydrogen, will show very weak spectral lines or none at all.

The hydrogen lines in the visible part of the spectrum (called *Balmer lines*) are strongest in stars with intermediate temperatures—not too hot and not too cold. Calculations show that the optimum temperature for producing visible hydrogen lines is about 10,000 K. At this temperature, an appreciable number of hydrogen atoms are excited to the second energy level. They can then absorb additional photons, rise to still-higher levels of excitation, and produce a dark absorption line. Similarly, every other chemical element, in each of its possible stages of ionization, has a characteristic temperature at which it is most effective in producing absorption lines in any particular part of the spectrum.

Classification of Stellar Spectra

Astronomers use the patterns of lines observed in stellar spectra to sort stars into a spectral class. Because a star's temperature determines which absorption lines are present in its spectrum, these spectral classes are a measure of its surface temperature. There are seven standard spectral classes. From hottest to coldest, these seven spectral classes are designated O, B, A, F, G, K, and M. Recently, astronomers have added three additional classes for even cooler objects—L, T, and Y.

At this point, you may be looking at these letters with wonder and asking yourself why astronomers didn't call the spectral types A, B, C, and so on. You will see, as we tell you the history, that it's an instance where tradition won out over common sense.

In the 1880s, Williamina Fleming devised a system to classify stars based on the strength of hydrogen absorption lines. Spectra with the strongest lines were classified as "A" stars, the next strongest "B," and so on down the alphabet to "O" stars, in which the hydrogen lines were very weak. But we saw above that hydrogen lines alone are not a good indicator for classifying stars, since their lines disappear from the visible light spectrum when the stars get too hot or too cold.

In the 1890s, Annie Jump Cannon revised this classification system, focusing on just a few letters from the original system: A, B, F, G, K, M, and O. Instead of starting over, Cannon also rearranged the existing classes—in order of decreasing temperature—into the sequence we have learned: O, B, A, F, G, K, M. As you can read in the feature on Annie Cannon: Classifier of the Stars in this chapter, she classified around 500,000 stars over her lifetime, classifying up to three stars per minute by looking at the stellar spectra.

Link to Learning

For a deep dive into spectral types, explore the interactive project at the Sloan Digital Sky Survey in which you can practice classifying stars yourself.

To help astronomers remember this crazy order of letters, Cannon created a mnemonic, "Oh Be A Fine Girl, Kiss Me." (If you prefer, you can easily substitute "Guy" for "Girl.") Other mnemonics, which we hope will not be relevant for you, include "Oh Brother, Astronomers Frequently Give Killer Midterms" and "Oh Boy, An F Grade Kills Me!" With the new L, T, and Y spectral classes, the mnemonic might be expanded to "Oh Be A Fine Girl (Guy), Kiss Me Like That, Yo!"

Each of these spectral classes, except possibly for the Y class which is still being defined, is further subdivided into 10 subclasses designated by the numbers 0 through 9. A B0 star is the hottest type of B star; a B9 star is the coolest type of B star and is only slightly hotter than an A0 star.

And just one more item of vocabulary: for historical reasons, astronomers call all the elements heavier than helium *metals*, even though most of them do not show metallic properties. (If you are getting annoyed at the peculiar jargon that astronomers use, just bear in mind that every field of human activity tends to develop its own specialized vocabulary. Just try reading a credit card or social media agreement form these days without training in law!)

Let's take a look at some of the details of how the spectra of the stars change with temperature. (It is these details that allowed Annie Cannon to identify the spectral types of stars as quickly as three per minute!) As Figure 17.5 shows, in the hottest O stars (those with temperatures over 28,000 K), only lines of ionized helium and highly ionized atoms of other elements are conspicuous. Hydrogen lines are strongest in A stars with atmospheric temperatures of about 10,000 K. Ionized metals provide

the most conspicuous lines in stars with temperatures from 6000 to 7500 K (spectral type F). In the coolest M stars (below 3500 K), absorption bands of titanium oxide and other molecules are very strong. By the way, the spectral class assigned to the Sun is G2. The sequence of spectral classes is summarized in Table 17.2.

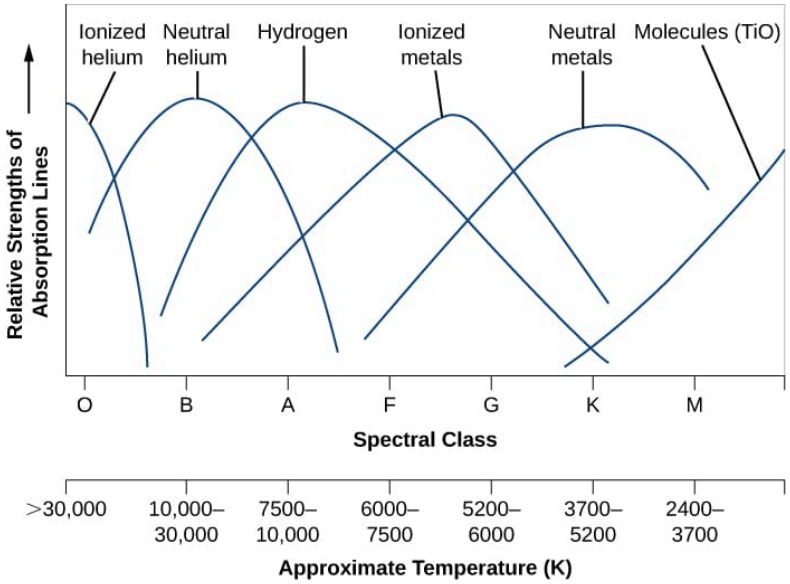


Figure 17.5 : Absorption Lines in Stars of Different Temperatures. This graph shows the strengths of absorption lines of different chemical species (atoms, ions, molecules) as we move from hot (left) to cool (right) stars. The sequence of spectral types is also shown.

Link to Learning

Use the Spectrum Explorer to see how each element's absorption lines change with temperature.

Table 17.2: Spectral Classes for Stars

Spectral Class	Color	Approximate Temperature (K)	Principal Features	Examples
O	Blue	> 30,000	Neutral and ionized helium lines, weak hydrogen lines	10 Lacertae
B	Blue-white	10,000–30,000	Neutral helium lines, strong hydrogen lines	Rigel, Spica
A	White	7500–10,000	Strongest hydrogen lines, weak ionized calcium lines, weak ionized metal (e.g., iron, magnesium) lines	Sirius, Vega
F	Yellow-white	6000–7500	Strong hydrogen lines, strong ionized calcium lines, weak sodium lines, many ionized metal lines	Canopus, Procyon
G	Yellow	5200–6000	Weaker hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of ionized and neutral metals	Sun, Capella
K	Orange	3700–5200	Very weak hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of neutral metals	Arcturus, Aldebaran
M	Red	2400–3700	Strong lines of neutral metals and molecular bands of titanium oxide dominate	Betelgeuse, Antares
L	Red	1300–2400	Metal hydride lines, alkali metal lines (e.g., sodium, potassium, rubidium)	Teide 1
T	Magenta	700–1300	Methane lines	Gliese 229B

Spectral Class	Color	Approximate Temperature (K)	Principal Features	Examples
Y	Infrared[1]	< 700	Ammonia lines	WISE 1828+2650

To see how spectral classification works, let's use Figure 17.5. Suppose you have a spectrum in which the hydrogen lines are about half as strong as those seen in an A star. Looking at the lines in our figure, you see that the star could be either a B star or a G star. But if the spectrum also contains helium lines, then it is a B star, whereas if it contains lines of ionized iron and other metals, it must be a G star.

If you look at Figure 17.6, you can see that you, too, could assign a spectral class to a star whose type was not already known. All you have to do is match the pattern of spectral lines to a standard star (like the ones shown in the figure) whose type has already been determined.

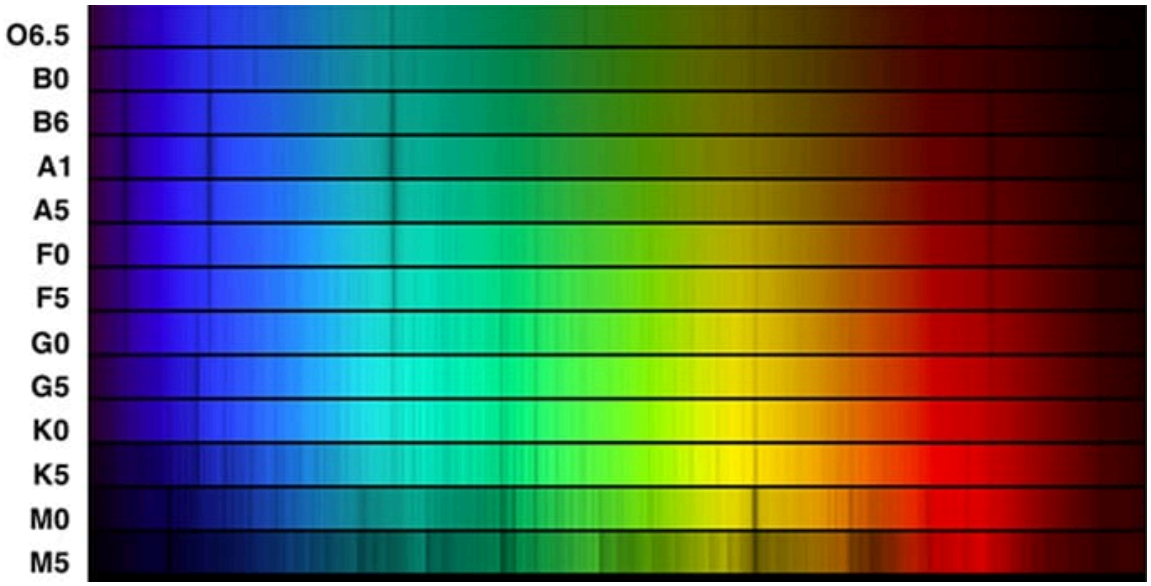


Figure 17.6 : Spectra of Stars with Different Spectral Classes. This image compares the spectra of the different spectral classes. The spectral class assigned to each of these stellar spectra is listed at the left of the picture. The strongest four lines seen at spectral type A1 (one in the red, one in the blue-green, and two in the blue) are Balmer lines of hydrogen. Note how these lines weaken at both higher and lower temperatures, as Figure 17.5 also indicates. The strong pair of closely spaced lines in the yellow in the cool stars is due to neutral sodium (one of the neutral metals in Figure 17.5). (Credit: modification of work by NOAO/AURA/NSF)

Both colors and spectral classes can be used to estimate the temperature of a star. Spectra are harder to measure because the light has to be bright enough to be spread out into all colors of the rainbow, and detectors must be sensitive enough to respond to individual wavelengths. In order to measure colors, the detectors need only respond to the many wavelengths that pass simultaneously through the colored filters that have been chosen—that is, to *all* the blue light or *all* the yellow-green light.

Voyagers in Astronomy: Annie Cannon: Classifier of the Stars

Annie Jump Cannon was born in Delaware in 1863 (Figure 17.7). In 1880, she went to Wellesley College, one of the new breed of US colleges opening up to educate young women. Wellesley, only 5 years old at the time, had the second student physics lab in the country and provided excellent training in basic science. After college, Cannon spent a decade with her parents but was very dissatisfied, longing to do scientific work. After her mother's death in 1893, she returned to Wellesley as a teaching assistant and also to take courses at Radcliffe, the women's college associated with Harvard.



Figure 17.7 : Annie Jump Cannon (1863–1941). Cannon is well-known for her classifications of stellar spectra. (credit: modification of work by Smithsonian Institution)

In the late 1800s, the director of the Harvard Observatory, Edward C. Pickering, needed lots of help with his ambitious program of classifying stellar spectra. The basis for these studies was a monumental collection of nearly a million photographic spectra of stars, obtained from many years of observations made at Harvard College Observatory in Massachusetts as well as at its remote observing stations in South America and South Africa. Pickering quickly discovered that educated young women could be hired as assistants for one-third or one-fourth the salary paid to men, and they would often put up with working conditions and repetitive tasks that men with the same education would not tolerate. These women became known as the Harvard Computers. (We should emphasize that astronomers were not alone in reaching such conclusions about the relatively new idea of upper-class, educated women working outside the home: women were exploited and undervalued in many fields. This is a legacy from which our society is just beginning to emerge.)

Cannon was hired by Pickering as one of the "computers" to help with the classification of spectra. She became so good at it that she could visually examine and determine the spectral types of several hundred stars per hour (dictating her conclusions to an assistant). She made many discoveries while investigating the Harvard photographic plates, including 300 variable stars (stars whose luminosity changes periodically). But her main legacy is a marvelous catalog of spectral types for hundreds of thousands of stars, which served as a foundation for much of twentieth-century astronomy.

In 1911, a visiting committee of astronomers reported that "she is the one person in the world who can do this work quickly and accurately" and urged Harvard to give Cannon an official appointment in keeping with her skill and renown. Not until 1938, however, did Harvard appoint her an astronomer at the university; she was then 75 years old.

Cannon received the first honorary degree Oxford awarded to a woman, and she became the first woman to be elected an officer of the American Astronomical Society, the main professional organization of astronomers in the US. She generously donated the money from one of the major prizes she had won to found a special award for women in astronomy, now known as the Annie Jump Cannon Prize. True to form, she continued classifying stellar spectra almost to the very end of her life in 1941.

Spectral Classes L, T, and Y

The scheme devised by Cannon worked well until 1988, when astronomers began to discover objects even cooler than M9-type stars. We use the word *object* because many of the new discoveries are not true stars. A star is defined as an object that during some part of its lifetime derives 100% of its energy from the same process that makes the Sun shine—the fusion of hydrogen nuclei (protons) into helium. Objects with masses less than about 7.5% of the mass of our Sun (about 0.075 M_{Sun}) do not become hot enough for hydrogen fusion to take place. Even before the first such "failed star" was found, this class of objects, with masses intermediate between stars and planets, was given the name brown dwarfs.

Brown dwarfs are very difficult to observe because they are extremely faint and cool, and they put out most of their light in the infrared part of the spectrum. It was only after the construction of very large telescopes, like the Keck telescopes in Hawaii, and the development of very sensitive infrared detectors, that the search for brown dwarfs succeeded. The first brown dwarf was discovered in 1988, and, as of the summer of 2015, there are more than 2200 known brown dwarfs.

Initially, brown dwarfs were given spectral classes like $M10^{+}$ or "much cooler than M9," but so many are now known that it

is possible to begin assigning spectral types. The hottest brown dwarfs are given types L0–L9 (temperatures in the range 2400–1300 K), whereas still cooler (1300–700 K) objects are given types T0–T9 (see Figure 17.8). In class L brown dwarfs, the lines of titanium oxide, which are strong in M stars, have disappeared. This is because the L dwarfs are so cool that atoms and molecules can gather together into dust particles in their atmospheres; the titanium is locked up in the dust grains rather than being available to form molecules of titanium oxide. Lines of steam (hot water vapor) are present, along with lines of carbon monoxide and neutral sodium, potassium, cesium, and rubidium. Methane (CH_4) lines are strong in class-T brown dwarfs, as methane exists in the atmosphere of the giant planets in our own solar system.

In 2009, astronomers discovered ultra-cool brown dwarfs with temperatures of 500–600 K. These objects exhibited absorption lines due to ammonia (NH_3), which are not seen in T dwarfs. A new spectral class, Y, was created for these objects. As of 2021, over 50 brown dwarfs belonging to spectral class Y have been discovered, some with temperatures comparable to that of the human body (about 300 K).[2]

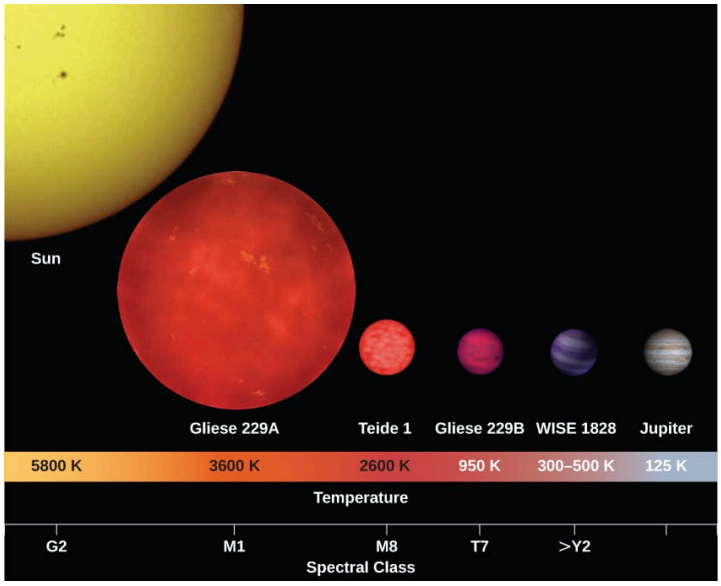


Figure 17.8 : Brown Dwarfs. This illustration shows the sizes and surface temperatures of brown dwarfs Teide 1, Gliese 229B, and WISE1828 in relation to the Sun, a red dwarf star (Gliese 229A), and Jupiter. (credit: modification of work by MPIA/V. Joergens)

Most brown dwarfs start out with atmospheric temperatures and spectra like those of true stars with spectral classes of M6.5 and later, even though the brown dwarfs are not hot and dense enough in their interiors to fuse hydrogen. In fact, the spectra of brown dwarfs and true stars are so similar from spectral types late M through L that it is not possible to distinguish the two types of objects based on spectra alone. An independent measure of mass is required to determine whether a specific object is a brown dwarf or a very low mass star. Since brown dwarfs cool steadily throughout their lifetimes, the spectral type of a given brown dwarf changes with time over a billion years or more from late M through L, T, and Y spectral types.

Low-Mass Brown Dwarfs vs. High-Mass Planets

An interesting property of brown dwarfs is that they are all about the same radius as Jupiter, regardless of their masses. Amazingly, this covers a range of masses from about 13 to 80 times the mass of Jupiter (M_J). This can make distinguishing a low-mass brown dwarf from a high-mass planet very difficult.

So, what is the difference between a low-mass brown dwarf and a high-mass planet? The International Astronomical Union considers the distinctive feature to be *deuterium fusion*. Although brown dwarfs do not sustain regular (proton-proton) hydrogen fusion, they are capable of fusing deuterium (a rare form of hydrogen with one proton and one neutron in its nucleus). The fusion of deuterium can happen at a lower temperature than the fusion of hydrogen. If an object has enough mass to fuse deuterium (about 13 M_J or 0.012 M_{Sun}), it is a brown dwarf. Objects with less than 13 M_J do not fuse deuterium and are usually considered planets.

Footnotes

- 1Absorption by sodium and potassium atoms makes Y dwarfs appear a bit less red than L dwarfs.
- 2The record-holder for the coldest Type Y dwarf is WISE 0855-0714, at around 250 K.

This page titled 12.4: The Spectra of Stars (and Brown Dwarfs) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 17.4: The Spectra of Stars (and Brown Dwarfs) by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion

Learning Objectives

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

- Understand how astronomers can learn about a star's radius and composition by studying its spectrum
- Explain how astronomers can measure the motion and rotation of a star using the Doppler effect
- Describe the proper motion of a star and how it relates to a star's space velocity

Analyzing the spectrum of a star can teach us all kinds of things in addition to its temperature. We can measure its detailed chemical composition as well as the pressure in its atmosphere. From the pressure, we get clues about its size. We can also measure its motion toward or away from us and estimate its rotation.

Clues to the Size of a Star

As we shall see in The Stars: A Celestial Census, stars come in a wide variety of sizes. At some periods in their lives, stars can expand to enormous dimensions. Stars of such exaggerated size are called giants. Luckily for the astronomer, stellar spectra can be used to distinguish giants from run-of-the-mill stars (such as our Sun).

Suppose you want to determine whether a star is a giant. A giant star has a large, extended photosphere. Because it is so large, a giant star's atoms are spread over a great volume, which means that the density of particles in the star's photosphere is low. As a result, the pressure in a giant star's photosphere is also low. This low pressure affects the spectrum in two ways. First, a star with a lower-pressure photosphere shows narrower spectral lines than a star of the same temperature with a higher-pressure photosphere (Figure 17.9). The difference is large enough that careful study of spectra can tell which of two stars at the same temperature has a higher pressure (and is thus more compressed) and which has a lower pressure (and thus must be extended). This effect is due to collisions between particles in the star's photosphere—more collisions lead to broader spectral lines. Collisions will, of course, be more frequent in a higher-density environment. Think about it like traffic—collisions are much more likely during rush hour, when the density of cars is high.

Second, more atoms are ionized in a giant star than in a star like the Sun with the same temperature. The ionization of atoms in a star's outer layers is caused mainly by photons, and the amount of energy carried by photons is determined by temperature. But how long atoms *stay* ionized depends in part on pressure. Compared with what happens in the Sun (with its relatively dense photosphere), ionized atoms in a giant star's photosphere are less likely to pass close enough to electrons to interact and combine with one or more of them, thereby becoming neutral again. Ionized atoms, as we discussed earlier, have different spectra from atoms that are neutral.

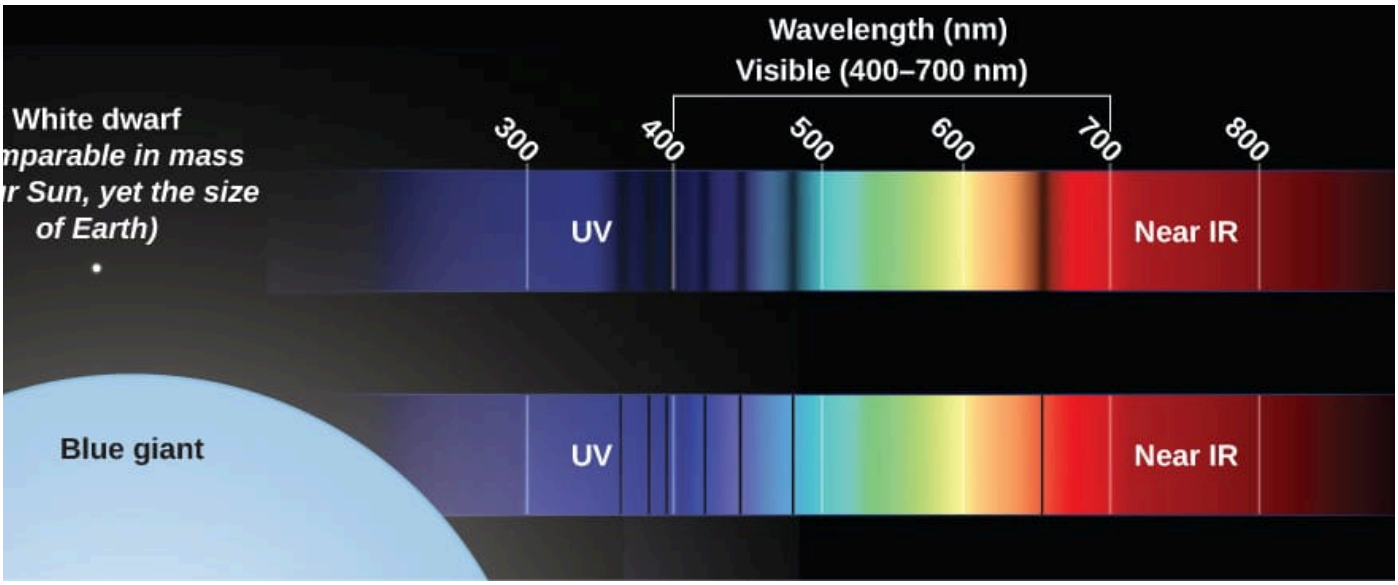


Figure 17.9 : Spectral Lines. This figure illustrates one difference in the spectral lines from stars of the same temperature but different pressures. A giant star with a very-low-pressure photosphere shows very narrow spectral lines (bottom), whereas a smaller star with a higher-pressure photosphere shows much broader spectral lines (top). (credit: modification of work by NASA, ESA, A. Field, and J. Kalirai (STScI))

Abundances of the Elements

Absorption lines of a majority of the known chemical elements have now been identified in the spectra of the Sun and stars. If we see lines of iron in a star's spectrum, for example, then we know immediately that the star must contain iron.

Note that the *absence* of an element's spectral lines does not necessarily mean that the element itself is absent. As we saw, the temperature and pressure in a star's atmosphere will determine what types of atoms are able to produce absorption lines. Only if the physical conditions in a star's photosphere are such that lines of an element *should* (according to calculations) be there can we conclude that the absence of observable spectral lines implies low abundance of the element.

Suppose two stars have identical temperatures and pressures, but the lines of, say, sodium are stronger in one than in the other. Stronger lines mean that there are more atoms in the stellar photosphere absorbing light. Therefore, we know immediately that the star with stronger sodium lines contains more sodium. Complex calculations are required to determine exactly how much more, but those calculations can be done for any element observed in any star with any temperature and pressure.

Of course, astronomy textbooks such as ours always make these things sound a bit easier than they really are. If you look at the stellar spectra such as those in Figure 17.6, you may get some feeling for how hard it is to decode all of the information contained in the thousands of absorption lines. First of all, it has taken many years of careful laboratory work on Earth to determine the precise wavelengths at which hot gases of each element have their spectral lines. Long books and computer databases have been compiled to show the lines of each element that can be seen at each temperature. Second, stellar spectra usually have many lines from a number of elements, and we must be careful to sort them out correctly. Sometimes nature is unhelpful, and lines of different elements have identical wavelengths, thereby adding to the confusion. And third, as we saw in the chapter on Radiation and Spectra, the motion of the star can change the observed wavelength of each of the lines. So, the observed wavelengths may not match laboratory measurements exactly. In practice, analyzing stellar spectra is a demanding, sometimes frustrating task that requires both training and skill.

Studies of stellar spectra have shown that hydrogen makes up about three-quarters of the mass of most stars. Helium is the second-most abundant element, making up almost a quarter of a star's mass. Together, hydrogen and helium make up from 96 to 99% of the mass; in some stars, they amount to more than 99.9%. Among the 4% or less of "heavy elements," oxygen, carbon, neon, iron, nitrogen, silicon, magnesium, and sulfur are among the most abundant. Generally, but not invariably, the elements of lower atomic weight are more abundant than those of higher atomic weight.

Take a careful look at the list of elements in the preceding paragraph. Two of the most abundant are hydrogen and oxygen (which make up water); add carbon and nitrogen and you are starting to write the prescription for the chemistry of an astronomy student. We are made of elements that are common in the universe—just mixed together in a far more sophisticated form (and a much

cooler environment) than in a star.

As we mentioned in The Spectra of Stars (and Brown Dwarfs) section, astronomers use the term “metals” to refer to all elements heavier than hydrogen and helium. The fraction of a star’s mass that is composed of these elements is referred to as the star’s *metallicity*. The metallicity of the Sun, for example, is 0.02, since 2% of the Sun’s mass is made of elements heavier than helium.

Appendix K lists how common each element is in the universe (compared to hydrogen); these estimates are based primarily on investigation of the Sun, which is a typical star. Some very rare elements, however, have not been detected in the Sun. Estimates of the amounts of these elements in the universe are based on laboratory measurements of their abundance in primitive meteorites, which are considered representative of unaltered material condensed from the solar nebula (see the Cosmic Samples and the Origin of the Solar System chapter).

Radial Velocity

When we measure the spectrum of a star, we determine the wavelength of each of its lines. If the star is not moving with respect to the Sun, then the wavelength corresponding to each element will be the same as those we measure in a laboratory here on Earth. But if stars are moving toward or away from us, we must consider the *Doppler effect* (see The Doppler Effect section). We should see all the spectral lines of moving stars shifted toward the red end of the spectrum if the star is moving away from us, or toward the blue (violet) end if it is moving toward us (Figure 17.10). The greater the shift, the faster the star is moving. Such motion, along the line of sight between the star and the observer, is called radial velocity and is usually measured in kilometers per second.

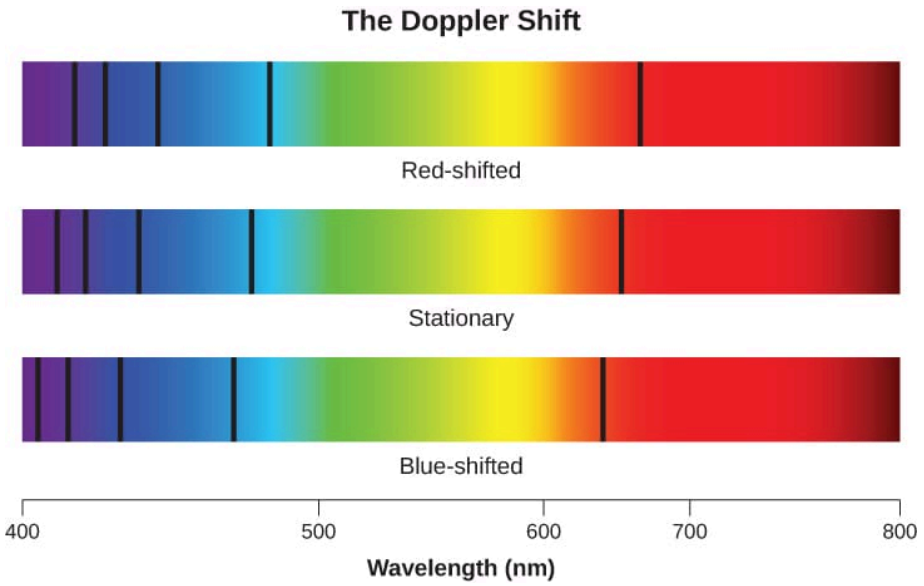


Figure 17.10 : Doppler-Shifted Stars. When the spectral lines of a moving star shift toward the red end of the spectrum, we know that the star is moving away from us. If they shift toward the blue end, the star is moving toward us.

William Huggins, pioneering yet again, in 1868 made the first radial velocity determination of a star. He observed the Doppler shift in one of the hydrogen lines in the spectrum of Sirius and found that this star is moving toward the solar system. (Better measurements now indicate that Sirius is actually moving away, but Huggins is still celebrated as a pioneer in making such measurements.) Today, radial velocity can be measured for any star bright enough for its spectrum to be observed. As we will see in The Stars: A Celestial Census, radial velocity measurements of double stars are crucial in deriving stellar masses.

Link to Learning

In the Spectrum Constructor, you can add absorption lines to the spectrum. Then adjust the velocity; negative indicates the star is moving toward the observer, and positive indicates it is moving away. Explore how the absorption lines are seen to shift to shorter and longer wavelengths.

Proper Motion

There is another type of motion stars can have that cannot be detected with stellar spectra. Unlike radial motion, which is along our line of sight (i.e., toward or away from Earth), this motion, called proper motion, is *transverse*: that is, across our line of sight. We see it as a change in the relative positions of the stars on the celestial sphere (Figure 17.11). These changes are very slow. Even the star with the largest proper motion takes 200 years to change its position in the sky by an amount equal to the width of the full Moon, and the motions of other stars are smaller yet.

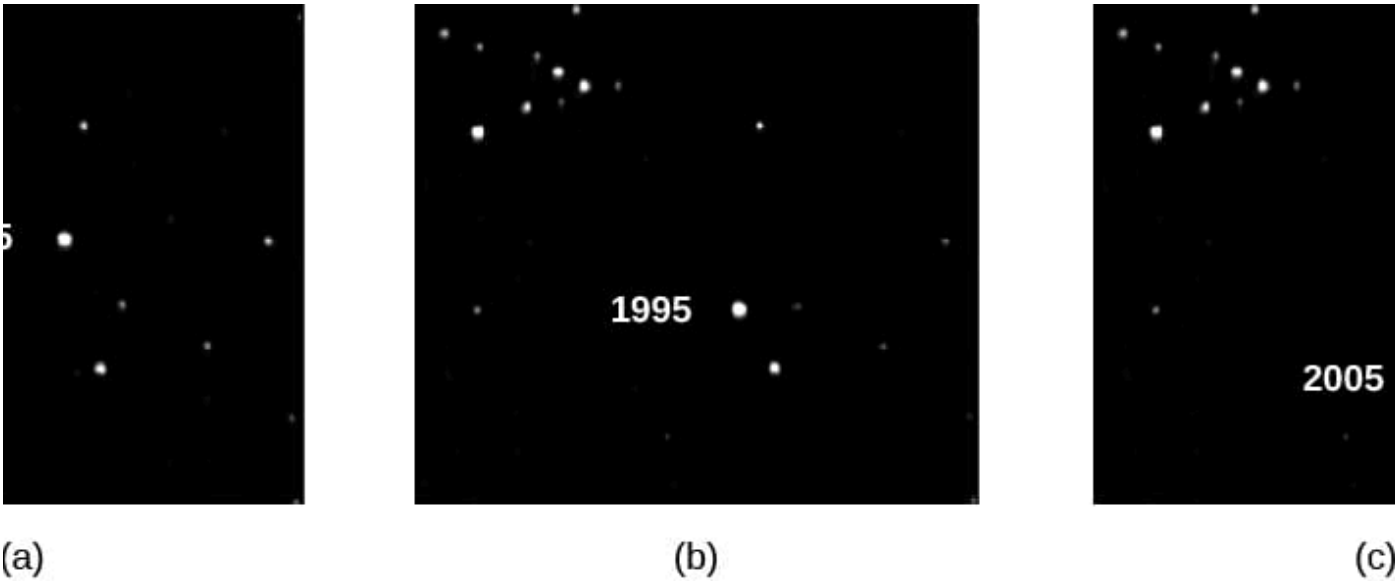


Figure 17.11 : Large Proper Motion. Three photographs of Barnard's star, the star with the largest known proper motion, show how this faint star has moved over a period of 20 years. (modification of work by Steve Quirk)

For this reason, with our naked eyes, we do not notice any change in the positions of the bright stars during the course of a human lifetime. If we could live long enough, however, the changes would become obvious. For example, some 50,000 years from now, terrestrial observers will find the handle of the Big Dipper unmistakably more bent than it is now (Figure 17.12).

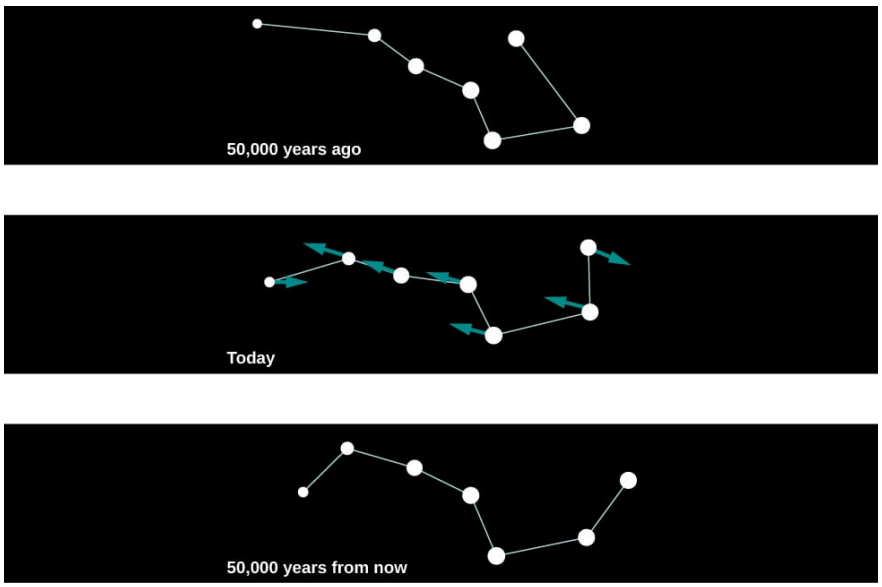


Figure 17.12 : Changes in the Big Dipper. This figure shows changes in the appearance of the Big Dipper due to proper motion of the stars over 100,000 years.

We measure the proper motion of a star in arcseconds (1/3600 of a degree) per year. That is, the measurement of proper motion tells us only by how much of an angle a star has changed its position on the celestial sphere. If two stars at different distances

are moving at the same velocity perpendicular to our line of sight, the closer one will show a larger shift in its position on the celestial sphere in a year's time. As an analogy, imagine you are standing at the side of a freeway. Cars will appear to whiz past you. If you then watch the traffic from a vantage point half a mile away, the cars will move much more slowly across your field of vision. In order to convert this angular motion to a velocity, we need to know how far away the star is.

To know the true space velocity of a star—that is, its total speed and the direction in which it is moving through space relative to the Sun—we must know its radial velocity, proper motion, and distance (Figure 17.13). A star's space velocity can also, over time, cause its distance from the Sun to change significantly. Over several hundred thousand years, these changes can be large enough to affect the apparent brightnesses of nearby stars. Today, Sirius, in the constellation Canis Major (the Big Dog) is the brightest star in the sky, but 100,000 years ago, the star Canopus in the constellation Carina (the Keel) was the brightest one. A little over 200,000 years from now, Sirius will have moved away and faded somewhat, and Vega, the bright blue star in Lyra, will take over its place of honor as the brightest star in Earth's skies.

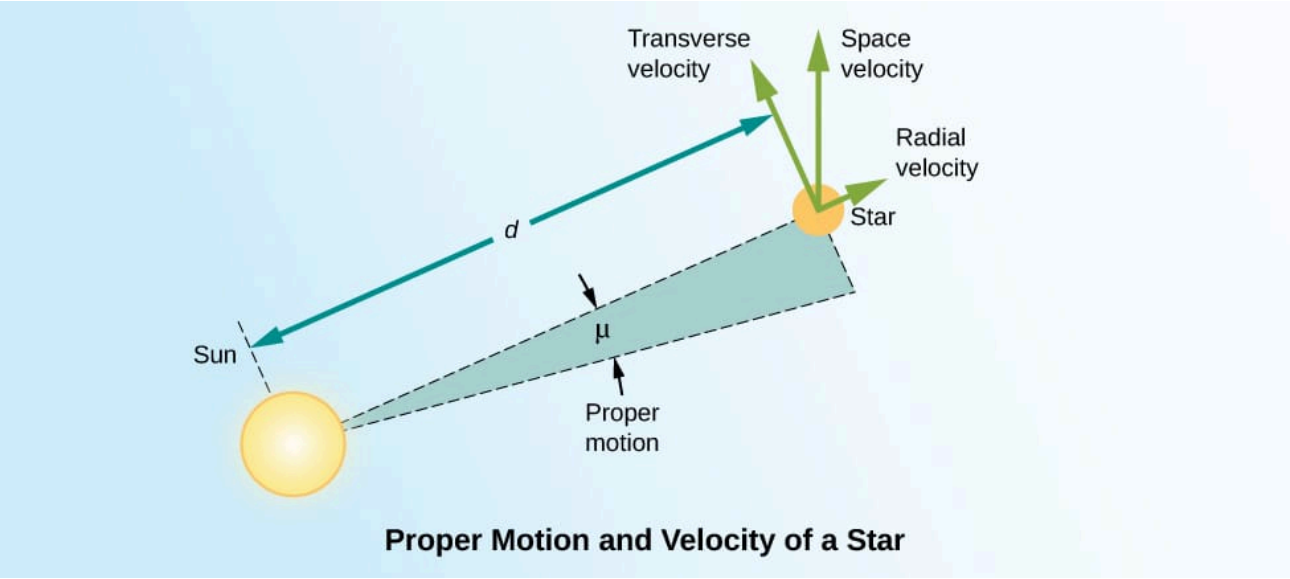


Figure 17.13 : Space Velocity and Proper Motion. This figure shows the true space velocity of a star. The radial velocity is the component of the space velocity projected along the line of sight from the Sun to a star. The transverse velocity is a component of the space velocity projected on the sky. What astronomers measure is proper motion (μ), which is the change in the apparent direction on the sky measured in fractions of a degree. To convert this change in direction to a speed in, say, kilometers per second, it is necessary to also know the distance (*d*) from the Sun to the star.

Rotation

We can also use the Doppler effect to measure how fast a star rotates. If an object is rotating, then one of its sides is approaching us while the other is receding (unless its axis of rotation happens to be pointed exactly toward us). This is clearly the case for the Sun or a planet; we can observe the light from either the approaching or receding edge of these nearby objects and directly measure the Doppler shifts that arise from the rotation.

Stars, however, are so far away that they all appear as unresolved points. The best we can do is to analyze the light from the entire star at once. Due to the Doppler effect, the lines in the light that come from the side of the star rotating toward us are shifted to shorter wavelengths and the lines in the light from the opposite edge of the star are shifted to longer wavelengths. You can think of each spectral line that we observe as the sum or composite of spectral lines originating from different speeds with respect to us. Each point on the star has its own Doppler shift, so the absorption line we see from the whole star is actually much wider than it would be if the star were not rotating. If a star is rotating rapidly, there will be a greater spread of Doppler shifts and all its spectral lines should be quite broad. In fact, astronomers call this effect *line broadening*, and the amount of broadening can tell us the speed at which the star rotates (Figure 17.14).

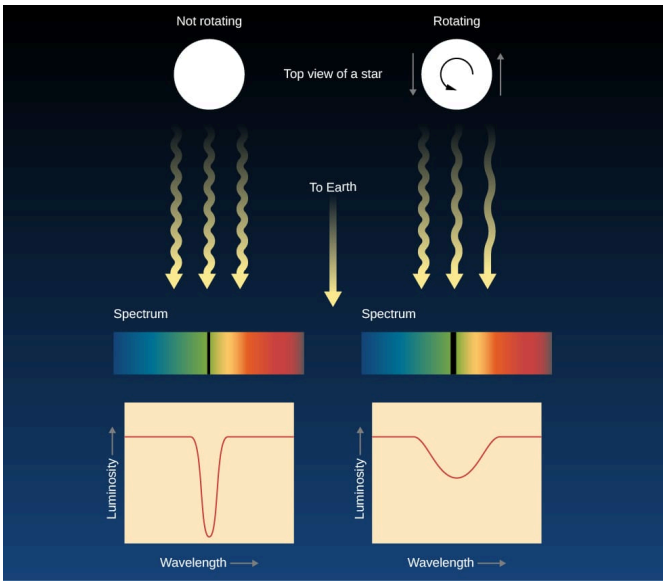


Figure 17.14 : Using a Spectrum to Determine Stellar Rotation. A rotating star will show broader spectral lines than a nonrotating star.

Measurements of the widths of spectral lines show that many stars rotate faster than the Sun, some with periods of less than a day! These rapid rotators spin so fast that their shapes are “flattened” into what we call *oblate spheroids*. An example of this is the star Vega, which rotates once every 12.5 hours. Vega’s rotation flattens its shape so much that its diameter at the equator is 23% wider than its diameter at the poles (Figure 17.15). The Sun, with its rotation period of about a month, rotates rather slowly. Studies have shown that stars decrease their rotational speed as they age. Young stars rotate very quickly, with rotational periods of days or less. Very old stars can have rotation periods of several months.

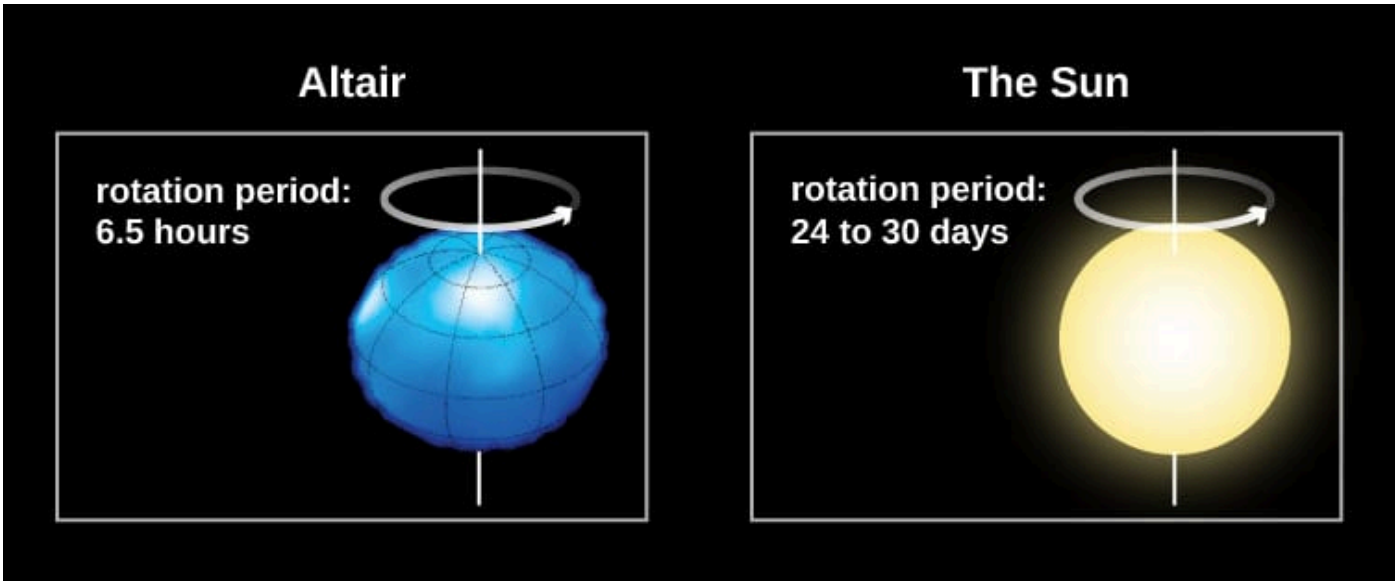


Figure 17.15 : Comparison of Rotating Stars. This illustration compares the more rapidly rotating star Altair to the slower rotating Sun.

As you can see, spectroscopy is an extremely powerful technique that helps us learn all kinds of information about stars that we simply could not gather any other way. We will see in later chapters that these same techniques can also teach us about galaxies, which are the most distant objects that can we observe. Without spectroscopy, we would know next to nothing about the universe beyond the solar system.

Making Connections: Astronomy and Philanthropy

Throughout the history of astronomy, contributions from wealthy patrons of the science have made an enormous difference in building new instruments and carrying out long-term research projects. Edward Pickering's stellar classification project, which was to stretch over several decades, was made possible by major donations from Anna Draper. She was the widow of Henry Draper, a physician who was one of the most accomplished amateur astronomers of the nineteenth century and the first person to successfully photograph the spectrum of a star. Anna Draper gave several hundred thousand dollars to Harvard Observatory. As a result, the great spectroscopic survey is still known as the Henry Draper Memorial, and many stars are still referred to by their "HD" numbers in that catalog (such as HD 209458).

In the 1870s, the eccentric piano builder and real estate magnate James Lick (Figure 17.16) decided to leave some of his fortune to build the world's largest telescope. When, in 1887, the pier to house the telescope was finished, Lick's body was entombed in it. Atop the foundation rose a 36-inch refractor, which for many years was the main instrument at the Lick Observatory near San Jose.

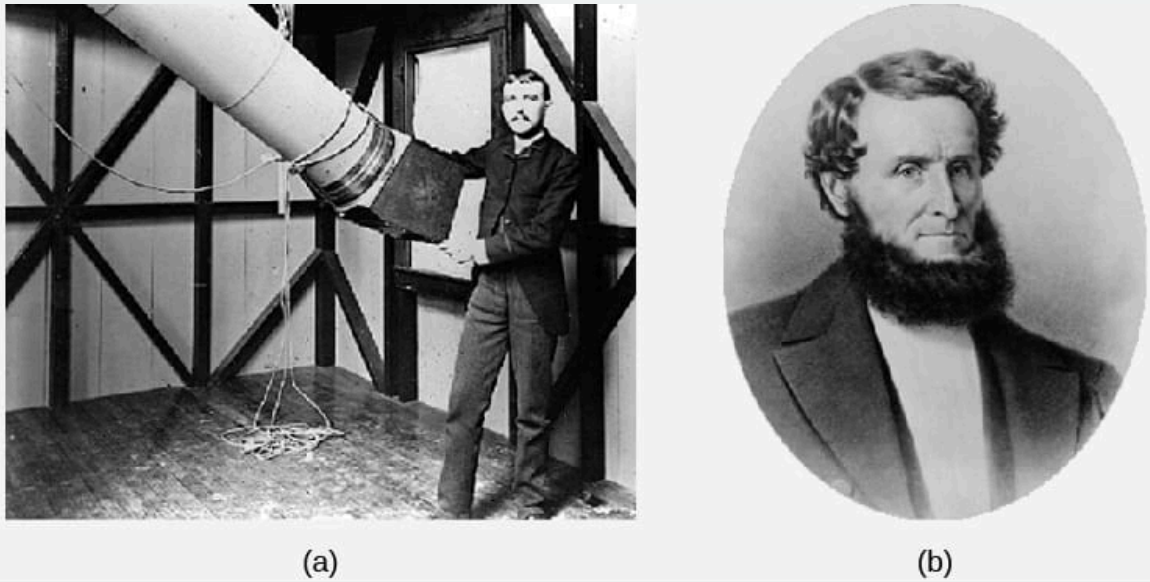


Figure 17.16 : Henry Draper (1837–1882) and James Lick (1796–1876). (a) Draper stands next to a telescope used for photography. After his death, his widow funded further astronomy work in his name. (b) Lick was a philanthropist who provided funds to build a 36-inch refractor not only as a memorial to himself but also to aid in further astronomical research.

The Lick telescope remained the largest in the world until 1897, when George Ellery Hale persuaded railroad millionaire Charles Yerkes to finance the construction of a 40-inch telescope near Chicago. More recently, Howard Keck, whose family made its fortune in the oil industry, gave $70 million from his family foundation to the California Institute of Technology to help build the world's largest telescope atop the 14,000-foot peak of Maunakea in Hawaii (see the chapter on Astronomical Instruments to learn more about these telescopes). The Keck Foundation was so pleased with what is now called the Keck telescope that they gave $74 million more to build Keck II, another 10-meter reflector on the same volcanic peak.

Now, if any of you become millionaires or billionaires, and astronomy has sparked your interest, do keep an astronomical instrument or project in mind as you plan your estate. But frankly, private philanthropy could not possibly support the full enterprise of scientific research in astronomy. Much of our exploration of the universe is financed by federal agencies such as the National Science Foundation and NASA in the United States, and by similar government agencies in the other countries. In this way, all of us, through a very small share of our tax dollars, are philanthropists for astronomy.

This page titled 12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 17.5: Using Spectra to Measure Stellar Radius, Composition, and Motion by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.6: A Stellar Census

Learning Objectives

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

- Explain why the stars visible to the unaided eye are not typical
- Describe the distribution of stellar masses found close to the Sun

Before we can make our own survey, we need to agree on a unit of distance appropriate to the objects we are studying. The stars are all so far away that kilometers (and even astronomical units) would be very cumbersome to use; so—as discussed in Science and the Universe: A Brief Tour—astronomers use a much larger "measuring stick" called the *light-year*. A light-year is the distance that light (the fastest signal we know) travels in 1 year. Since light covers an astounding 300,000 kilometers per second, and since there are a lot of seconds in 1 year, a light-year is a very large quantity: 9.5 trillion (9.5×10^{12}) kilometers to be exact. (Bear in mind that the light-year is a unit of *distance* even though the term *year* appears in it.) If you drove at the legal US speed limit without stopping for food or rest, you would not arrive at the end of a light-year in space until roughly 12 million years had passed. And the closest star is more than 4 light-years away.

Notice that we have not yet said much about how such enormous distances can be measured. That is a complicated question, to which we will return in Celestial Distances. For now, let us assume that distances have been measured for stars in our cosmic vicinity so that we can proceed with our census.

Small Is Beautiful—Or at Least More Common

When we do a census of people in the United States, we count the inhabitants by neighborhood. We can try the same approach for our stellar census and begin with our own immediate neighborhood. As we shall see, we run into two problems—just as we do with a census of human beings. First, it is hard to be sure we have counted *all* the inhabitants; second, our local neighborhood may not contain all possible types of people.

Table 18.1 shows an estimate of the number of stars of each spectral type[1] in our own local neighborhood—within 21 light-years of the Sun. (The Milky Way Galaxy, in which we live, is about 100,000 light-years in diameter, so this figure really applies to a *very* local neighborhood, one that contains a *tiny* fraction of all the billions of stars in the Milky Way.) You can see that there are many more low-luminosity (and hence low mass) stars than high-luminosity ones. Only three of the stars in our local neighborhood (one F type and two A types) are significantly more luminous and more massive than the Sun. This is truly a case where small triumphs over large—at least in terms of numbers. The Sun is more massive than the vast majority of stars in our vicinity.

Table 18.1: Stars within 21 Light-Years of the Sun

Spectral Type	Number of Stars
A	2
F	1
G	7
K	17
M	94
White dwarfs	8
Brown dwarfs	33

This table is based on data published through 2015, and it is likely that more faint objects remain to be discovered (see Figure 18.2). Along with the L and T brown dwarfs already observed in our neighborhood, astronomers expect to find perhaps hundreds of additional T dwarfs and Y dwarfs. Many of these are likely to be even cooler than the coolest currently known T dwarf. The

reason the lowest-mass dwarfs are so hard to find is that they put out very little light—ten thousand to a million times less light than the Sun. Only recently has our technology progressed to the point that we can detect these dim, cool objects.

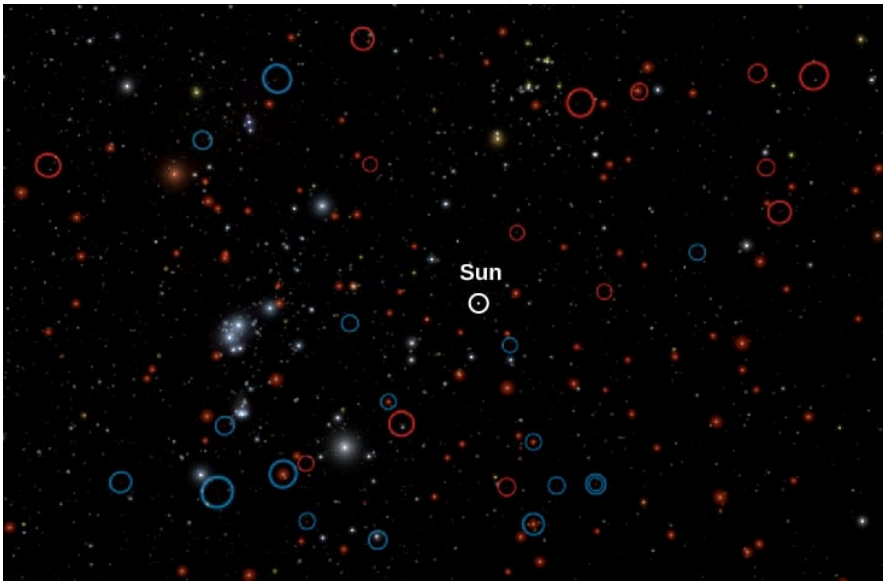


Figure 18.2 : Dwarf Simulation. This computer simulation shows the stars in our neighborhood as they would be seen from a distance of 30 light-years away. The Sun is in the center. All the brown dwarfs are circled; those found earlier are circled in blue, the ones found recently with the WISE infrared telescope in space (whose scientists put this diagram together) are circled in red. The common M stars, which are red and faint, are made to look brighter than they really would be so that you can see them in the simulation. Note that luminous hot stars like our Sun are very rare. (credit: modification of work by NASA/ JPL-Caltech)

To put all this in perspective, we note that even though the stars counted in the table are our closest neighbors, you can't just look up at the night sky and see them without a telescope; stars fainter than the Sun cannot be seen with the unaided eye unless they are *very* nearby. For example, stars with luminosities ranging from 1/100 to 1/10,000 the luminosity of the Sun (L_{Sun}) are very common, but a star with a luminosity of 1/100 L_{Sun} would have to be within 5 light-years to be visible to the naked eye—and only three stars (all in one system) are this close to us. The nearest of these three stars, Proxima Centauri, still cannot be seen without a telescope because it has such a low luminosity.

Astronomers are working hard these days to complete the census of our local neighborhood by finding our faintest neighbors. Recent discoveries of nearby stars have relied heavily upon infrared telescopes that are able to find these many cool, low-mass stars. You should expect the number of known stars within 21 light-years of the Sun to keep increasing as more and better surveys are undertaken.

Bright Does Not Necessarily Mean Close

If we confine our census to the local neighborhood, we will miss many of the most interesting kinds of stars. After all, the neighborhood in which you live does not contain all the types of people—distinguished according to age, education, income, race, and so on—that live in the entire country. For example, a few people do live to be over 100 years old, but there may be no such individual within several miles of where you live. In order to sample the full range of the human population, you would have to extend your census to a much larger area. Similarly, some types of stars simply are not found nearby.

A clue that we are missing something in our stellar census comes from the fact that only six of the 20 stars that appear brightest in our sky—Sirius, Vega, Altair, Alpha Centauri, Fomalhaut, and Procyon—are found within 26 light-years of the Sun (Figure 18.3). Why are we missing most of the brightest stars when we take our census of the local neighborhood?

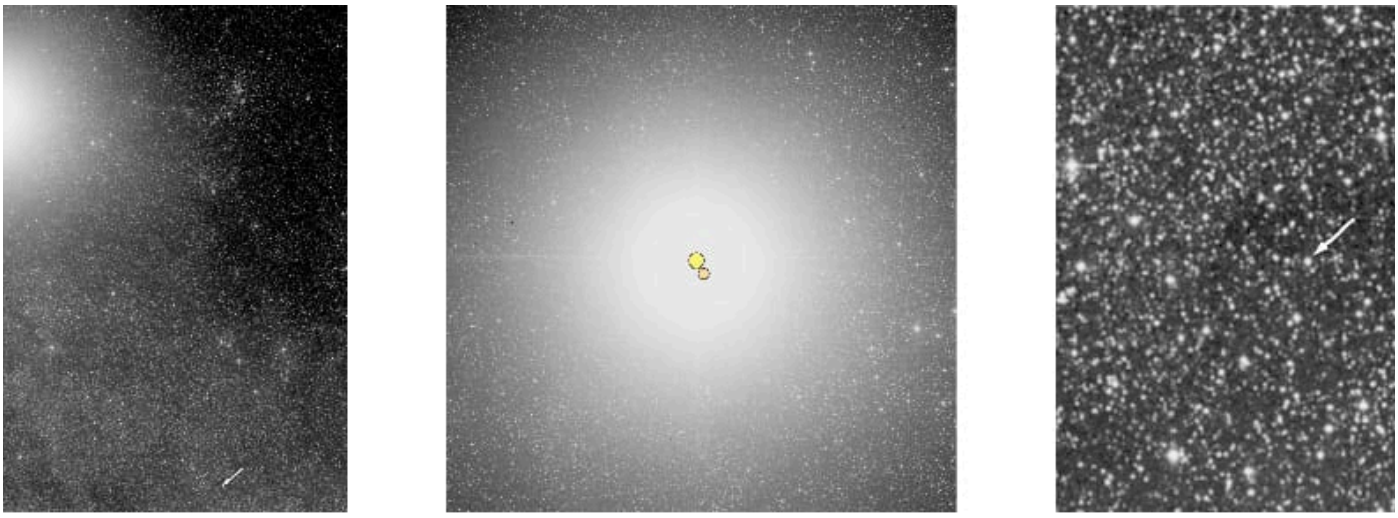


Figure 18.3 : The Closest Stars. (a) This image, taken with a wide-angle telescope at the European Southern Observatory in Chile, shows the system of three stars that is our nearest neighbor. (b) Two bright stars that are close to each other (Alpha Centauri A and B) blend their light together. (c) Indicated with an arrow (since you'd hardly notice it otherwise) is the much fainter Proxima Centauri star, which is spectral type M. (credit: modification of work by ESO)

The answer, interestingly enough, is that the stars that appear brightest are *not* the ones closest to us. The brightest stars look the way they do because they emit a very large amount of energy—so much, in fact, that they do not have to be nearby to look brilliant. You can confirm this by looking at Appendix J, which gives distances for the 20 stars that appear brightest from Earth. The most distant of these stars is more than *1000 light-years* from us. In fact, it turns out that most of the stars visible without a telescope are hundreds of light-years away and many times more luminous than the Sun. Among the 9000 stars visible to the unaided eye, only about 50 are intrinsically fainter than the Sun. Note also that several of the stars in Appendix J are spectral type B, a type that is completely missing from Table 18.1.

The most luminous of the bright stars listed in Appendix J emit more than 50,000 times more energy than does the Sun. These highly luminous stars are missing from the solar neighborhood because they are very rare. None of them happens to be in the tiny volume of space immediately surrounding the Sun, and only this small volume was surveyed to get the data shown in Table 18.1.

For example, let's consider the most luminous stars—those 100 or more times as luminous as the Sun. Although such stars are rare, they are visible to the unaided eye, even when hundreds to thousands of light-years away. A star with a luminosity 10,000 times greater than that of the Sun can be seen without a telescope out to a distance of 5000 light-years. The volume of space included within a distance of 5000 light-years, however, is enormous; so even though highly luminous stars are intrinsically rare, many of them are readily visible to our unaided eye.

The contrast between these two samples of stars, those that are close to us and those that can be seen with the unaided eye, is an example of a selection effect. When a population of objects (stars in this example) includes a great variety of different types, we must be careful what conclusions we draw from an examination of any particular subgroup. Certainly we would be fooling ourselves if we assumed that the stars visible to the unaided eye are characteristic of the general stellar population; this subgroup is heavily weighted to the most luminous stars. It requires much more effort to assemble a complete data set for the nearest stars, since most are so faint that they can be observed only with a telescope. However, it is only by doing so that astronomers are able to know about the properties of the vast majority of the stars, which are actually much smaller and fainter than our own Sun. In the next section, we will look at how we measure some of these properties.

Footnotes

- 1The spectral types of stars were defined and discussed in Analyzing Starlight.

This page titled 12.6: A Stellar Census is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax.

- 18.2: A Stellar Census by OpenStax is licensed CC BY 4.0.

12.7: Measuring Stellar Masses

Learning Objectives

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

- Distinguish the different types of binary star systems
- Understand how we can apply Newton's version of Kepler's third law to derive the sum of star masses in a binary star system
- Apply the relationship between stellar mass and stellar luminosity to determine the physical characteristics of a star

The mass of a star—how much material it contains—is one of its most important characteristics. If we know a star's mass, as we shall see, we can estimate how long it will shine and what its ultimate fate will be. Yet the mass of a star is very difficult to measure directly. Somehow, we need to put a star on the cosmic equivalent of a scale.

Luckily, not all stars live like the Sun, in isolation from other stars. About half the stars are binary stars—two stars that orbit each other, bound together by gravity. Masses of binary stars can be calculated from measurements of their orbits, just as the mass of the Sun can be derived by measuring the orbits of the planets around it (see Orbits and Gravity).

Binary Stars

Before we discuss in more detail how mass can be measured, we will take a closer look at stars that come in pairs. The first binary star was discovered in 1650, less than half a century after Galileo began to observe the sky with a telescope. John Baptiste Riccioli (1598–1671), an Italian astronomer, noted that the star Mizar, in the middle of the Big Dipper's handle, appeared through his telescope as two stars. Since that discovery, thousands of binary stars have been cataloged. (Astronomers call any pair of stars that appear to be close to each other in the sky *double stars*, but not all of these form a true binary, that is, not all of them are physically associated. Some are just chance alignments of stars that are actually at different distances from us.) Although stars most commonly come in pairs, there are also triple and quadruple systems.

One well-known binary star is Castor, located in the constellation of Gemini. By 1804, astronomer William Herschel, who also discovered the planet Uranus, had noted that the fainter component of Castor had slightly changed its position relative to the brighter component. (We use the term "component" to mean a member of a star system.) Here was evidence that one star was moving around another. It was actually the first evidence that gravitational influences exist outside the solar system. The orbital motion of a binary star is shown in Figure 18.4. A binary star system in which both of the stars can be seen with a telescope is called a visual binary.

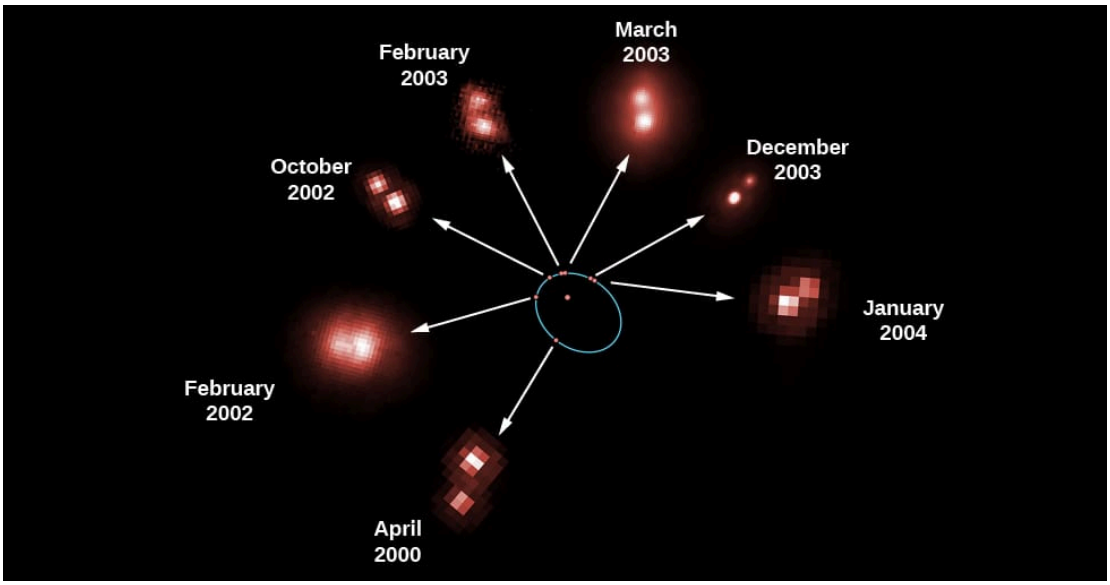


Figure 18.4 : Revolution of a Binary Star. This figure shows seven observations of the mutual revolution of two stars, one a brown dwarf and one an ultra-cool L dwarf. Each red dot on the orbit, which is shown by the blue ellipse, corresponds to the position of one of the dwarfs relative to the other. The reason that the pair of stars looks different on the different dates is that some images were taken with the Hubble Space Telescope and others were taken from the ground. The arrows point to the actual observations that correspond to the positions of each red dot. From these observations, an international team of astronomers directly measured the mass of an ultra-cool brown dwarf star for the first time. Barely the size of the planet Jupiter, the dwarf star weighs in at just 8.5% of the mass of our Sun. (credit: modification of work by ESA/ NASA and Herve Bouy (Max-Planck-Institut für Extraterrestrische Physik/ESO, Germany))

Edward C. Pickering (1846–1919), at Harvard, discovered a second class of binary stars in 1889—a class in which only one of the stars is actually seen directly. He was examining the spectrum of Mizar and found that the dark absorption lines in the brighter star's spectrum were usually double. Not only were there two lines where astronomers normally saw only one, but the spacing of the lines was constantly changing. At times, the lines even became single. Pickering correctly deduced that the brighter component of Mizar, called Mizar A, is itself really two stars that revolve about each other in a period of 104 days. A star like Mizar A, which appears as a single star when photographed or observed visually through the telescope, but which spectroscopy shows really to be a double star, is called a spectroscopic binary.

Mizar, by the way, is a good example of just how complex such star systems can be. Mizar has been known for centuries to have a faint companion called Alcor, which can be seen without a telescope. Mizar and Alcor form an *optical double*—a pair of stars that appear close together in the sky but do not orbit each other. Through a telescope, as Riccioli discovered in 1650, Mizar can be seen to have another, closer companion that does orbit it; Mizar is thus a visual binary. The two components that make up this visual binary, known as Mizar A and Mizar B, are both spectroscopic binaries. So, Mizar is really a quadruple system of stars.

Strictly speaking, it is not correct to describe the motion of a binary star system by saying that one star orbits the other. Gravity is a *mutual* attraction. Each star exerts a gravitational force on the other, with the result that both stars orbit a point between them called the *center of mass*. Imagine that the two stars are seated at either end of a seesaw. The point at which the fulcrum would have to be located in order for the seesaw to balance is the center of mass, and it is always closer to the more massive star (Figure 18.5).

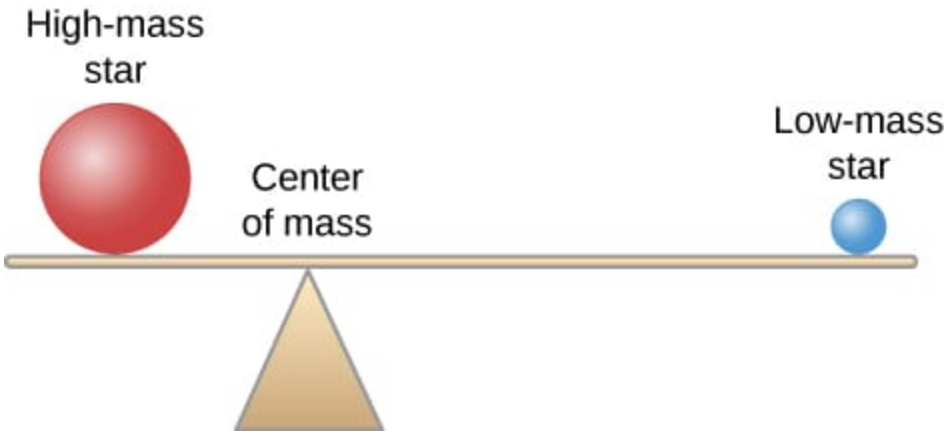


Figure 18.5 : Binary Star System. In a binary star system, both stars orbit their center of mass. The image shows the relative positions of two, different-mass stars from their center of mass, similar to how two masses would have to be located on a seesaw in order to keep it level. The star with the higher mass will be found closer to the center of mass, while the star with the lower mass will be farther from it.

Figure 18.6 shows two stars (A and B) moving around their center of mass, along with one line in the spectrum of each star that we observe from the system at different times. When one star is approaching us relative to the center of mass, the other star is receding from us. In the top left illustration, star A is moving toward us, so the line in its spectrum is Doppler-shifted toward the blue end of the spectrum. Star B is moving away from us, so its line shows a redshift. When we observe the composite spectrum of the two stars, the line appears double. When the two stars are both moving across our line of sight (neither away from nor toward us), they both have the same radial velocity (that of the pair's center of mass); hence, the spectral lines of the two stars come together. This is shown in the two bottom illustrations in Figure 18.6.

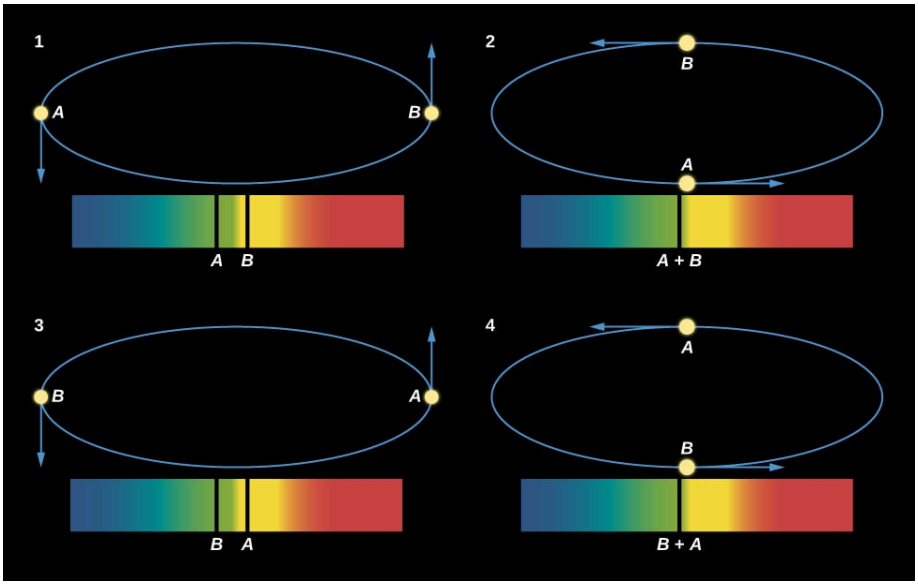


Figure 18.6 : Motions of Two Stars Orbiting Each Other and What the Spectrum Shows. We see changes in velocity because when one star is moving toward Earth, the other is moving away; half a cycle later, the situation is reversed. Doppler shifts cause the spectral lines to move back and forth. In diagrams 1 and 3, lines from both stars can be seen well separated from each other. When the two stars are moving perpendicular to our line of sight (that is, they are not moving either toward or away from us), the two lines are exactly superimposed, and so in diagrams 2 and 4, we see only a single spectral line. Note that in the diagrams, the orbit of the star pair is tipped slightly with respect to the viewer (or if the viewer were looking at it in the sky, the orbit would be tilted with respect to the viewer's line of sight). If the orbit were exactly in the plane of the page or screen (or the sky), then it would look nearly circular, but we would see no change in radial velocity (no part of the motion would be toward us or away from us.) If the orbit were perpendicular to the plane of the page or screen, then the stars would appear to move back and forth in a straight line, and we would see the largest-possible radial velocity variations. This diagram has been simplified for clarity. For more realistic situations, see the animation in the following Link to Learning feature box.

A plot showing how the velocities of the stars change with time is called a *radial velocity curve*; the curve for the binary system in Figure 18.6 is shown in Figure 18.7.

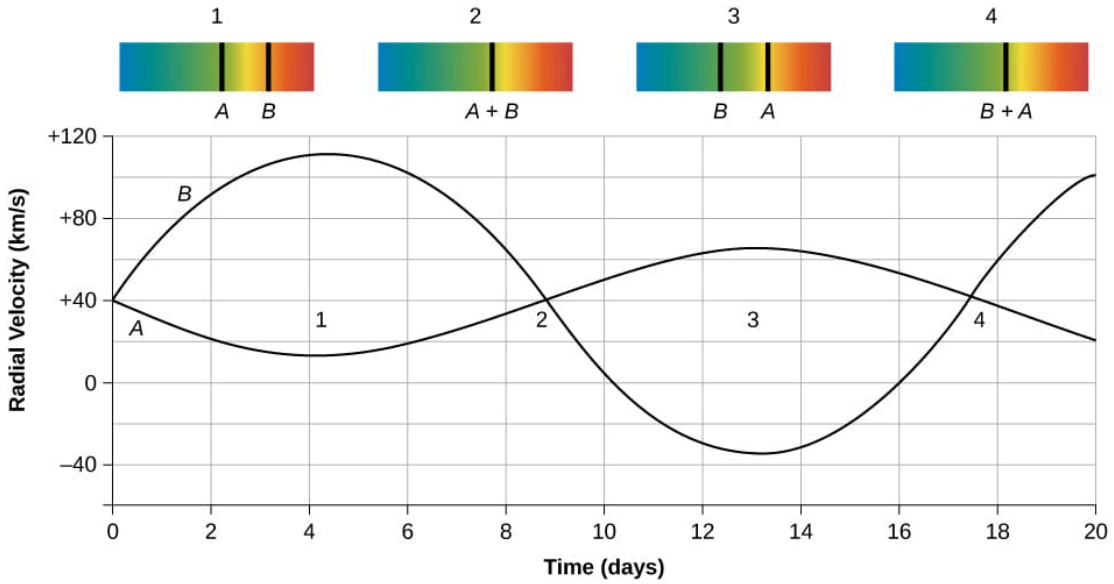


Figure 18.7 : Radial Velocities in a Spectroscopic Binary System. These curves plot the radial velocities of two stars in a spectroscopic binary system, showing how the stars alternately approach and recede from Earth. Note that positive velocity means the star is moving away from us, and negative velocity means the star is moving toward us. The center of mass of the system itself is also moving away from us, indicated by the positive velocity of 40 kilometers per second. The positions on the curve corresponding to the illustrations in Figure 18.6 are marked with the diagram number (1–4).

Link to Learning

This animation lets you follow the orbits of a binary star system in various combinations of the masses of the two stars.

Masses from the Orbits of Binary Stars

We can estimate the masses of binary star systems using Newton's reformulation of Kepler's third law (discussed in Newton's Universal Law of Gravitation). Kepler found that the time a planet takes to go around the Sun is related by a specific mathematical formula to its distance from the Sun. In our binary star situation, if two objects are in mutual revolution, then the period (*P*) with which they go around each other is related to the semimajor axis (*D*) of the orbit of one with respect to the other, according to this equation

$$D^3 = (M_1 + M_2)P^2$$

where *D* is in astronomical units, *P* is measured in years, and $M_1 + M_2$ is the sum of the masses of the two stars in units of the Sun's mass. This is a very useful formula for astronomers; it says that if we can observe the size of the orbit and the period of mutual revolution of the stars in a binary system, we can calculate the sum of their masses.

Most spectroscopic binaries have periods ranging from a few days to a few months, with separations of usually less than 1 AU between their member stars. Recall that an AU is the distance from Earth to the Sun, so this is a small separation and very hard to see at the distances of stars. This is why many of these systems are known to be double only through careful study of their spectra.

We can analyze a radial velocity curve (such as the one in Figure 18.7) to determine the masses of the stars in a spectroscopic binary. This is complex in practice but not hard in principle. We measure the speeds of the stars from the Doppler effect. We then determine the period—how long the stars take to go through an orbital cycle—from the velocity curve. Knowing how fast the stars are moving and how long they take to go around tells us the circumference of the orbit and, hence, the separation of the stars in kilometers or astronomical units. From Kepler's law, the period and the separation allow us to calculate the sum of the stars' masses.

Of course, knowing the sum of the masses is not as useful as knowing the mass of each star separately. But the relative orbital speeds of the two stars can tell us how much of the total mass each star has. As we saw in our seesaw analogy, the more massive

star is closer to the center of mass and therefore has a smaller orbit. Therefore, it moves more slowly to get around in the same time compared to the more distant, lower-mass star. If we sort out the speeds relative to each other, we can sort out the masses relative to each other. In practice, we also need to know how the binary system is oriented in the sky to our line of sight, but if we do, and the just-described steps are carried out carefully, the result is a calculation of the masses of each of the two stars in the system.

To summarize, a good measurement of the motion of two stars around a common center of mass, combined with the laws of gravity, allows us to determine the masses of stars in such systems. These mass measurements are absolutely crucial to developing a theory of how stars evolve. One of the best things about this method is that it is independent of the location of the binary system. It works as well for stars 100 light-years away from us as for those in our immediate neighborhood.

To take a specific example, Sirius is one of the few binary stars in Appendix J for which we have enough information to apply Kepler's third law:

$$D^3 = (M_1 + M_2)P^2$$

In this case, the two stars, the one we usually call Sirius and its very faint companion, are separated by about 20 AU and have an orbital period of about 50 years. If we place these values in the formula we would have

$$(20)^3 = (M_1 + M_2)(50)^2$$
$$8000 = (M_1 + M_2)(2500)$$

This can be solved for the sum of the masses:

$$M_1 + M_2 = \frac{8000}{2500} = 3.2$$

Therefore, the sum of masses of the two stars in the Sirius binary system is 3.2 times the Sun's mass. In order to determine the individual mass of each star, we would need the velocities of the two stars and the orientation of the orbit relative to our line of sight.

The Range of Stellar Masses

How large can the mass of a star be? Stars more massive than the Sun are rare. None of the stars within 30 light-years of the Sun has a mass greater than four times that of the Sun. Searches at large distances from the Sun have led to the discovery of a few stars with masses up to about 100 times that of the Sun, and a handful of stars (a few out of several billion) may have masses as large as 250 solar masses. However, most stars have less mass than the Sun.

According to theoretical calculations, the smallest mass that a true star can have is about 1/12 that of the Sun. By a "true" star, astronomers mean one that becomes hot enough to fuse protons to form helium (as discussed in The Sun: A Nuclear Powerhouse). Objects with masses between roughly 1/100 and 1/12 that of the Sun may produce energy for a brief time by means of nuclear reactions involving deuterium, but they do not become hot enough to fuse protons. Such objects are intermediate in mass between stars and planets and have been given the name brown dwarfs (Figure 18.8). Brown dwarfs are similar to Jupiter in radius but have masses from approximately 13 to 80 times larger than the mass of Jupiter.[2]



Figure 18.8 : Brown Dwarfs in Orion. These images, taken with the Hubble Space Telescope, show the region surrounding the Trapezium star cluster inside the star-forming region called the Orion Nebula. (a) No brown dwarfs are seen in the visible light image, both because they put out very little light in the visible and because they are hidden within the clouds of dust in this region. (b) This image was taken in infrared light, which can make its way to us through the dust. The faintest objects in this image are brown dwarfs with masses between 13 and 80 times the mass of Jupiter. (credit a: NASA, C.R. O'Dell and S.K. Wong (Rice University); credit b: NASA; K.L. Luhman (Harvard-Smithsonian Center for Astrophysics) and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thompson (Steward Observatory))

Still-smaller objects with masses less than about 1/100 the mass of the Sun (or 10 Jupiter masses) are called planets. They may radiate energy produced by the radioactive elements that they contain, and they may also radiate heat generated by slowly compressing under their own weight (a process called gravitational contraction). However, their interiors will never reach temperatures high enough for any nuclear reactions, to take place. Jupiter, whose mass is about 1/1000 the mass of the Sun, is unquestionably a planet, for example. Until the 1990s, we could only detect planets in our own solar system, but now we have thousands of them elsewhere as well. (We will discuss these exciting observations in The Birth of Stars and the Discovery of Planets outside the Solar System.)

The Mass-Luminosity Relation

Now that we have measurements of the characteristics of many different types of stars, we can search for relationships among the characteristics. For example, we can ask whether the mass and luminosity of a star are related. It turns out that for most stars, they are: The more massive stars are generally also the more luminous. This relationship, known as the mass-luminosity relation, is shown graphically in Figure 18.9. Each point represents a star whose mass and luminosity are both known. The horizontal position on the graph shows the star's mass, given in units of the Sun's mass, and the vertical position shows its luminosity in units of the Sun's luminosity.

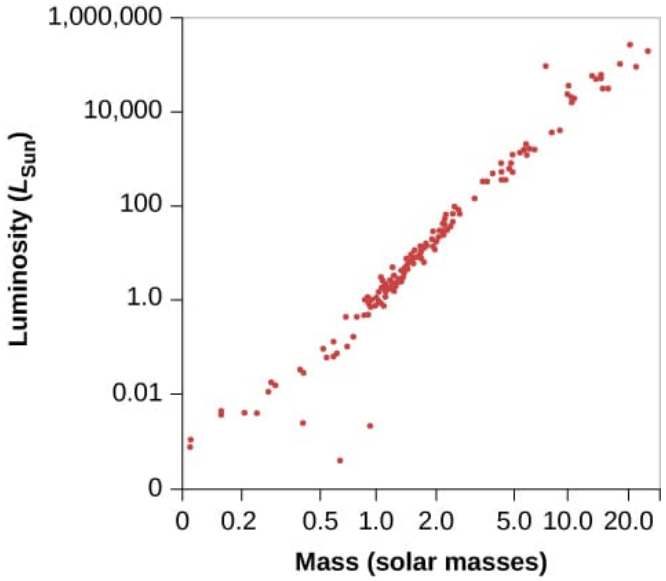


Figure 18.9 : Mass-Luminosity Relation. The plotted points show the masses and luminosities of stars. The three points lying below the sequence of points are all white dwarf stars.

We can also say this in mathematical terms.

$$L \sim M^{3.9}$$

It's a reasonably good approximation to say that luminosity (expressed in units of the Sun's luminosity) varies as the fourth power of the mass (in units of the Sun's mass). (The symbol ~ means the two quantities are proportional.) If two stars differ in mass by a factor of 2, then the more massive one will be 2^4, or about 16 times brighter; if one star is 1/3 the mass of another, it will be approximately 81 times less luminous.

Example 18.1: Calculating the Mass from the Luminosity of a Star

The mass-luminosity formula can be rewritten so that a value of mass can be determined if the luminosity is known.

Solution

First, we must get our units right by expressing both the mass and the luminosity of a star in units of the Sun's mass and luminosity:

$$L/L_{\text{Sun}} = (M/M_{\text{Sun}})^4$$

Now we can take the 4th root of both sides, which is equivalent to taking both sides to the 1/4 = 0.25 power. The formula in this case would be:

$$M/M_{\text{Sun}} = (L/L_{\text{Sun}})^{0.25} = (L/L_{\text{Sun}})^{0.25}$$

Exercise 12.7.1

In the previous section, we determined the sum of the masses of the two stars in the Sirius binary system (Sirius and its faint companion) using Kepler's third law to be 3.2 solar masses. Using the mass-luminosity relationship, calculate the mass of each individual star.

Answer available at the end of this section.

Notice how good this mass-luminosity relationship is. Most stars (see Figure 18.9) fall along a line running from the lower-left (low

mass, low luminosity) corner of the diagram to the upper-right (high mass, high luminosity) corner. About 90% of all stars obey the mass-luminosity relation. Later, we will explore why such a relationship exists and what we can learn from the roughly 10% of stars that “disobey” it.

Link to Learning

This visualization illustrates how all the properties of a star are determined by the mass that the star happened to form with and how these properties change when you change the mass.

Footnotes

- 2Exactly where to put the dividing line between planets and brown dwarfs is a subject of some debate among astronomers as we write this book (as is, in fact, the exact definition of each of these objects). Even those who accept deuterium fusion (see The Birth of Stars and the Discovery of Planets outside the Solar System) as the crucial issue for brown dwarfs concede that, depending on the composition of the star and other factors, the lowest mass for such a dwarf could be anywhere from 11 to 16 Jupiter masses.

Answers

Exercise $12.7.1$

In Appendix J, Sirius is listed with a luminosity 23 times that of the Sun. This value can be inserted into the mass-luminosity relationship to get the mass of Sirius: $M/M_{\text{Sun}} = 23^{0.25} = 2.2$
The mass of the companion star to Sirius is then $3.2 - 2.2 = 1.0$ solar mass.

This page titled 12.7: Measuring Stellar Masses is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 18.3: Measuring Stellar Masses by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.8: Diameters of Stars

Learning Objectives

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

- Describe the methods used to determine star diameters
- Identify the parts of an eclipsing binary star light curve that correspond to the diameters of the individual components

It is easy to measure the diameter of the Sun. Its angular diameter—that is, its apparent size on the sky—is about 1/2°. If we know the angle the Sun takes up in the sky and how far away it is, we can calculate its true (linear) diameter, which is 1.39 million kilometers, or about 109 times the diameter of Earth.

Unfortunately, the Sun is the only star whose angular diameter is easily measured. All the other stars are so far away that they look like pinpoints of light through even the largest ground-based telescopes. (They often seem to be bigger, but that is merely distortion introduced by turbulence in Earth's atmosphere.) Luckily, there are several techniques that astronomers can use to estimate the sizes of stars.

Stars Blocked by the Moon

One technique, which gives very precise diameters but can be used for only a few stars, is to observe the dimming of light that occurs when the Moon passes in front of a star. What astronomers measure (with great precision) is the time required for the star's brightness to drop to zero as the edge of the Moon moves across the star's disk. Since we know how rapidly the Moon moves in its orbit around Earth, it is possible to calculate the angular diameter of the star. If the distance to the star is also known, we can calculate its diameter in kilometers. This method works only for fairly bright stars that happen to lie along the zodiac, where the Moon (or, much more rarely, a planet) can pass in front of them as seen from Earth.

Eclipsing Binary Stars

Accurate sizes for a large number of stars come from measurements of eclipsing binary star systems, and so we must make a brief detour from our main story to examine this type of star system. Some binary stars are lined up in such a way that, when viewed from Earth, each star passes in front of the other during every revolution (Figure 18.10). When one star blocks the light of the other, preventing it from reaching Earth, the brightness of the system decreases, and astronomers say that an eclipse has occurred.

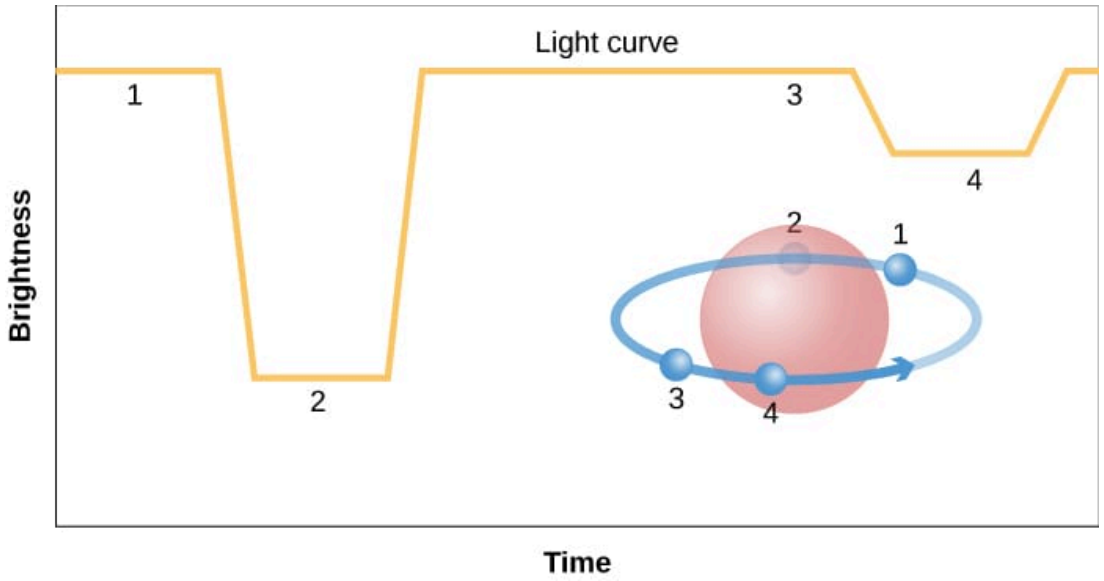


Figure 18.10 : Light Curve of an Eclipsing Binary. The light curve of an eclipsing binary star system shows how the combined light from both stars changes due to eclipses over the time span of an orbit. This light curve shows the behavior of a hypothetical eclipsing binary star with total eclipses (one star passes directly in front of and behind the other). The numbers indicate parts of the light curve corresponding to various positions of the smaller star in its orbit. In this diagram, we have assumed that the smaller star is also the hotter one so that it emits more flux (energy per second per square meter) than the larger one. When the smaller, hotter star goes behind the larger one, its light is completely blocked, and so there is a strong dip in the light curve. When the smaller star goes in front of the bigger one, a small amount of light from the bigger star is blocked, so there is a smaller dip in the light curve.

The discovery of the first eclipsing binary helped solve a long-standing puzzle in astronomy. The star Algol, in the constellation of Perseus, changes its brightness in an odd but regular way. Normally, Algol is a fairly bright star, but at intervals of 2 days, 20 hours, 49 minutes, it fades to one-third of its regular brightness. After a few hours, it brightens to normal again. This effect is easily seen, even without a telescope, if you know what to look for.

In 1783, a young English astronomer named John Goodricke (1764–1786) made a careful study of Algol (see the feature on John Goodricke for a discussion of his life and work). Even though Goodricke could neither hear nor speak, he made a number of major discoveries in the 21 years of his brief life. He suggested that Algol's unusual brightness variations might be due to an invisible companion that regularly passes in front of the brighter star and blocks its light. Unfortunately, Goodricke had no way to test this idea, since it was not until about a century later that equipment became good enough to measure Algol's spectrum.

In 1889, the German astronomer Hermann Vogel (1841–1907) demonstrated that, like Mizar, Algol is a spectroscopic binary. The spectral lines of Algol were not observed to be double because the fainter star of the pair gives off too-little light compared with the brighter star for its lines to be conspicuous in the composite spectrum. Nevertheless, the periodic shifting back and forth of the brighter star's lines gave evidence that it was revolving about an unseen companion. (The lines of both components need not be visible for a star to be recognized as a spectroscopic binary.)

The discovery that Algol is a spectroscopic binary verified Goodricke's hypothesis. The plane in which the stars revolve is turned nearly edgewise to our line of sight, and each star passes in front of the other during every revolution. (The eclipse of the fainter star in the Algol system is not very noticeable because the part of it that is covered contributes little to the total light of the system. This second eclipse can, however, be detected by careful measurements.)

Any binary star produces eclipses if viewed from the proper direction, near the plane of its orbit, so that one star passes in front of the other (see Figure 18.10). But from our vantage point on Earth, only a few binary star systems are oriented in this way.

Making Connections: Astronomy and Mythology: Algol the Demon Star and Perseus the Hero

The name Algol comes from the Arabic *Ras al Ghul*, meaning "the demon's head."[3] The word "ghoul" in English has the same derivation. As discussed in Observing the Sky: The Birth of Astronomy, many of the bright stars have Arabic names because during the long dark ages in medieval Europe, it was Arabic astronomers who preserved and expanded the Greek

and Roman knowledge of the skies. The reference to the demon is part of the ancient Greek legend of the hero Perseus, who is commemorated by the constellation in which we find Algol and whose adventures involve many of the characters associated with the northern constellations.

Perseus was one of the many half-god heroes fathered by Zeus (Jupiter in the Roman version), the king of the gods in Greek mythology. Zeus had, to put it delicately, a roving eye and was always fathering somebody or other with a human maiden who caught his fancy. (Perseus derives from *Per Zeus*, meaning "fathered by Zeus.") Set adrift with his mother by an (understandably) upset stepfather, Perseus grew up on an island in the Aegean Sea. The king there, taking an interest in Perseus' mother, tried to get rid of the young man by assigning him an extremely difficult task.

In a moment of overarching pride, a beautiful young woman named Medusa had compared her golden hair to that of the goddess Athena (Minerva for the Romans). The Greek gods did not take kindly to being compared to mere mortals, and Athena turned Medusa into a gorgon: a hideous, evil creature with writhing snakes for hair and a face that turned anyone who looked at it into stone. Perseus was given the task of slaying this demon, which seemed like a pretty sure way to get him out of the way forever.

But because Perseus had a god for a father, some of the other gods gave him tools for the job, including Athena's reflective shield and the winged sandals of Hermes (Mercury in the Roman story). By flying over her and looking only at her reflection, Perseus was able to cut off Medusa's head without ever looking at her directly. Taking her head (which, conveniently, could still turn onlookers to stone even without being attached to her body) with him, Perseus continued on to other adventures.

He next came to a rocky seashore, where boasting had gotten another family into serious trouble with the gods. Queen Cassiopeia had dared to compare her own beauty to that of the Nereids, sea nymphs who were daughters of Poseidon (Neptune in Roman mythology), the god of the sea. Poseidon was so offended that he created a sea-monster named Cetus to devastate the kingdom. King Cepheus, Cassiopeia's beleaguered husband, consulted the oracle, who told him that he must sacrifice his beautiful daughter Andromeda to the monster.

When Perseus came along and found Andromeda chained to a rock near the sea, awaiting her fate, he rescued her by turning the monster to stone. (Scholars of mythology actually trace the essence of this story back to far-older legends from ancient Mesopotamia, in which the god-hero Marduk vanquishes a monster named Tiamat. Symbolically, a hero like Perseus or Marduk is usually associated with the Sun, the monster with the power of night, and the beautiful maiden with the fragile beauty of dawn, which the Sun releases after its nightly struggle with darkness.)

Many of the characters in these Greek legends can be found as constellations in the sky, not necessarily resembling their namesakes but serving as reminders of the story. For example, vain Cassiopeia is sentenced to be very close to the celestial pole, rotating perpetually around the sky and hanging upside down every winter. The ancients imagined Andromeda still chained to her rock (it is much easier to see the chain of stars than to recognize the beautiful maiden in this star grouping). Perseus is next to her with the head of Medusa swinging from his belt. Algol represents this gorgon head and has long been associated with evil and bad fortune in such tales. Some commentators have speculated that the star's change in brightness (which can be observed with the unaided eye) may have contributed to its unpleasant reputation, with the ancients regarding such a change as a sort of evil "wink."

Diameters of Eclipsing Binary Stars

We now turn back to the main thread of our story to discuss how all this can be used to measure the sizes of stars. The technique involves making a light curve of an eclipsing binary, a graph that plots how the brightness changes with time. Let us consider a hypothetical binary system in which the stars are very different in size, like those illustrated in Figure 18.11. To make life easy, we will assume that the orbit is viewed exactly edge-on.

Even though we cannot see the two stars separately in such a system, the light curve can tell us what is happening. When the smaller star just starts to pass behind the larger star (a point we call *first contact*), the brightness begins to drop. The eclipse becomes total (the smaller star is completely hidden) at the point called *second contact*. At the end of the total eclipse (*third contact*), the smaller star begins to emerge. When the smaller star has reached *last contact*, the eclipse is completely over.

To see how this allows us to measure diameters, look carefully at Figure 18.11. During the time interval between the first and second contacts, the smaller star has moved a distance equal to its own diameter. During the time interval from the first to third contacts, the smaller star has moved a distance equal to the diameter of the larger star. If the spectral lines of both stars are

visible in the spectrum of the binary, then the speed of the smaller star with respect to the larger one can be measured from the Doppler shift. But knowing the speed with which the smaller star is moving and how long it took to cover some distance can tell the span of that distance—in this case, the diameters of the stars. The speed multiplied by the time interval from the first to second contact gives the diameter of the smaller star. We multiply the speed by the time between the first and third contacts to get the diameter of the larger star.

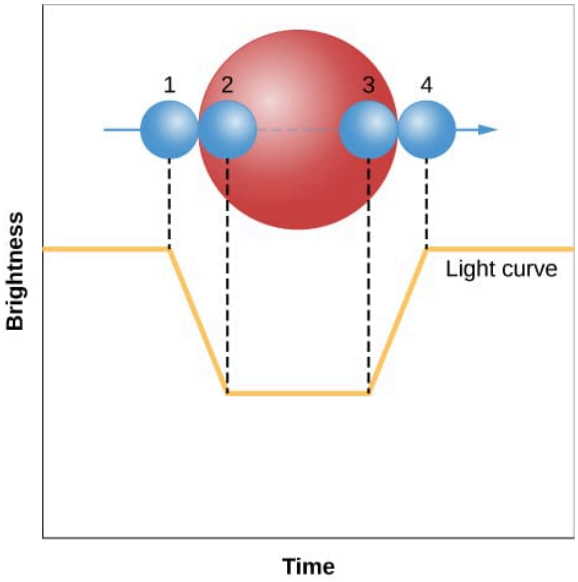


Figure 18.11 : Light Curve of an Edge-On Eclipsing Binary. Here we see the light curve of a hypothetical eclipsing binary star whose orbit we view exactly edge-on, in which the two stars fully eclipse each other. From the time intervals between contacts, it is possible to estimate the diameters of the two stars.

In actuality, the situation with eclipsing binaries is often a bit more complicated: orbits are generally not seen exactly edge-on, and the light from each star may be only partially blocked by the other. Furthermore, binary star orbits, just like the orbits of the planets, are ellipses, not circles. However, all these effects can be sorted out from very careful measurements of the light curve.

Link to Learning

The Eclipsing Binary Simulator allows you to explore how the eclipse timing can be used to determine the size of stars in a binary pair. Other traits can be explored as well, such as their mass, separation, and surface temperatures.

Using the Radiation Law to Get the Diameter

Another method for measuring star diameters makes use of the Stefan-Boltzmann law for the relationship between energy radiated and temperature (see Radiation and Spectra). In this method, the *energy flux* (energy emitted per second per square meter by a blackbody, like the Sun) is given by

$$F = \sigma T^4$$

where σ is a constant and *T* is the temperature. The surface area of a sphere (like a star) is given by

$$A = 4\pi R^2$$

The luminosity (*L*) of a star is then given by its surface area in square meters times the energy flux:

$$L = (A \times F)$$

Previously, we determined the masses of the two stars in the Sirius binary system. Sirius gives off 8200 times more energy than its fainter companion star, although both stars have nearly identical temperatures. The extremely large difference in luminosity is due to the difference in radius, since the temperatures and hence the energy fluxes for the two stars are nearly the same. To

determine the relative sizes of the two stars, we take the ratio of the corresponding luminosities:

$$\frac{L_{\text{Sirius}}}{L_{\text{companion}}} = \frac{(A_{\text{Sirius}} \times F_{\text{Sirius}})}{(A_{\text{companion}} \times F_{\text{companion}})}$$

$$= \frac{A_{\text{Sirius}}}{A_{\text{companion}}} = \frac{4\pi {R^2}_{\text{Sirius}}}{4\pi {R^2}_{\text{companion}}} = \frac{{R^2}_{\text{Sirius}}}{{R^2}_{\text{companion}}}$$

$$\frac{L_{\text{Sirius}}}{L_{\text{companion}}} = 8200 = \frac{{R^2}_{\text{Sirius}}}{{R^2}_{\text{companion}}}$$

Therefore, the relative sizes of the two stars can be found by taking the square root of the relative luminosity. Since $\sqrt{8200} = 91$, the radius of Sirius is 91 times larger than the radius of its faint companion.

The method for determining the radius shown here requires both stars be visible, which is not always the case.

Link to Learning

Use the Stellar Luminosity Simulato to explore the relationship between a star's surface temperature, luminosity, and radius. Move the sliders to see what happens. Try to make two stars with the same luminosity but different surface temperatures.

Stellar Diameters

The results of many stellar size measurements over the years have shown that most nearby stars are roughly the size of the Sun, with typical diameters of a million kilometers or so. Faint stars, as we might have expected, are generally smaller than more luminous stars. However, there are some dramatic exceptions to this simple generalization.

A few of the very luminous stars, those that are also red (indicating relatively low surface temperatures), turn out to be truly enormous. These stars are called, appropriately enough, giant stars or supergiant stars. An example is Betelgeuse, the second brightest star in the constellation of Orion and one of the dozen brightest stars in our sky. Its diameter, remarkably, is greater than 10 AU (1.5 *billion* kilometers!), large enough to fill the entire inner solar system almost as far out as Jupiter. In Stars from Adolescence to Old Age, we will look in detail at the evolutionary process that leads to the formation of such giant and supergiant stars.

Link to Learning

Watch this star size comparison video for a striking visual that highlights the size of stars versus planets and the range of sizes among stars.

Footnotes

- 3Fans of Batman comic books and movies will recognize that this name was given to an archvillain in the series.

This page titled 12.8: Diameters of Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 18.4: Diameters of Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.9: The H-R Diagram

Learning Objectives

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

- Identify the physical characteristics of stars that are used to create an H–R diagram, and describe how those characteristics vary among groups of stars
- Discuss the physical properties of most stars found at different locations on the H–R diagram, such as radius, and for main sequence stars, mass

In this chapter and Analyzing Starlight, we described some of the characteristics by which we might classify stars and how those are measured. These ideas are summarized in Table 18.2. We have also given an example of a relationship between two of these characteristics in the mass-luminosity relation. When the characteristics of large numbers of stars were measured at the beginning of the twentieth century, astronomers were able to begin a deeper search for patterns and relationships in these data.

Table 18.2: Measuring the Characteristics of Stars

Characteristic	Technique
Surface temperature	1. Determine the color (very rough). 2. Measure the spectrum and get the spectral type.
Chemical composition	Determine which lines are present in the spectrum.
Luminosity	Measure the apparent brightness and compensate for distance.
Radial velocity	Measure the Doppler shift in the spectrum.
Rotation	Measure the width of spectral lines.
Mass	Measure the period and radial velocity curves of spectroscopic binary stars.
Diameter	1. Measure the way a star's light is blocked by the Moon. 2. Measure the light curves and Doppler shifts for eclipsing binary stars.

To help understand what sorts of relationships might be found, let's look briefly at a range of data about human beings. If you want to understand humans by comparing and contrasting their characteristics—without assuming any previous knowledge of these strange creatures—you could try to determine which characteristics lead you in a fruitful direction. For example, you might plot the heights of a large sample of humans against their weights (which is a measure of their mass). Such a plot is shown in Figure 18.12 and it has some interesting features. In the way we have chosen to present our data, height increases upward, whereas weight increases to the left. Notice that humans are not randomly distributed in the graph. Most points fall along a sequence that goes from the upper left to the lower right.

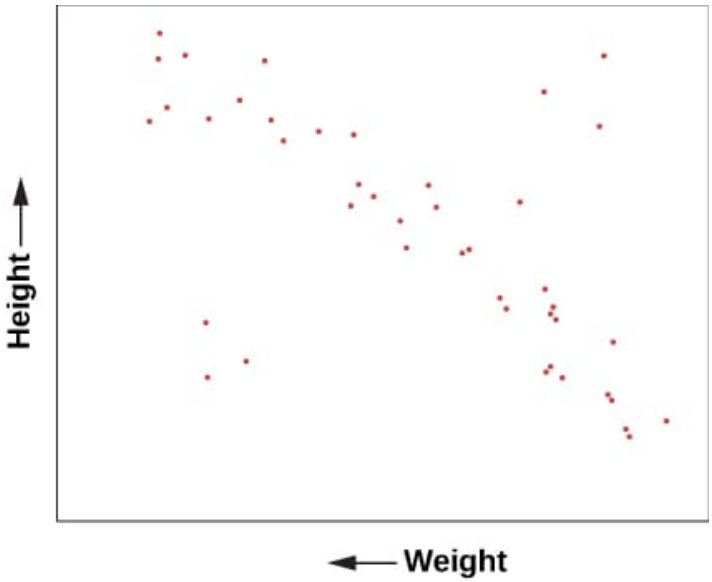


Figure 18.12 : Height versus Weight. The plot of the heights and weights of a representative group of human beings. Most points lie along a "main sequence" representing most people, but there are a few exceptions.

We can conclude from this graph that human height and weight are related. Generally speaking, taller human beings weigh more, whereas shorter ones weigh less. This makes sense if you are familiar with the structure of human beings. Typically, if we have bigger bones, we have more flesh to fill out our larger frame. It's not mathematically exact—there is a wide range of variation—but it's not a bad overall rule. And, of course, there are some dramatic exceptions. You occasionally see a short human who is very overweight and would thus be more to the bottom left of our diagram than the average sequence of people. Or you might have a very tall, skinny fashion model with great height but relatively small weight, who would be found near the upper right.

A similar diagram has been found extremely useful for understanding the lives of stars. In 1913, American astronomer Henry Norris Russell plotted the luminosities of stars against their spectral classes (a way of denoting their surface temperatures). This investigation, and a similar independent study in 1911 by Danish astronomer Ejnar Hertzsprung, led to the extremely important discovery that the temperature and luminosity of stars are related (Figure 18.13).

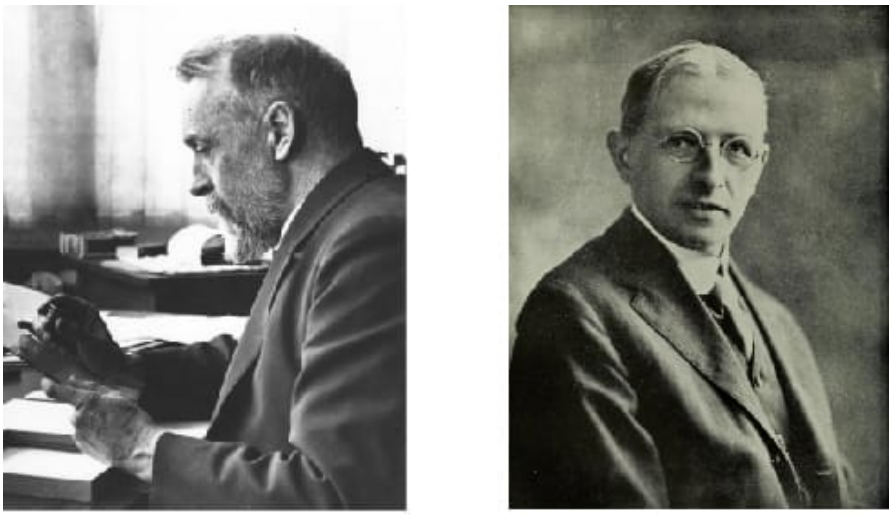


Figure 18.13 : Hertzsprung (1873–1967) and Russell (1877–1957). (a) Ejnar Hertzsprung and (b) Henry Norris Russell independently discovered the relationship between the luminosity and surface temperature of stars that is summarized in what is now called the H–R diagram.

Voyagers in Astronomy

Henry Norris Russell

When Henry Norris Russell graduated from Princeton University, his work had been so brilliant that the faculty decided to create a new level of honors degree beyond “summa cum laude” for him. His students later remembered him as a man whose thinking was three times faster than just about anybody else’s. His memory was so phenomenal, he could correctly quote an enormous number of poems and limericks, the entire Bible, tables of mathematical functions, and almost anything he had learned about astronomy. He was nervous, active, competitive, critical, and very articulate; he tended to dominate every meeting he attended. In outward appearance, he was an old-fashioned product of the nineteenth century who wore high-top black shoes and high starched collars, and carried an umbrella every day of his life. His 264 papers were enormously influential in many areas of astronomy.

Born in 1877, the son of a Presbyterian minister, Russell showed early promise. When he was 12, his family sent him to live with an aunt in Princeton so he could attend a top preparatory school. He lived in the same house in that town until his death in 1957 (interrupted only by a brief stay in Europe for graduate work). He was fond of recounting that both his mother and his maternal grandmother had won prizes in mathematics, and that he probably inherited his talents in that field from their side of the family.

Before Russell, American astronomers devoted themselves mainly to surveying the stars and making impressive catalogs of their properties, especially their spectra (as described in Analyzing Starlight. Russell began to see that interpreting the spectra of stars required a much more sophisticated understanding of the physics of the atom, a subject that was being developed by European physicists in the 1910s and 1920s. Russell embarked on a lifelong quest to ascertain the physical conditions inside stars from the clues in their spectra; his work inspired, and was continued by, a generation of astronomers, many trained by Russell and his collaborators.

Russell also made important contributions in the study of binary stars and the measurement of star masses, the origin of the solar system, the atmospheres of planets, and the measurement of distances in astronomy, among other fields. He was an influential teacher and popularizer of astronomy, writing a column on astronomical topics for *Scientific American* magazine for more than 40 years. He and two colleagues wrote a textbook for college astronomy classes that helped train astronomers and astronomy enthusiasts over several decades. That book set the scene for the kind of textbook you are now reading, which not only lays out the facts of astronomy but also explains how they fit together. Russell gave lectures around the country, often emphasizing the importance of understanding modern physics in order to grasp what was happening in astronomy.

Harlow Shapley, director of the Harvard College Observatory, called Russell “the dean of American astronomers.” Russell was certainly regarded as the leader of the field for many years and was consulted on many astronomical problems by colleagues from around the world. Today, one of the highest recognitions that an astronomer can receive is an award from the American Astronomical Society called the Russell Prize, set up in his memory.

Features of the H–R Diagram

Following Hertzsprung and Russell, let us plot the temperature (or spectral class) of a selected group of nearby stars against their luminosity and see what we find (Figure 18.14). Such a plot is frequently called the *Hertzsprung–Russell diagram*, abbreviated H–R diagram. It is one of the most important and widely used diagrams in astronomy, with applications that extend far beyond the purposes for which it was originally developed more than a century ago.

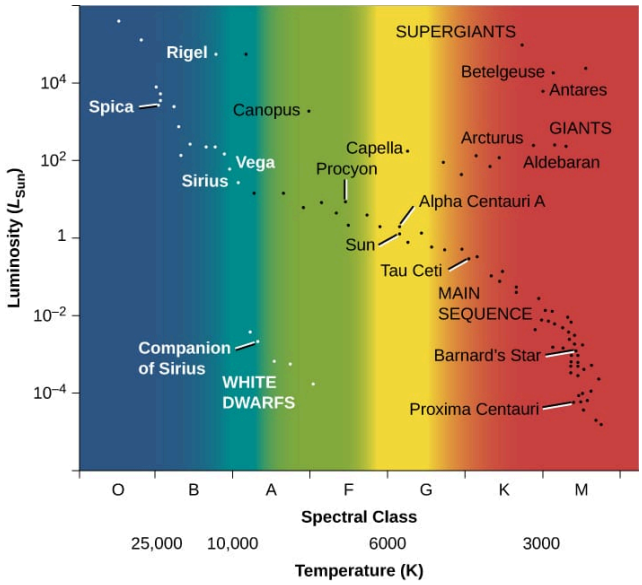


Figure 18.14 : H–R Diagram for a Selected Sample of Stars. In such diagrams, luminosity is plotted along the vertical axis. Along the horizontal axis, we can plot either temperature or spectral type (also sometimes called spectral class). Several of the brightest stars are identified by name. Most stars fall on the main sequence.

It is customary to plot H–R diagrams in such a way that temperature increases toward the left and luminosity toward the top. Notice the similarity to our plot of height and weight for people (Figure 18.12). Stars, like people, are not distributed over the diagram at random, as they would be if they exhibited all combinations of luminosity and temperature. Instead, we see that the stars cluster into certain parts of the H–R diagram. The great majority are aligned along a narrow sequence running from the upper left (hot, highly luminous) to the lower right (cool, less luminous). This band of points is called the main sequence. It represents a relationship between *temperature* and *luminosity* that is followed by most stars. We can summarize this relationship by saying that hotter stars are more luminous than cooler ones.

A number of stars, however, lie above the main sequence on the H–R diagram, in the upper-right region, where stars have low temperature and high luminosity. How can a star be at once cool, meaning each square meter on the star does not put out all that much energy, and yet very luminous? The only way is for the star to be enormous—to have so many square meters on its surface that the *total* energy output is still large. These stars must be *giants* or *supergiants*, the stars of huge diameter we discussed earlier.

There are also some stars in the lower-left corner of the diagram, which have high temperature and low luminosity. If they have high surface temperatures, each square meter on that star puts out a lot of energy. How then can the overall star be dim? It must be that it has a very small total surface area; such stars are known as white dwarfs (white because, at these high temperatures, the colors of the electromagnetic radiation that they emit blend together to make them look bluish-white). We will say more about these puzzling objects in a moment. Figure 18.15 is a schematic H–R diagram for a large sample of stars, drawn to make the different types more apparent.

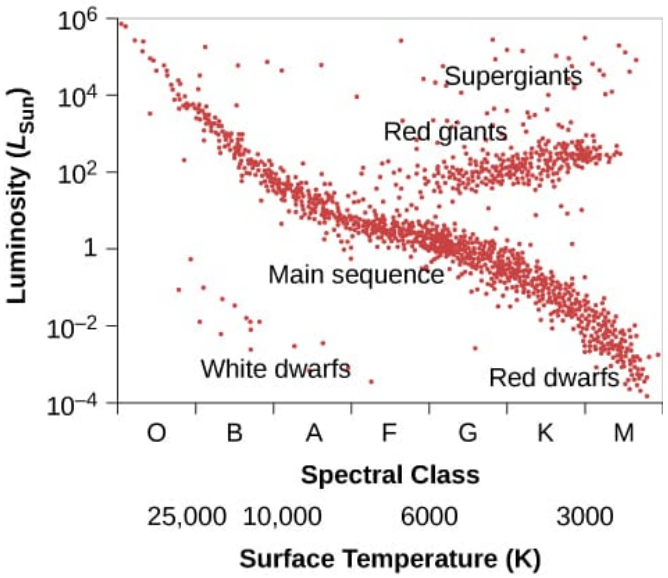


Figure 18.15 : Schematic H–R Diagram for Many Stars. Ninety percent of all stars on such a diagram fall along a narrow band called the main sequence. A minority of stars are found in the upper right; they are both cool (and hence red) and bright, and must be giants. Some stars fall in the lower left of the diagram; they are both hot and dim, and must be white dwarfs.

Now, think back to our discussion of star surveys. It is difficult to plot an H–R diagram that is truly representative of all stars because most stars are so faint that we cannot see those outside our immediate neighborhood. The stars plotted in Figure 18.14 were selected because their distances are known. This sample omits many intrinsically faint stars that are nearby but have not had their distances measured, so it shows fewer faint main-sequence stars than a "fair" diagram would. To be truly representative of the stellar population, an H–R diagram should be plotted for all stars within a certain distance. Unfortunately, our knowledge is reasonably complete only for stars within 10 to 20 light-years of the Sun, among which there are no giants or supergiants. Still, from many surveys (and more can now be done with new, more powerful telescopes), we estimate that about 90% of the true stars overall (excluding brown dwarfs) in our part of space are main-sequence stars, about 10% are white dwarfs, and fewer than 1% are giants or supergiants.

These estimates can be used directly to understand the lives of stars. Permit us another quick analogy with people. Suppose we survey people just like astronomers survey stars, but we want to focus our attention on the location of young people, ages 6 to 18 years. Survey teams fan out and take data about where such youngsters are found at all times during a 24-hour day. Some are found in the local pizza parlor, others are asleep at home, some are at the movies, and many are in school. After surveying a very large number of young people, one of the things that the teams determine is that, averaged over the course of the 24 hours, one-third of all youngsters are found in school.

How can they interpret this result? Does it mean that two-thirds of students are truants and the remaining one-third spend all their time in school? No, we must bear in mind that the survey teams counted youngsters throughout the full 24-hour day. Some survey teams worked at night, when most youngsters were at home asleep, and others worked in the late afternoon, when most youngsters were on their way home from school (and more likely to be enjoying a pizza). If the survey was truly representative, we *can* conclude, however, that if an average of one-third of all youngsters are found in school, then humans ages 6 to 18 years must spend about one-third of *their time* in school.

We can do something similar for stars. We find that, on average, 90% of all stars are located on the main sequence of the H–R diagram. If we can identify some activity or life stage with the main sequence, then it follows that stars must spend 90% of their lives in that activity or life stage.

Understanding the Main Sequence

In The Sun: A Nuclear Powerhouse, we discussed the Sun as a representative star. We saw that what stars such as the Sun "do for a living" is to convert protons into helium deep in their interiors via the process of nuclear fusion, thus producing energy. The fusion of protons to helium is an excellent, long-lasting source of energy for a star because the bulk of every star consists of hydrogen atoms, whose nuclei are protons.

Our computer models of how stars evolve over time show us that a typical star will spend about 90% of its life fusing the abundant hydrogen in its core into helium. This then is a good explanation of why 90% of all stars are found on the main sequence in the H–R diagram. But if all the stars on the main sequence are doing the same thing (fusing hydrogen), why are they distributed along a sequence of points? That is, why do they differ in luminosity and surface temperature (which is what we are plotting on the H–R diagram)?

To help us understand how main-sequence stars differ, we can use one of the most important results from our studies of model stars. Astrophysicists have been able to show that the structure of stars that are in equilibrium and derive all their energy from nuclear fusion is completely and uniquely determined by just two quantities: the *total mass* and the *composition* of the star. This fact provides an interpretation of many features of the H–R diagram.

Imagine a cluster of stars forming from a cloud of interstellar "raw material" whose chemical composition is similar to the Sun's. (We'll describe this process in more detail in The Birth of Stars and Discovery of Planets outside the Solar System, but for now, the details will not concern us.) In such a cloud, all the clumps of gas and dust that become stars begin with the same chemical composition and differ from one another only in mass. Now suppose that we compute a model of each of these stars for the time at which it becomes stable and derives its energy from nuclear reactions, but before it has time to alter its composition appreciably as a result of these reactions.

The models calculated for these stars allow us to determine their luminosities, temperatures, and sizes. If we plot the results from the models—one point for each model star—on the H–R diagram, we get something that looks just like the main sequence we saw for real stars.

And here is what we find when we do this. The model stars with the largest masses are the hottest and most luminous, and they are located at the upper left of the diagram.

The least-massive model stars are the coolest and least luminous, and they are placed at the lower right of the plot. The other model stars all lie along a line running diagonally across the diagram. In other words, *the main sequence turns out to be a sequence of stellar masses*.

This makes sense if you think about it. The most massive stars have the most gravity and can thus compress their centers to the greatest degree. This means they are the hottest inside and the best at generating energy from nuclear reactions deep within. As a result, they shine with the greatest luminosity and have the hottest surface temperatures. The stars with lowest mass, in turn, are the coolest inside and least effective in generating energy. Thus, they are the least luminous and wind up being the coolest on the surface. Our Sun lies somewhere in the middle of these extremes (as you can see in Figure 18.14). The characteristics of representative main-sequence stars (excluding brown dwarfs, which are not true stars) are listed in Table 18.3.

Table 18.3: Characteristics of Main-Sequence Stars

Spectral Type	Mass (Sun = 1)	Luminosity (Sun = 1)	Temperature	Radius (Sun = 1)
O5	40	7×10^5	40,000 K	18
B0	16	2.7×10^5	28,000 K	7
A0	3.3	55	10,000 K	2.5
F0	1.7	5	7500 K	1.4
G0	1.1	1.4	6000 K	1.1
K0	0.8	0.35	5000 K	0.8
M0	0.4	0.05	3500 K	0.6

Note that we've seen this 90% figure come up before. This is exactly what we found earlier when we examined the mass-luminosity relation (Figure 18.9). We observed that 90% of all stars seem to follow the relationship; these are the 90% of all stars that lie on the main sequence in our H–R diagram. Our models and our observations agree.

What about the other stars on the H–R diagram—the giants and supergiants, and the white dwarfs? As we will see in the next few chapters, these are what main-sequence stars turn into as they age: They are the later stages in a star's life. As a star consumes its nuclear fuel, its source of energy changes, as do its chemical composition and interior structure. These changes cause the star to alter its luminosity and surface temperature so that it no longer lies on the main sequence on our diagram. Because stars

spend much less time in these later stages of their lives, we see fewer stars in those regions of the H–R diagram.

Extremes of Stellar Luminosities, Diameters, and Densities

We can use the H–R diagram to explore the extremes in size, luminosity, and density found among the stars. Such extreme stars are not only interesting to fans of the *Guinness Book of World Records*; they can teach us a lot about how stars work. For example, we saw that the most massive main-sequence stars are the most luminous ones. We know of a few extreme stars that are a million times more luminous than the Sun, with masses that exceed 100 times the Sun's mass. These superluminous stars, which are at the upper left of the H–R diagram, are exceedingly hot, very blue stars of spectral type O. These are the stars that would be the most conspicuous at vast distances in space.

The cool supergiants in the upper corner of the H–R diagram are as much as 10,000 times as luminous as the Sun. In addition, these stars have diameters very much larger than that of the Sun. As discussed above, some supergiants are so large that if the solar system could be centered in one, the star's surface would lie beyond the orbit of Mars (see Figure 18.16). We will have to ask, in coming chapters, what process can make a star swell up to such an enormous size, and how long these "swollen" stars can last in their distended state.

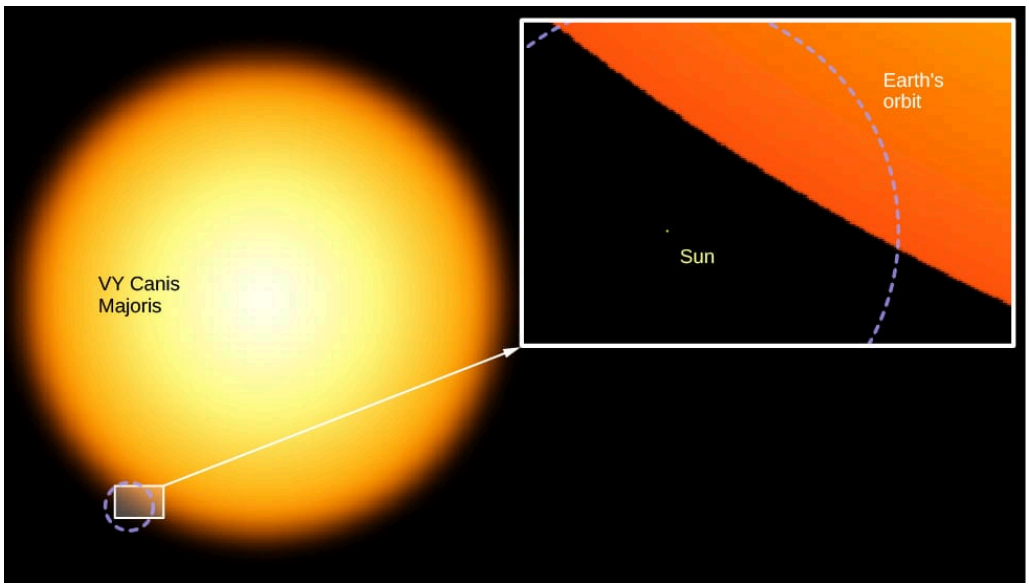


Figure 18.16 : The Sun and a Supergiant. Here you see how small the Sun looks in comparison to one of the largest known stars: VY Canis Majoris, a supergiant.

In contrast, the very common red, cool, low-luminosity stars at the lower end of the main sequence are much smaller and more compact than the Sun. An example of such a red dwarf is Ross 614B, with a surface temperature of 2700 K and only 1/2000 of the Sun's luminosity. We call such a star a dwarf because its diameter is only 1/10 that of the Sun. A star with such a low luminosity also has a low mass (about 1/12 that of the Sun). This combination of mass and diameter means that it is so compressed that the star has an average density about 80 times that of the Sun. Its density must be higher, in fact, than that of any known solid found on the surface of Earth. (Despite this, the star is made of gas throughout because its center is so hot.)

The faint, red, main-sequence stars are not the stars of the most extreme densities, however. The white dwarfs, at the lower-left corner of the H–R diagram, have densities many times greater still.

The White Dwarfs

The first white dwarf star was detected in 1862. Called Sirius B, it forms a binary system with Sirius A, the brightest-appearing star in the sky. It eluded discovery and analysis for a long time because its faint light tends to be lost in the glare of nearby Sirius A (Figure 18.17). (Since Sirius is often called the Dog Star—being the brightest star in the constellation of Canis Major, the big dog—Sirius B is sometimes nicknamed the Pup.)

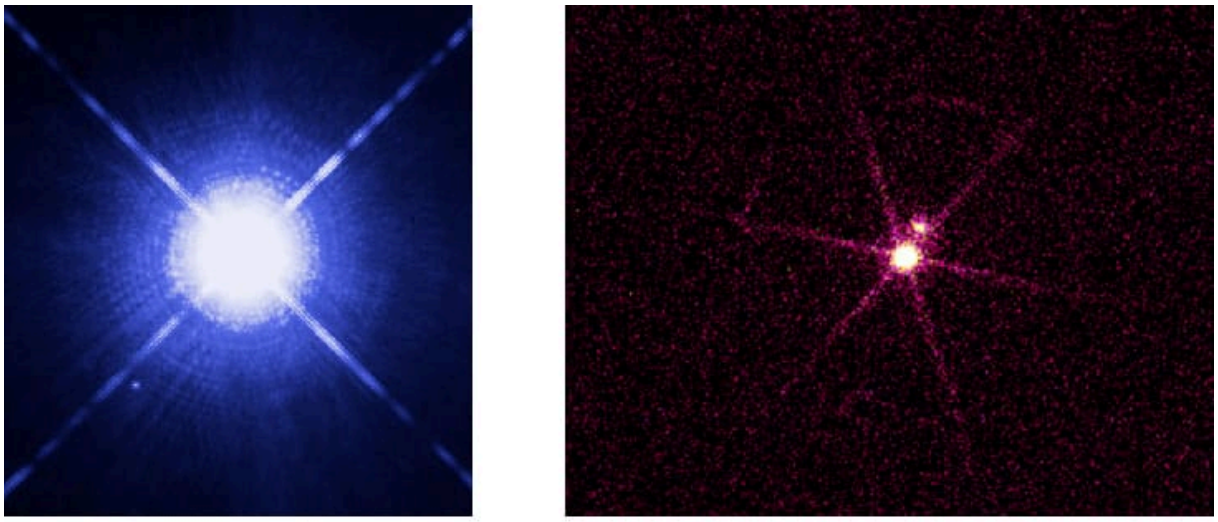


Figure 18.17 : Two Views of Sirius and Its White Dwarf Companion. (a) The (visible light) image, taken with the Hubble Space Telescope, shows bright Sirius A, and, below it and off to its left, faint Sirius B. (b) This image of the Sirius star system was taken with the Chandra X-Ray Telescope. Now, the bright object is the white dwarf companion, Sirius B. Sirius A is the faint object above it; what we are seeing from Sirius is probably not actually X-ray radiation but rather ultraviolet light that has leaked into the detector. Note that the ultraviolet intensities of these two objects are completely reversed from the situation in visible light because Sirius B is hotter and emits more higher-frequency radiation. (credit a: modification of work by NASA, H.E. Bond and E. Nelan (Space Telescope Science Institute), M. Barstow and M. Burleigh (University of Leicester) and J.B. Holberg (University of Arizona); credit b: modification of work by NASA/SAO/CXC)

We have now found thousands of white dwarfs. Table 18.1 shows that about 7% of the true stars (spectral types O–M) in our local neighborhood are white dwarfs. A good example of a typical white dwarf is the nearby star 40 Eridani B. Its surface temperature is a relatively hot 12,000 K, but its luminosity is only 1/275 L_{Sun}. Calculations show that its radius is only 1.4% of the Sun's, or about the same as that of Earth, and its volume is 2.5×10^{-6} that of the Sun. Its mass, however, is 0.57 times the Sun's mass, just a little more than half. To fit such a substantial mass into so tiny a volume, the star's density must be about 210,000 times the density of the Sun, or more than 300,000 g/cm^3. A teaspoonful of this material would have a mass of some 1.6 tons! At such enormous densities, matter cannot exist in its usual state; we will examine the particular behavior of this type of matter in The Death of Stars. For now, we just note that white dwarfs are dying stars, reaching the end of their productive lives and ready for their stories to be over.

The British astrophysicist (and science popularizer) Arthur Eddington (1882–1944) described the first known white dwarf this way:

> The message of the companion of Sirius, when decoded, ran: "I am composed of material three thousand times denser than anything you've ever come across. A ton of my material would be a little nugget you could put in a matchbox." What reply could one make to something like that? Well, the reply most of us made in 1914 was, "Shut up; don't talk nonsense."

Today, however, astronomers not only accept that stars as dense as white dwarfs exist but (as we will see) have found even denser and stranger objects in their quest to understand the evolution of different types of stars.

Link to Learning

Use this interactive H-R diagram to explore the sizes of stars based on their placement. Overlay the closest stars and the brightest stars, and consider where they lie on the figure. How many stars are among both the closest and brightest stars?

This page titled 12.9: The H-R Diagram is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 18.5: The H-R Diagram by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.10: Star Clusters

Learning Objectives

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

- Explain how star clusters help us understand the stages of stellar evolution
- List the different types of star clusters and describe how they differ in number of stars, structure, and age
- Explain why the chemical composition of globular clusters is different from that of open clusters

The preceding description of stellar evolution is based on calculations. However, no star completes its main-sequence lifetime or its evolution to a red giant quickly enough for us to observe these structural changes as they happen. Fortunately, nature has provided us with an indirect way to test our calculations.

Instead of observing the evolution of a single star, we can look at a group or *cluster* of stars. We look for a group of stars that is very close together in space, held together by gravity, often moving around a common center. Then it is reasonable to assume that the individual stars in the group all formed at nearly the same time, from the same cloud, and with the same composition. We expect that these stars will differ only in mass. And their masses determine how quickly they go through each stage of their lives.

Since stars with higher masses evolve more quickly, we can find clusters in which massive stars have already completed their main-sequence phase of evolution and become red giants, while stars of lower mass in the same cluster are still on the main sequence, or even—if the cluster is very young—undergoing pre-main-sequence gravitational contraction. We can see many stages of stellar evolution among the members of a single cluster, and we can see whether our models can explain why the H–R diagrams of clusters of different ages look the way they do.

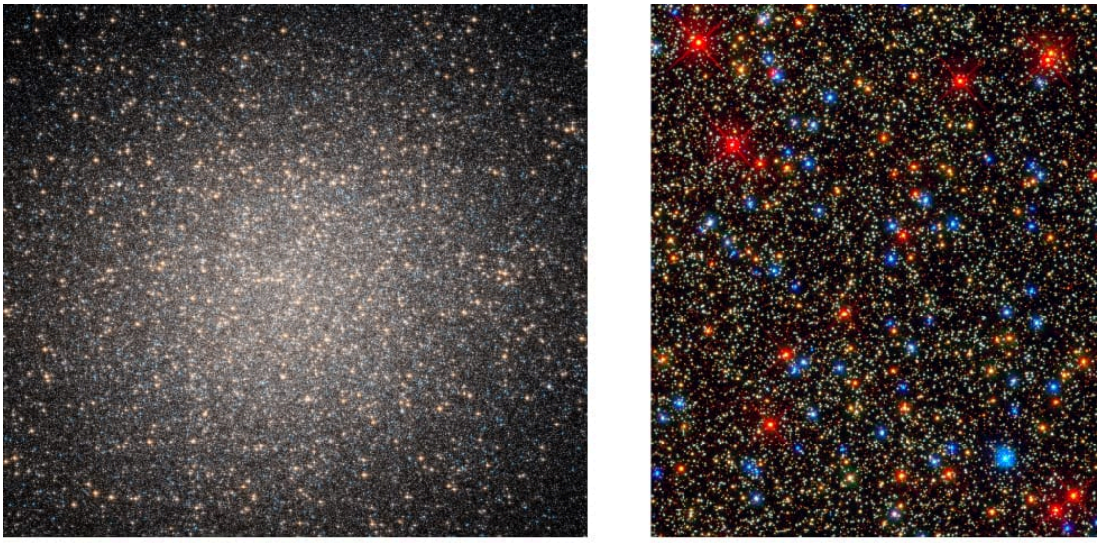
The three basic types of clusters astronomers have discovered are globular clusters, open clusters, and stellar associations. Their properties are summarized in Table 22.3. As we will see in the next section of this chapter, globular clusters contain only very old stars, whereas open clusters and associations contain young stars.

Table 22.3: Characteristics of Star Clusters

Characteristic	Globular Clusters	Open Clusters	Associations
Number in the Galaxy	150	Thousands	Thousands
Location in the Galaxy	Halo and central bulge	Disk (and spiral arms)	Spiral arms
Diameter (in light-years)	50–450	<30	100–500
Mass M_{Sun}	10^4–10^6	10^2–10^3	10^2–10^3
Number of stars	10^4–10^6	50–1000	10^2–10^4
Color of brightest stars	Red	Red or blue	Blue
Luminosity of cluster (L_{Sun})	10^4–10^6	10^2–10^6	10^4–10^7
Typical ages	Billions of years	A few hundred million years to, in the case of unusually large clusters, more than a billion years	Up to about 10^7 years

Globular Clusters

Globular clusters were given this name because they are nearly symmetrical round systems of, typically, hundreds of thousands of stars. The most massive globular cluster in our own Galaxy is Omega Centauri, which is about 16,000 light-years away and contains several million stars (Figure 22.6). Note that the brightest stars in this cluster, which are red giants that have already completed the main-sequence phase of their evolution, are red-orange in color. These stars have typical surface temperatures around 4000 K. As we will see, globular clusters are among the oldest parts of our Milky Way Galaxy.



(a) (b)

Figure 22.6 : Omega Centauri. (a) Located at about 16,000 light-years away, Omega Centauri is the most massive globular cluster in our Galaxy. It contains several million stars. (b) This image, taken with the Hubble Space Telescope, zooms in near the center of Omega Centauri. The image is about 6.3 light-years wide. The most numerous stars in the image, which are yellow-white in color, are main-sequence stars similar to our Sun. The brightest stars are red giants that have begun to exhaust their hydrogen fuel and have expanded to about 100 times the diameter of our Sun. The blue stars have started helium fusion. (credit a: modification of work by NASA, ESA and the Hubble Heritage Team (STScI/AURA); credit b: modification of work by NASA, ESA, and the Hubble SM4 ERO Team)

What would it be like to live inside a globular cluster? In the dense central regions, the stars would be roughly a million times closer together than in our own neighborhood. If Earth orbited one of the inner stars in a globular cluster, the nearest stars would be light-months, not light-years, away. They would still appear as points of light, but would be brighter than any of the stars we see in our own sky. The Milky Way would probably be difficult to see through the bright haze of starlight produced by the cluster.

About 150 globular clusters are known in our Galaxy. Most of them are in a spherical halo (or cloud) surrounding the flat disk formed by the majority of our Galaxy's stars. All the globular clusters are very far from the Sun, and some are found at distances of 60,000 light-years or more from the main disk of the Milky Way. The diameters of globular star clusters range from 50 light-years to more than 450 light-years.

Open Clusters

Open clusters are found in the disk of the Galaxy. They have a range of ages, some as old as, or even older than, our Sun. The youngest open clusters are still associated with the interstellar matter from which they formed. Open clusters are smaller than globular clusters, usually having diameters of less than 30 light-years, and they typically contain only several dozen to several hundreds of stars (Figure 22.7). The stars in open clusters usually appear well separated from one another, even in the central regions, which explains why they are called "open." Our Galaxy contains thousands of open clusters, but we can see only a small fraction of them. Interstellar dust, which is also concentrated in the disk, dims the light of more distant clusters so much that they are undetectable.



Figure 22.7 : Jewel Box (NGC 4755). This open cluster of young, bright stars is about 6400 light-years away from the Sun. Note the contrast in color between the bright yellow supergiant and the hot blue main-sequence stars. The name comes from John Herschel's nineteenth-century description of it as "a casket of variously colored precious stones." (credit: ESO/Y. Beletsky)

Although the individual stars in an open cluster can survive for billions of years, they typically remain together as a cluster for only a few million years, or at most, a few hundred million years. There are several reasons for this. In small open clusters, the average speed of the member stars within the cluster may be higher than the cluster's escape velocity,[1] and the stars will gradually "evaporate" from the cluster. Close encounters of member stars may also increase the velocity of one of the members beyond the escape velocity. Every few hundred million years or so, the cluster may have a close encounter with a giant molecular cloud, and the gravitational force exerted by the cloud may tear the cluster apart.

Several open clusters are visible to the unaided eye. Most famous among them is the Pleiades (Figure 20.13), which appears as a tiny group of six stars (some people can see even more than six, and the Pleiades is sometimes called the Seven Sisters). This cluster is arranged like a small dipping spoon and is seen in the constellation of Taurus, the bull. A good pair of binoculars shows dozens of stars in the cluster, and a telescope reveals hundreds. (A car company, Subaru, takes its name from the Japanese term for this cluster; you can see the star group on the Subaru logo.)

The Hyades is another famous open cluster in the constellation of Taurus the bull. To the naked eye, it appears as a V-shaped group of faint stars marking the face of the bull. Telescopes show that Hyades actually contains more than 200 stars.

Stellar Associations

An association is a group of extremely young stars, typically containing 5 to 50 hot, bright O and B stars scattered over a region of space some 100–500 light-years in diameter. As an example, most of the stars in the constellation Orion form one of the nearest stellar associations. Associations also contain hundreds to thousands of low-mass stars, but these are much fainter and less conspicuous. The presence of really hot, luminous stars indicates that star formation in the association has occurred in the last million years or so. Since O stars go through their entire lives in only about a million years, they would not still be around unless star formation has occurred recently. It is therefore not surprising that associations are found in regions rich in the gas and dust required to form new stars. It's like a brand new building still surrounded by some of the construction materials used to build it and with the landscape still showing signs of construction. On the other hand, because associations, like ordinary open clusters, lie in regions occupied by dusty interstellar matter, many are hidden from our view.

Footnotes

- 1Escape velocity is the speed needed to overcome the gravity of some object or group of objects. The rockets we send up from Earth, for example, must travel faster than the escape velocity of our planet to be able to get to other worlds.

This page titled 12.10: Star Clusters is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 22.3: Star Clusters by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

12.11: Checking Out the Theory

Learning Objectives

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

- Explain how the H–R diagram of a star cluster can be related to the cluster's age and the stages of evolution of its stellar members
- Describe how the main-sequence turnoff of a cluster reveals its age

In the previous section, we indicated that open clusters are younger than globular clusters, and associations are typically even younger. In this section, we will show how we determine the ages of these star clusters. The key observation is that the stars in these different types of clusters are found in different places in the H–R diagram, and we can use their locations in the diagram in combination with theoretical calculations to estimate how long they have lived.

H–R Diagrams of Young Clusters

What does theory predict for the H–R diagram of a cluster whose stars have recently condensed from an interstellar cloud? Remember that at every stage of evolution, massive stars evolve more quickly than their lower-mass counterparts. After a few million years ("recently" for astronomers), the most massive stars should have completed their contraction phase and be on the main sequence, while the less massive ones should be off to the right, still on their way to the main sequence. These ideas are illustrated in Figure 22.8, which shows the H–R diagram calculated by R. Kippenhahn and his associates at Munich University for a hypothetical cluster with an age of 3 million years.

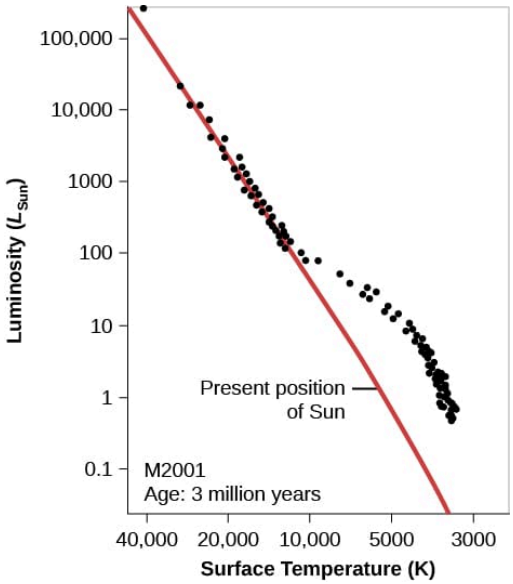


Figure 22.8 : Young Cluster H–R Diagram. We see an H–R diagram for a hypothetical young cluster with an age of 3 million years. Note that the high-mass (high-luminosity) stars have already arrived at the main-sequence stage of their lives, while the lower-mass (lower-luminosity) stars are still contracting toward the zero-age main sequence (the red line) and are not yet hot enough to derive all of their energy from the fusion of hydrogen.

There are real star clusters that fit this description. The first to be studied (in about 1950) was NGC 2264, which is still associated with the region of gas and dust from which it was born (Figure 22.9).



Figure 22.9 : Young Cluster NGC 2264. Located about 2600 light-years from us, this region of newly formed stars, known as the Christmas Tree Cluster, is a complex mixture of hydrogen gas (which is ionized by hot embedded stars and shown in red), dark obscuring dust lanes, and brilliant young stars. The image shows a scene about 30 light-years across. (credit: ESO)

The NGC 2264 cluster's H–R diagram is shown in Figure 22.10. The cluster in the middle of the Orion Nebula (shown in Figure 21.4 and Figure 21.5) is in a similar stage of evolution.

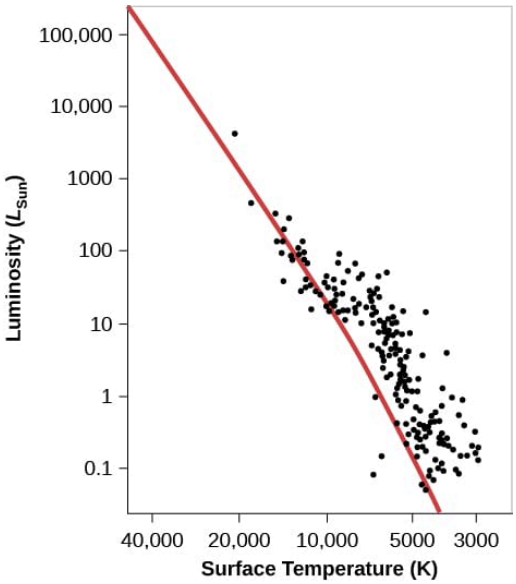


Figure 22.10 : NGC 2264 H–R Diagram. Compare this H–R diagram to that in Figure 22.8; although the points scatter a bit more here, the theoretical and observational diagrams are remarkably, and satisfyingly, similar.

As clusters get older, their H–R diagrams begin to change. After a short time (less than a million years after they reach the main sequence), the most massive stars use up the hydrogen in their cores and evolve off the main sequence to become red giants and supergiants. As more time passes, stars of lower mass begin to leave the main sequence and make their way to the upper right of the H–R diagram.

Link to Learning

To see the evolution of a star cluster in a dwarf galaxy, you can watch this brief animation of how its H–R diagram changes.

Figure 22.11 is a photograph of NGC 3293, a cluster that is about 10 million years old. The dense clouds of gas and dust are gone. One massive star has evolved to become a red giant and stands out as an especially bright orange member of the cluster.



Figure 22.11 : NGC 3293. All the stars in an open star cluster like NGC 3293 form at about the same time. The most massive stars, however, exhaust their nuclear fuel more rapidly and hence evolve more quickly than stars of low mass. As stars evolve, they become redder. The bright orange star in NGC 3293 is the member of the cluster that has evolved most rapidly. (credit: ESO/G. Beccari)

Figure 22.12 shows the H–R diagram of the open cluster M41, which is roughly 100 million years old; by this time, a significant number of stars have moved off to the right and become red giants. Note the gap that appears in this H–R diagram between the stars near the main sequence and the red giants. A gap does not necessarily imply that stars avoid a region of certain temperatures and luminosities. In this case, it simply represents a domain of temperature and luminosity through which stars evolve very quickly. We see a gap for M41 because at this particular moment, we have not caught a star in the process of scurrying across this part of the diagram.

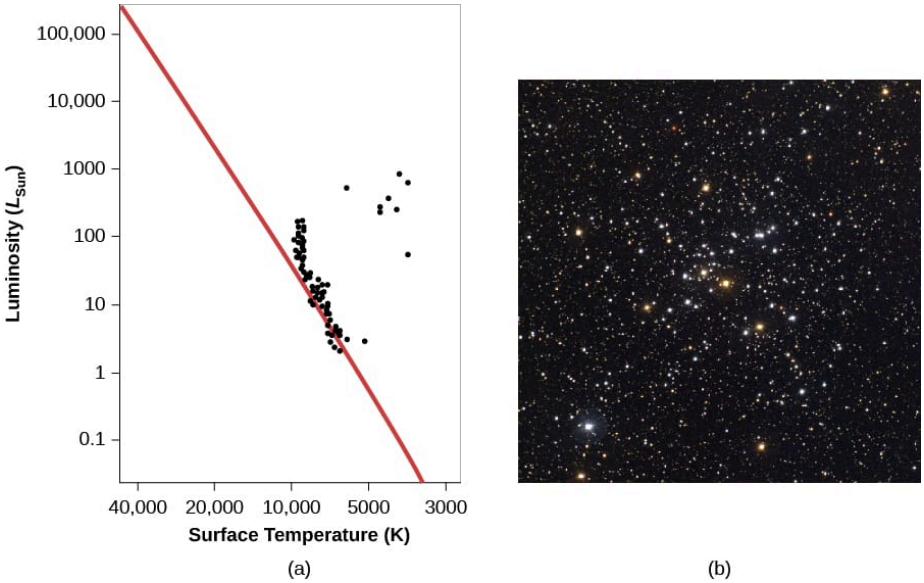


Figure 22.12 : Cluster M41. (a) Cluster M41 is older than NGC 2264 (see Figure 22.10) and contains several red giants. Some of its more massive stars are no longer close to the zero-age main sequence (red line). (b) This ground-based photograph shows the open cluster M41. Note that it contains several orange-color stars. These are stars that have exhausted hydrogen in their centers, and have swelled up to become red giants. (credit b: modification of work by NOAO/AURA/NSF)

H–R Diagrams of Older Clusters

After 4 billion years have passed, many more stars, including stars that are only a few times more massive than the Sun, have left the main sequence (Figure 22.13). This means that no stars are left near the top of the main sequence; only the low-mass stars near the bottom remain. The older the cluster, the lower the point on the main sequence (and the lower the mass of the stars) where stars begin to move toward the red giant region. The location in the H–R diagram where the stars have begun to leave the main sequence is called the main-sequence turnoff.

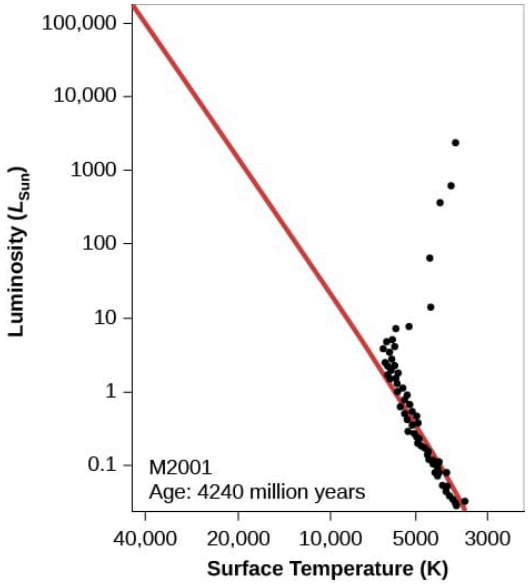


Figure 22.13 : H–R Diagram for an Older Cluster. We see the H–R diagram for a hypothetical older cluster at an age of 4.24 billion years. Note that most of the stars on the upper part of the main sequence have turned off toward the red-giant region. And the most massive stars in the cluster have already died and are no longer on the diagram.

The oldest clusters of all are the globular clusters. Figure 22.14 shows the H–R diagram of globular cluster 47 Tucanae. Notice that the luminosity and temperature scales are different from those of the other H–R diagrams in this chapter. In Figure 22.13, for example, the luminosity scale on the left side of the diagram goes from 0.1 to 100,000 times the Sun's luminosity. But in Figure 22.14, the luminosity scale has been significantly reduced in extent. So many stars in this old cluster have had time to turn off the main sequence that only the very bottom of the main sequence remains.

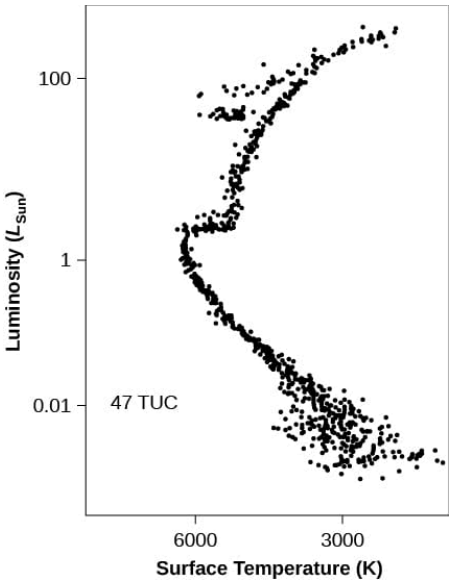


Figure 22.14 : Cluster 47 Tucanae. This H–R diagram is for the globular cluster 47. Note that the scale of luminosity differs from that of the other H–R diagrams in this chapter. We are only focusing on the lower portion of the main sequence, the only part where stars still remain in this old cluster.

Link to Learning

Check out this brief NASA video with a 3-D visualization of how an H–R diagram is created for the globular cluster Omega Centauri.

Just how old are the different clusters we have been discussing? To get their actual ages (in years), we must compare the appearances of our *calculated* H–R diagrams of different ages to *observed* H–R diagrams of real clusters. In practice, astronomers use the position at the top of the main sequence (that is, the luminosity at which stars begin to move off the main sequence to become red giants) as a measure of the age of a cluster (the main-sequence turnoff we discussed previously). For example, we can compare the luminosities of the brightest stars that are still on the main sequence in Figure 22.10 and Figure 22.13.

Using this method, some associations and open clusters turn out to be as young as 1 million years old, while others are several hundred million years old. Once all of the interstellar matter surrounding a cluster has been used to form stars or has dispersed and moved away from the cluster, star formation ceases, and stars of progressively lower mass move off the main sequence, as shown in Figure 22.10, Figure 22.12, and Figure 22.13.

To our surprise, even the youngest of the globular clusters in our Galaxy are found to be older than the oldest open cluster. All of the globular clusters have main sequences that turn off at a luminosity less than that of the Sun. Star formation in these crowded systems ceased billions of years ago, and no new stars are coming on to the main sequence to replace the ones that have turned off (see Figure 22.15).

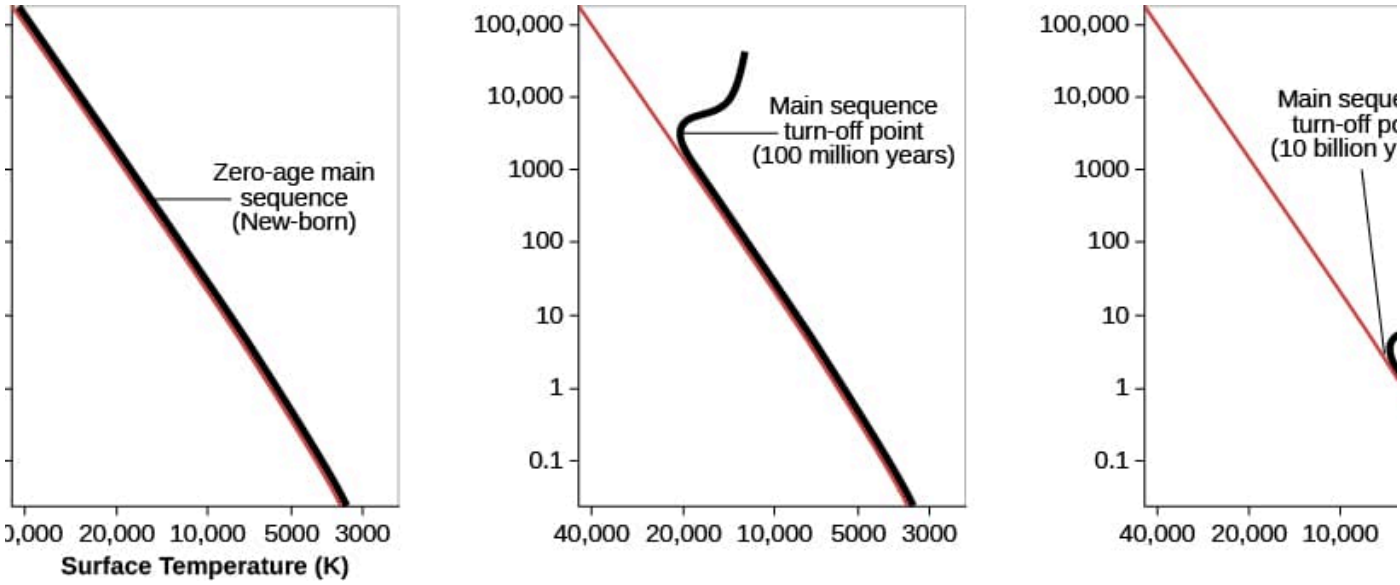


Figure 22.15 : H–R Diagrams for Clusters of Different Ages. This sketch shows how the turn-off point from the main sequence gets lower as we make H–R diagrams for clusters that are older and older.

Indeed, the globular clusters are the oldest structures in our Galaxy (and in other galaxies as well). The youngest have ages of about 11 billion years and some appear to be even older. Since these are the oldest objects we know of, this estimate is one of the best limits we have on the age of the universe itself—it must be at least 11 billion years old. We will return to the fascinating question of determining the age of the entire universe in the chapter on The Big Bang.

This page titled 12.11: Checking Out the Theory is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 22.4: Checking Out the Theory by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

13: STELLAR EVOLUTION

13.1: Star Formation
13.2: The H-R and the Study of Stellar Evolution
13.3: Evolution from the Main Sequence to Red Giants
13.4: Further Evolution of Stars
13.5: The Evolution of More Massive Stars

This page titled 13: Stellar Evolution is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

13.1: Star Formation

Learning Objectives

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

- Identify the sometimes-violent processes by which parts of a molecular cloud collapse to produce stars
- Recognize some of the structures seen in images of molecular clouds like the one in Orion
- Explain how the environment of a molecular cloud enables the formation of stars
- Describe how advancing waves of star formation cause a molecular cloud to evolve

As we begin our exploration of how stars are formed, let's review some basics about stars discussed in earlier chapters:

- Stable (main-sequence) stars such as our Sun maintain equilibrium by producing energy through nuclear fusion in their cores. The ability to generate energy by fusion defines a star.
- Each second in the Sun, approximately 600 million tons of hydrogen undergo fusion into helium, with about 4 million tons turning into energy in the process. This rate of hydrogen use means that eventually the Sun (and all other stars) will run out of central fuel.
- Stars come with many different masses, ranging from 1/12 solar masses (M_{Sun}) to roughly 100–200 M_{Sun}. There are far more low-mass than high-mass stars.
- The most massive main-sequence stars (spectral type O) are also the most luminous and have the highest surface temperature. The lowest-mass stars on the main sequence (spectral type M or L) are the least luminous and the coolest.
- A galaxy of stars such as the Milky Way contains enormous amounts of gas and dust—enough to make billions of stars like the Sun.

If we want to find stars still in the process of formation, we must look in places that have plenty of the raw material from which stars are assembled. Since stars are made of gas, we focus our attention (and our telescopes) on the dense and cold clouds of gas and dust that dot the Milky Way (see Figure 21.1 and Figure 21.2).



(a) (b)

Figure 21.2 : Pillars of Dust and Dense Globules in M16. (a) This Hubble Space Telescope image of the central regions of M16 (also known as the Eagle Nebula) shows huge columns of cool gas, (including molecular hydrogen, H2) and dust. These columns are of higher density than the surrounding regions and have resisted evaporation by the ultraviolet radiation from a cluster of hot stars just beyond the upper-right corner of this image. The tallest pillar is about 1 light-year long, and the M16 region is about 7000 light-years away from us. (b) This close-up view of one of the pillars shows some very dense globules, many of which harbor embryonic stars. Astronomers coined the term *evaporating gas globules* (EGGs) for these structures, in part so that they could say we found EGGs inside the Eagle Nebula. It is possible that because these EGGs are exposed to the relentless action of the radiation from nearby hot stars, some may not yet have collected enough material to form a star. (credit a : modification of work by NASA, ESA, and the Hubble Heritage Team (STScI/AURA); credit b: modification of work by NASA, ESA, STScI, J. Hester and P. Scowen (Arizona State University))

Molecular Clouds: Stellar Nurseries

As we saw in Between the Stars: Gas and Dust in Space, the most massive reservoirs of interstellar matter—and some of the most massive objects in the Milky Way Galaxy—are the giant molecular clouds. These clouds have cold interiors with characteristic temperatures of only 10–20 K; most of their gas atoms are bound into molecules. These clouds turn out to be the birthplaces of most stars in our Galaxy.

The masses of molecular clouds range from a thousand times the mass of the Sun to about 3 million solar masses. Molecular clouds have a complex filamentary structure, similar to cirrus clouds in Earth's atmosphere, but much less dense. The molecular cloud filaments can be up to 1000 light-years long. Within the clouds are cold, dense regions with typical masses of 50 to 500 times the mass of the Sun; we give these regions the highly technical name *clumps*. Within these clumps, there are even denser, smaller regions called cores. The cores are the embryos of stars. The conditions in these cores—low temperature and high density—are just what is required to make stars. Remember that the essence of the life story of any star is the ongoing competition between two forces: *gravity* and *pressure*. The force of gravity, pulling inward, tries to make a star collapse. Internal pressure produced by the motions of the gas atoms, pushing outward, tries to force the star to expand. When a star is first forming, low temperature (and hence, low pressure) and high density (hence, greater gravitational attraction) both work to give gravity the advantage. In order to form a star—that is, a dense, hot ball of matter capable of starting nuclear reactions deep within—we need a typical core of interstellar atoms and molecules to shrink in radius and increase in density by a factor of nearly 10^{20}. It is the force of gravity that produces this drastic collapse.

The Orion Molecular Cloud

Let's discuss what happens in regions of star formation by considering a nearby site where stars are forming right now. One of the best-studied stellar nurseries is in the constellation of Orion, The Hunter, about 1500 light-years away (Figure 21.3). The pattern of the hunter is easy to recognize by the conspicuous "belt" of three stars that mark his waist. The Orion molecular cloud is much larger than the star pattern and is truly an impressive structure. In its long dimension, it stretches over a distance of about 100 light-years. The total quantity of molecular gas is about 200,000 times the mass of the Sun. Most of the cloud does not glow with visible light but betrays its presence by the radiation that the dusty gas gives off at infrared and radio wavelengths.

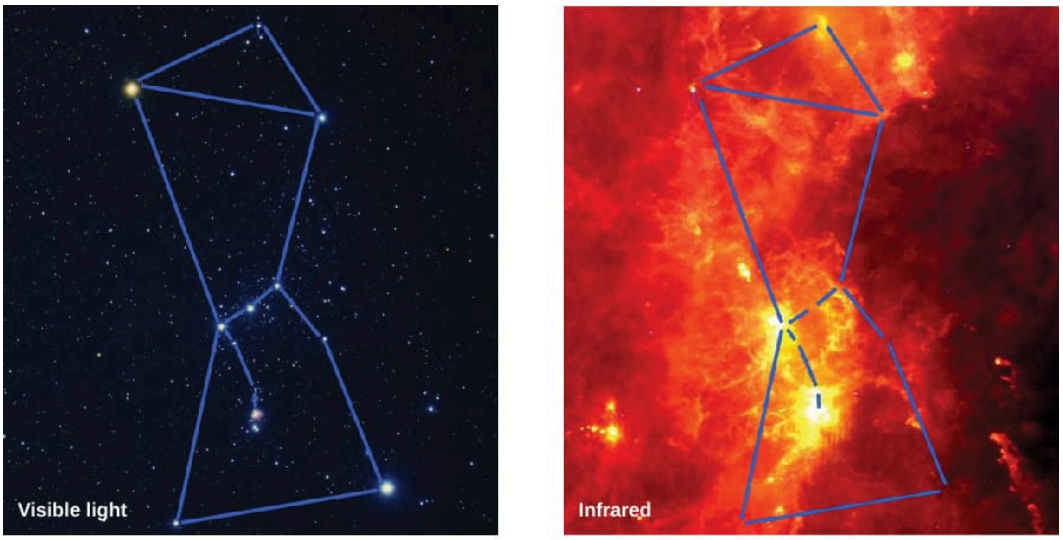


Figure 21.3 : Orion in Visible and Infrared. (a) The Orion star group was named after the legendary hunter in Greek mythology. Three stars close together in a link mark Orion's belt. The ancients imagined a sword hanging from the belt; the object at the end of the blue line in this sword is the Orion Nebula. (b) This wide-angle, infrared view of the same area was taken with the Infrared Astronomical Satellite. Heated dust clouds dominate in this false-color image, and many of the stars that stood out on part (a) are now invisible. An exception is the cool, red-giant star Betelgeuse, which can be seen as a yellowish point at the left vertex of the blue triangle (at Orion's left armpit). The large, yellow ring to the right of Betelgeuse is the remnant of an exploded star. The infrared image lets us see how large and full of cooler material the Orion molecular cloud really is. On the visible-light image at left, you see only two colorful regions of interstellar matter—the two, bright yellow splotches at the left end of and below Orion's belt. The lower one is the Orion Nebula and the higher one is the region of the Horsehead Nebula. (credit: modification of work by NASA, visible light: Akira Fujii; infrared: Infrared Astronomical Satellite)

The stars in Orion's belt are typically about 5 million years old, whereas the stars near the middle of the "sword" hanging from Orion's belt are only 300,000 to 1 million years old. The region about halfway down the sword where star formation is still taking place is called the Orion Nebula. About 2200 young stars are found in this region, which is only slightly larger than a dozen light-years in diameter. The Orion Nebula also contains a tight cluster of stars called the Trapezium (Figure 21.5). The brightest Trapezium stars can be seen easily with a small telescope.

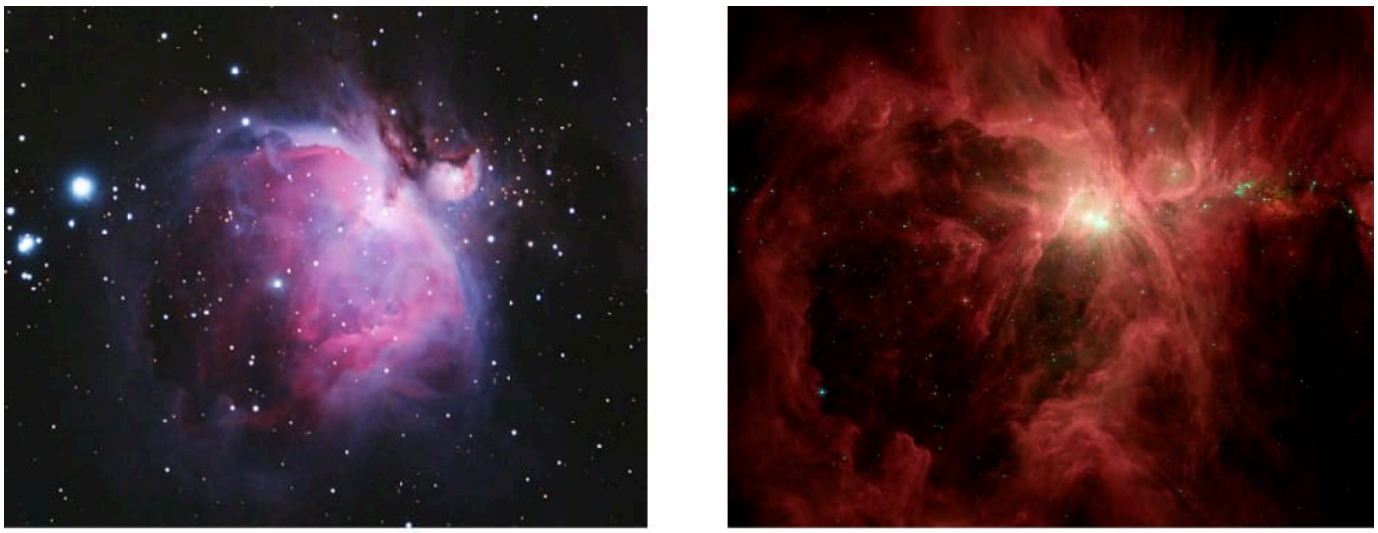


Figure 21.4 : Orion Nebula. (a) The Orion Nebula is shown in visible light. (b) With near-infrared radiation, we can see more detail within the dusty nebula since infrared can penetrate dust more easily than can visible light. (credit a: modification of work by Filip Lolić; credit b: modification of work by NASA/JPL-Caltech/T. Megeath (University of Toledo, Ohio))

Compare this with our own solar neighborhood, where the typical spacing between stars is about 3 light-years. Only a small number of stars in the Orion cluster can be seen with visible light, but infrared images—which penetrate the dust better—detect

the more than 2000 stars that are part of the group (Figure 21.5).



(a) (b)

Figure 21.5 : Central Region of the Orion Nebula. The Orion Nebula harbors some of the youngest stars in the solar neighborhood. At the heart of the nebula is the Trapezium cluster, which includes four very bright stars that provide much of the energy that causes the nebula to glow so brightly. In these images, we see a section of the nebula in (a) visible light and (b) infrared. The four bright stars in the center of the visible-light image are the Trapezium stars. Notice that most of the stars seen in the infrared are completely hidden by dust in the visible-light image. (credit a: modification of work by NASA, C.R. O'Dell and S.K. Wong (Rice University); credit b: modification of work by NASA; K.L. Luhman (Harvard-Smithsonian Center for Astrophysics); and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thompson (Steward Observatory, University of Arizona))

Studies of Orion and other star-forming regions show that star formation is not a very efficient process. In the region of the Orion Nebula, about 1% of the material in the cloud has been turned into stars. That is why we still see a substantial amount of gas and dust near the Trapezium stars. The leftover material is eventually heated, either by the radiation and winds from the hot stars that form or by explosions of the most massive stars. (We will see in later chapters that the most massive stars go through their lives very quickly and end by exploding.)

Link to Learning

Take a journey through the Orion Nebula to view a nice narrated video tour of this region.

Whether gently or explosively, the material in the neighborhood of the new stars is blown away into interstellar space. Older groups or clusters of stars can now be easily observed in visible light because they are no longer shrouded in dust and gas (Figure 21.6).



Figure 21.6 : Westerlund 2. This young cluster of stars known as Westerlund 2 formed within the Carina star-forming region about 2 million years ago. Stellar winds and pressure produced by the radiation from the hot stars within the cluster are blowing and sculpting the surrounding gas and dust. The nebula still contains many globules of dust. Stars are continuing to form within the denser globules and pillars of the nebula. This Hubble Space Telescope image includes near-infrared exposures of the star cluster and visible-light observations of the surrounding nebula. Colors in the nebula are dominated by the red glow of hydrogen gas, and blue-green emissions from glowing oxygen. (credit: NASA, ESA, the Hubble Heritage Team (STScI/AURA), A. Nota (ESA/STScI), and the Westerlund 2 Science Team)

Although we do not know what initially caused stars to begin forming in Orion, there is good evidence that the first generation of stars triggered the formation of additional stars, which in turn led to the formation of still more stars (Figure 21.7).

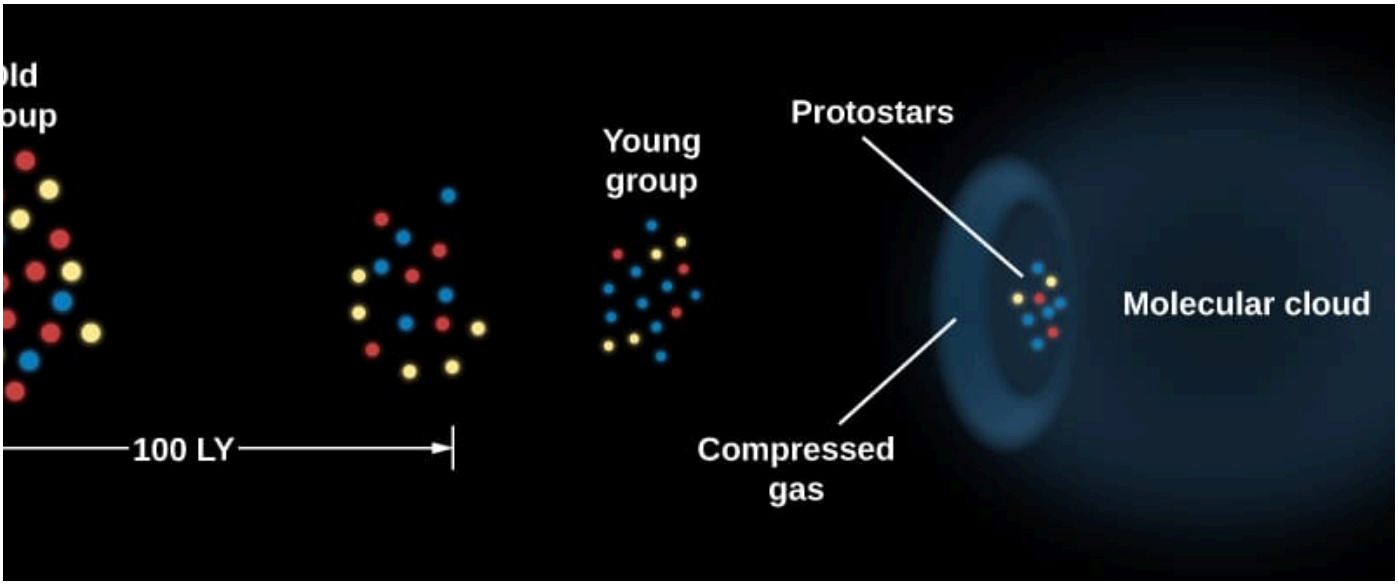


Figure 21.7 : Propagating Star Formation. Star formation can move progressively through a molecular cloud. The oldest group of stars lies to the left of the diagram and has expanded because of the motions of individual stars. Eventually, the stars in the group will disperse and no longer be recognizable as a cluster. The youngest group of stars lies to the right, next to the molecular cloud. This group of stars is only 1 to 2 million years old. The pressure of the hot, ionized gas surrounding these stars compresses the material in the nearby edge of the molecular cloud and initiates the gravitational collapse that will lead to the formation of more stars.

The basic idea of triggered star formation is this: when a massive star is formed, it emits a large amount of ultraviolet radiation and ejects high-speed gas in the form of a stellar wind. This injection of energy heats the gas around the stars and causes it to expand. When massive stars exhaust their supply of fuel, they explode, and the energy of the explosion also heats the gas. The hot gases pile into the surrounding cold molecular cloud, compressing the material in it and increasing its density. If this increase in density is large enough, gravity will overcome pressure, and stars will begin to form in the compressed gas. Such a

chain reaction—where the brightest and hottest stars of one area become the cause of star formation "next door"—seems to have occurred not only in Orion but also in many other molecular clouds.

There are many molecular clouds that form only (or mainly) low-mass stars. Because low-mass stars do not have strong winds and do not die by exploding, triggered star formation cannot occur in these clouds. There are also stars that form in relative isolation in small cores. Therefore, not all star formation is originally triggered by the death of massive stars. However, there are likely to be other possible triggers, such as spiral density waves and other processes we do not yet understand.

The Birth of a Star

Although regions such as Orion give us clues about how star formation begins, the subsequent stages are still shrouded in mystery (and a lot of dust). There is an enormous difference between the density of a molecular cloud core and the density of the youngest stars that can be detected. Direct observations of this collapse to higher density are nearly impossible for two reasons. First, the dust-shrouded interiors of molecular clouds where stellar births take place cannot be observed with visible light. Second, the timescale for the initial collapse—thousands of years—is very short, astronomically speaking. Since each star spends such a tiny fraction of its life in this stage, relatively few stars are going through the collapse process at any given time. Nevertheless, through a combination of theoretical calculations and the limited observations available, astronomers have pieced together a picture of what the earliest stages of stellar evolution are likely to be.

The first step in the process of creating stars is the formation of dense cores within a clump of gas and dust (Figure 21.8(a)). It is generally thought that all the material for the star comes from the core, the larger structure surrounding the forming star. Eventually, the gravitational force of the infalling gas becomes strong enough to overwhelm the pressure exerted by the cold material that forms the dense cores. The material then undergoes a rapid collapse, and the density of the core increases greatly as a result. During the time a dense core is contracting to become a true star, but before the fusion of protons to produce helium begins, we call the object a protostar.

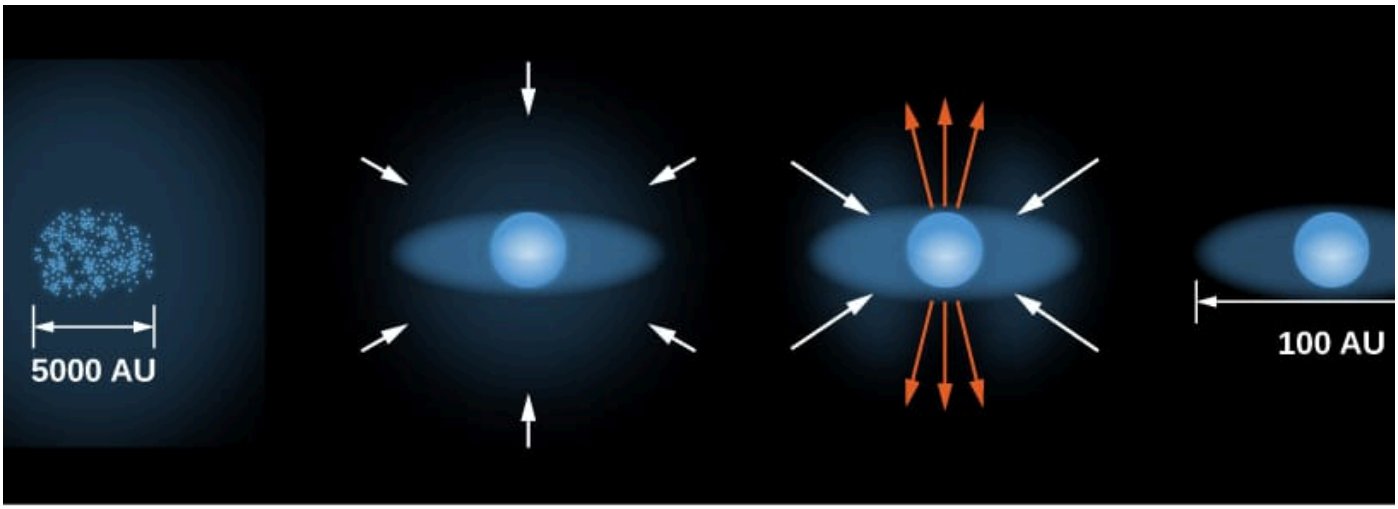


Figure 21.8 : Formation of a Star. (a) Dense cores form within a molecular cloud. (b) A protostar with a surrounding disk of material forms at the center of a dense core, accumulating additional material from the molecular cloud through gravitational attraction. (c) A stellar wind breaks out but is confined by the disk to flow out along the two poles of the star. (d) Eventually, this wind sweeps away the cloud material and halts the accumulation of additional material, and a newly formed star, surrounded by a disk, becomes observable. These sketches are not drawn to the same scale. The diameter of a typical envelope that is supplying gas to the newly forming star is about 5000 AU. The typical diameter of the disk is about 100 AU or slightly larger than the diameter of the orbit of Pluto.

The natural turbulence inside a clump tends to give any portion of it some initial spinning motion (even if it is very slow). As a result, each collapsing core is expected to spin. According to the law of conservation of angular momentum (discussed in the chapter on Orbits and Gravity), a rotating body spins more rapidly as it decreases in size. In other words, if the object can turn its material around a smaller circle, it can move that material more quickly—like a figure skater spinning more rapidly as she brings her arms in tight to her body. This is exactly what happens when a core contracts to form a protostar: as it shrinks, its rate of spin increases.

But all directions on a spinning sphere are not created equal. As the protostar rotates, it is much easier for material to fall right onto the poles (which spin most slowly) than onto the equator (where material moves around most rapidly). Therefore, gas and dust falling in toward the protostar's equator are "held back" by the rotation and form a whirling extended disk around the equator (part b in Figure 21.8). You may have observed this same "equator effect" on the amusement park ride in which you stand with your back to a cylinder that is spun faster and faster. As you spin really fast, you are pushed against the wall so strongly that you cannot possibly fall toward the center of the cylinder. Gas can, however, fall onto the protostar easily from directions away from the star's equator.

The protostar and disk at this stage are embedded in an envelope of dust and gas from which material is still falling onto the protostar. This dusty envelope blocks visible light, but infrared radiation can get through. As a result, in this phase of its evolution, the protostar itself is emitting infrared radiation and so is observable only in the infrared region of the spectrum. Once almost all of the available material has been accreted and the central protostar has reached nearly its final mass, it is given a special name: it is called a *T Tauri star*, named after one of the best studied and brightest members of this class of stars, which was discovered in the constellation of Taurus. (Astronomers have a tendency to name types of stars after the first example they discover or come to understand. It's not an elegant system, but it works.) Only stars with masses less than or similar to the mass of the Sun become T Tauri stars. Massive stars do not go through this stage, although they do appear to follow the formation scenario illustrated in Figure 21.8.

Winds and Jets

Recent observations suggest that T Tauri stars may actually be stars in a middle stage between protostars and hydrogen-fusing stars such as the Sun. High-resolution infrared images have revealed jets of material as well as *stellar winds* coming from some T Tauri stars, proof of interaction with their environment. A stellar wind consists mainly of protons (hydrogen nuclei) and electrons streaming away from the star at speeds of a few hundred kilometers per second (several hundred thousand miles per hour). When the wind first starts up, the disk of material around the star's equator blocks the wind in its direction. Where the wind particles *can* escape most effectively is in the direction of the star's poles.

Astronomers have actually seen evidence of these beams of particles shooting out in opposite directions from the polar regions of newly formed stars. In many cases, these beams point back to the location of a protostar that is still so completely shrouded in dust that we cannot yet see it (Figure 21.9).

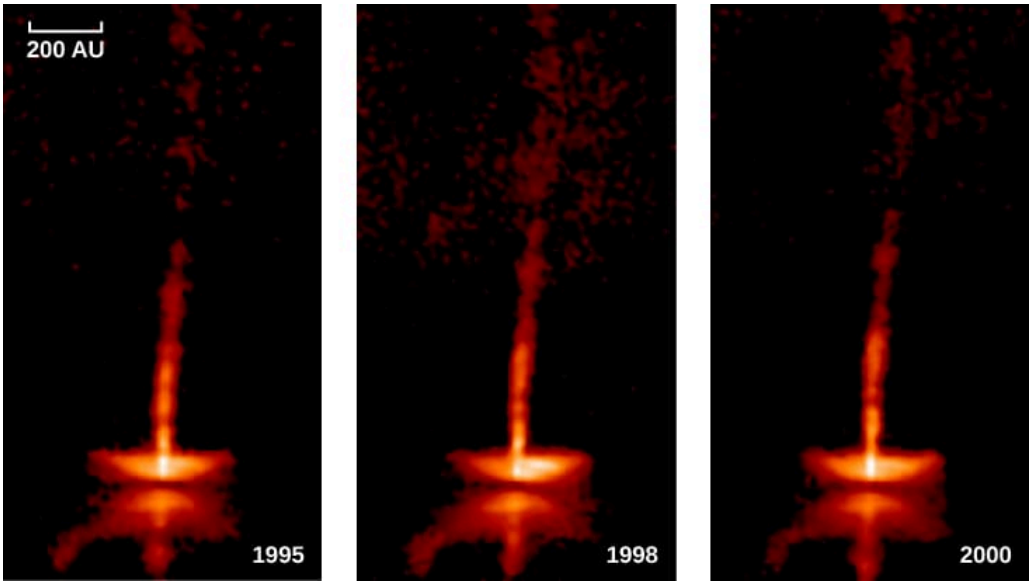


Figure 21.9 : Gas Jets Flowing away from a Protostar. Here we see the neighborhood of a protostar, known to us as HH 34 because it is a Herbig-Haro object. The star is about 450 light-years away and only about 1 million years old. Light from the star itself is blocked by a disk, which is larger than 60 billion kilometers in diameter and is seen almost edge-on. Jets are seen emerging perpendicular to the disk. The material in these jets is flowing outward at speeds up to 580,000 kilometers per hour. The series of three images shows changes during a period of 5 years. Every few months, a compact clump of gas is ejected, and its motion outward can be followed. The changes in the brightness of the disk may be due to motions of clouds within the disk that alternately block some of the light and then let it through. This image corresponds to the stage in the life of a protostar shown in part (c) of Figure 21.8. (credit: modification of work by Hubble Space Telescope, NASA, ESA)

On occasion, the jets of high-speed particles streaming away from the protostar collide with a somewhat-denser lump of gas nearby, excite its atoms, and cause them to emit light. These glowing regions, each of which is known as a Herbig-Haro (HH) object after the two astronomers who first identified them, allow us to trace the progress of the jet to a distance of a light-year or more from the star that produced it. Figure 21.10 shows two spectacular images of HH objects.

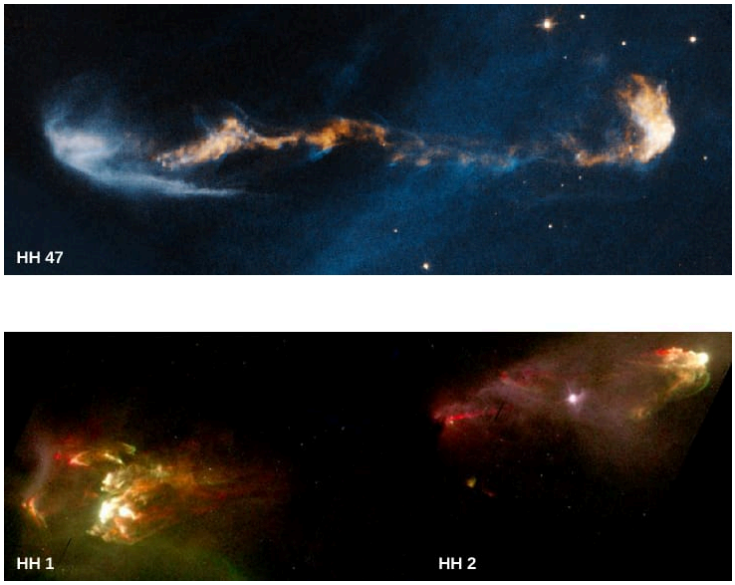


Figure 21.10 : Outflows from Protostars. These images were taken with the Hubble Space Telescope and show jets flowing outward from newly formed stars. In the HH47 image, a protostar 1500 light-years away (invisible inside a dust disk at the left edge of the image) produces a very complicated jet. The star may actually be wobbling, perhaps because it has a companion. Light from the star illuminates the white region at the left because light can emerge perpendicular to the disk (just as the jet does). At right, the jet is plowing into existing clumps of interstellar gas, producing a shock wave that resembles an arrowhead. The HH1/2 image shows a double-beam jet emanating from a protostar (hidden in a dust disk in the center) in the constellation of Orion. Tip to tip, these jets are more than 1 light-year long. The bright regions (first identified by Herbig and Haro) are places where the jet is a slamming into a clump of interstellar gas and causing it to glow. (credit “HH 47”: modification of work by NASA, ESA, and P. Hartigan (Rice University); credit “HH 1 and HH 2: modification of work by J. Hester, WFPC2 Team, NASA)

The wind from a forming star will ultimately sweep away the material that remains in the obscuring envelope of dust and gas, leaving behind the naked disk and protostar, which can then be seen with visible light. We should note that at this point, the protostar itself is still contracting slowly and has not yet reached the main-sequence stage on the H–R diagram (a concept introduced in the chapter The Stars: A Celestial Census). The disk can be detected directly when observed at infrared wavelengths or when it is seen silhouetted against a bright background (Figure 21.11).

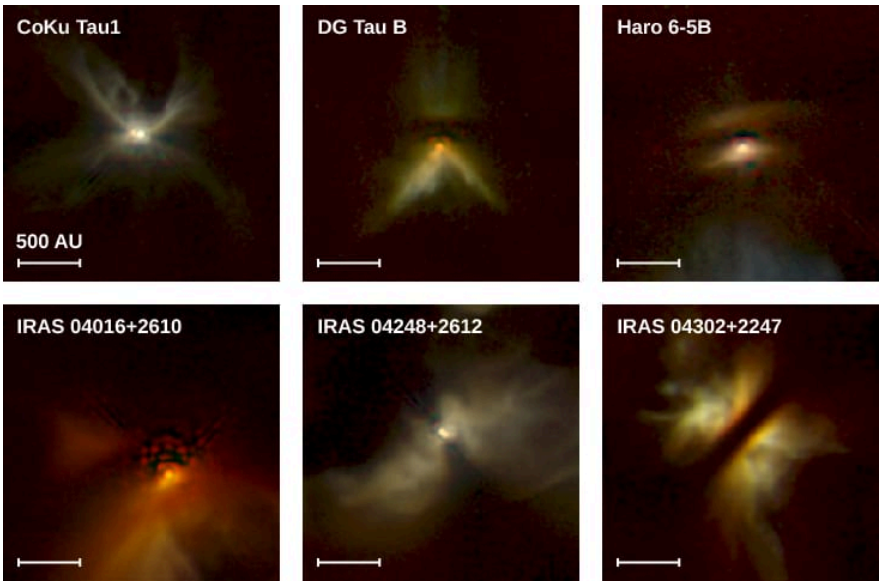


Figure 21.11 : Disks around Protostars. These Hubble Space Telescope infrared images show disks around young stars in the constellation of Taurus, in a region about 450 light-years away. In some cases, we can see the central star (or stars—some are binaries). In other cases, the dark, horizontal bands indicate regions where the dust disk is so thick that even infrared radiation from the star embedded within it cannot make its way through. The brightly glowing regions are starlight reflected from the upper and lower surfaces of the disk, which are less dense than the central, dark regions. (Credit: modification of work by D. Padgett (IPAC/Caltech), W. Brandner (IPAC), K. Stapelfeldt (JPL) and NASA)

This description of a protostar surrounded by a rotating disk of gas and dust sounds very much like what happened in our solar system when the Sun and planets formed. Indeed, one of the most important discoveries from the study of star formation in the last decade of the twentieth century was that disks are an inevitable byproduct of the process of creating stars. The next questions that astronomers set out to answer was: will the disks around protostars also form planets? And if so, how often? We will return to these questions later in this chapter.

To keep things simple, we have described the formation of single stars. Many stars, however, are members of binary or triple systems, where several stars are born together. In this case, the stars form in nearly the same way. Widely separated binaries may each have their own disk; close binaries may share a single disk.

This page titled 13.1: Star Formation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 21.2: Star Formation by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

13.2: The H-R and the Study of Stellar Evolution

Learning Objectives

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

- Determine the age of a protostar using an H–R diagram and the protostar's luminosity and temperature
- Explain the interplay between gravity and pressure, and how the contracting protostar changes its position in the H–R diagram as a result

One of the best ways to summarize all of these details about how a star or protostar changes with time is to use a Hertzsprung-Russell (H–R) diagram. Recall from The Stars: A Celestial Census that, when looking at an H–R diagram, the temperature (the horizontal axis) is plotted increasing toward the left. As a star goes through the stages of its life, its luminosity and temperature change. Thus, its position on the H–R diagram, in which luminosity is plotted against temperature, also changes. As a star ages, we must replot it in different places on the diagram. Therefore, astronomers often speak of a star *moving* on the H–R diagram, or of its evolution tracing out a path on the diagram. In this context, "tracing out a path" has nothing to do with the star's motion through space; this is just a shorthand way of saying that its temperature and luminosity change as it evolves.

Link to Learning

Watch an animation of the stars in the Omega Centauri cluster as they rearrange according to luminosity and temperature, forming a Hertzsprung-Russell (H–R) diagram.

To estimate just how much the luminosity and temperature of a star change as it ages, we must resort to calculations. Theorists compute a series of models for a star, with each successive model representing a later point in time. Stars may change for a variety of reasons. Protostars, for example, change in size because they are contracting, and their temperature and luminosity change as they do so. After nuclear fusion begins in the star's core (see Stars from Adolescence to Old Age), main-sequence stars change because they are using up their nuclear fuel.

Given a model that represents a star at one stage of its evolution, we can calculate what it will be like at a slightly later time. At each step, the model predicts the luminosity and size of the star, and from these values, we can figure out its surface temperature. A series of points on an H–R diagram, calculated in this way, allows us to follow the life changes of a star and hence is called its *evolutionary track*.

Evolutionary Tracks

Let's now use these ideas to follow the evolution of protostars that are on their way to becoming main-sequence stars. The evolutionary tracks of newly forming stars with a range of stellar masses are shown in Figure 21.12. These young stellar objects are not yet producing energy by nuclear reactions, but they derive energy from gravitational contraction—through the sort of process proposed for the Sun by Helmhotz and Kelvin in this last century (see the chapter on The Sun: A Nuclear Powerhouse).

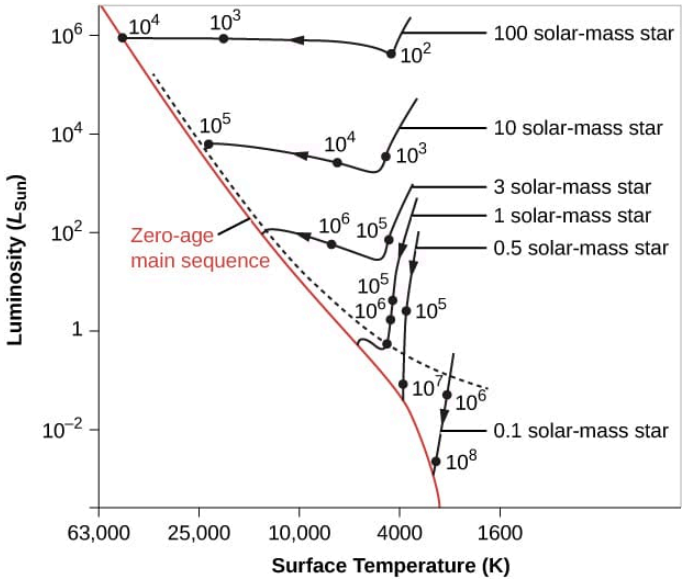


Figure 21.12 : Evolutionary Tracks for Contracting Protostars. Tracks are plotted on the H–R diagram to show how stars of different masses change during the early parts of their lives. The number next to each dark point on a track is the rough number of years it takes an embryo star to reach that stage (the numbers are the result of computer models and are therefore not well known). Note that the surface temperature (K) on the horizontal axis increases toward the left. You can see that the more mass a star has, the shorter time it takes to go through each stage. Stars above the dashed line are typically still surrounded by infalling material and are hidden by it.

Initially, a protostar remains fairly cool with a very large radius and a very low density. It is transparent to infrared radiation, and the heat generated by gravitational contraction can be radiated away freely into space. Because heat builds up slowly inside the protostar, the gas pressure remains low, and the outer layers fall almost unhindered toward the center. Thus, the protostar undergoes very rapid collapse, a stage that corresponds to the roughly vertical lines at the right of Figure 21.12. As the star shrinks, its surface area gets smaller, and so its total luminosity decreases. The rapid contraction stops only when the protostar becomes dense and opaque enough to trap the heat released by gravitational contraction.

When the star begins to retain its heat, the contraction becomes much slower, and changes inside the contracting star keep the luminosity of stars like our Sun roughly constant. The surface temperatures start to build up, and the star “moves” to the left in the H–R diagram. Stars first become visible only after the stellar wind described earlier clears away the surrounding dust and gas. This can happen during the rapid-contraction phase for low-mass stars, but high-mass stars remain shrouded in dust until they end their early phase of gravitational contraction (see the dashed line in Figure 21.12).

To help you keep track of the various stages that stars go through in their lives, it can be useful to compare the development of a star to that of a human being. (Clearly, you will not find an exact correspondence, but thinking through the stages in human terms may help you remember some of the ideas we are trying to emphasize.) Protostars might be compared to human embryos—as yet unable to sustain themselves but drawing resources from their environment as they grow. Just as the birth of a child is the moment it is called upon to produce its own energy (through eating and breathing), so astronomers say that a star is born when it is able to sustain itself through nuclear reactions (by making its own energy.)

When the star’s central temperature becomes high enough (about 12 million K) to fuse hydrogen into helium, we say that the star has reached the main sequence (a concept introduced in The Stars: A Celestial Census). It is now a full-fledged star, more or less in equilibrium, and its rate of change slows dramatically. Only the gradual depletion of hydrogen as it is transformed into helium in the core slowly changes the star’s properties.

The mass of a star determines exactly where it falls on the main sequence. As Figure 21.12 shows, massive stars on the main sequence have high temperatures and high luminosities. Low-mass stars have low temperatures and low luminosities.

Objects of extremely low mass never achieve high-enough central temperatures to ignite nuclear reactions. The lower end of the main sequence stops where stars have a mass just barely great enough to sustain nuclear reactions at a sufficient rate to stop gravitational contraction. This critical mass is calculated to be about 0.075 times the mass of the Sun. As we discussed in the chapter on Analyzing Starlight, objects below this critical mass are called either brown dwarfs or planets. At the other extreme, the upper end of the main sequence terminates at the point where the energy radiated by the newly forming massive star becomes so great that it halts the accretion of additional matter. The upper limit of stellar mass is between 100 and 200 solar masses.

Evolutionary Timescales

How long it takes a star to form depends on its mass. The numbers that label the points on each track in Figure 21.12 are the times, in years, required for the embryo stars to reach the stages we have been discussing. Stars of masses much higher than the Sun's reach the main sequence in a few thousand to a million years. The Sun required millions of years before it was born. Tens of millions of years are required for stars of lower mass to evolve to the lower main sequence. (We will see that this turns out to be a general principle: massive stars go through *all* stages of evolution faster than low-mass stars do.)

We will take up the subsequent stages in the life of a star in Stars from Adolescence to Old Age, examining what happens after stars arrive in the main sequence and begin a "prolonged adolescence" and "adulthood" of fusing hydrogen to form helium. But now we want to examine the connection between the formation of stars and planets.

This page titled 13.2: The H-R and the Study of Stellar Evolution is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 21.3: The H-R and the Study of Stellar Evolution by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

13.3: Evolution from the Main Sequence to Red Giants

Learning Objectives

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

- Explain the zero-age main sequence
- Describe what happens to main-sequence stars of various masses as they exhaust their hydrogen supply

One of the best ways to get a "snapshot" of a group of stars is by plotting their properties on an H–R diagram. We have already used the H–R diagram to follow the evolution of protostars up to the time they reach the main sequence. Now we'll see what happens next.

Once a star has reached the main-sequence stage of its life, it derives its energy almost entirely from the conversion of hydrogen to helium via the process of nuclear fusion in its core (see The Sun: A Nuclear Powerhouse). Since hydrogen is the most abundant element in stars, this process can maintain the star's equilibrium for a long time. Thus, all stars remain on the main sequence for most of their lives. Some astronomers like to call the main-sequence phase the star's "prolonged adolescence" or "adulthood" (continuing our analogy to the stages in a human life).

The left-hand edge of the main-sequence band in the H–R diagram is called the zero-age main sequence (see Figure 18.15). We use the term *zero-age* to mark the time when a star stops contracting, settles onto the main sequence, and begins to fuse hydrogen in its core. The zero-age main sequence is a continuous line in the H–R diagram that shows where stars of different masses but similar chemical composition can be found when they begin to fuse hydrogen.

Since only 0.7% of the hydrogen used in fusion reactions is converted into energy, fusion does not change the *total* mass of the star appreciably during this long period. It does, however, change the chemical composition in its central regions where nuclear reactions occur: hydrogen is gradually depleted, and helium accumulates. This change of composition changes the luminosity, temperature, size, and interior structure of the star. When a star's luminosity and temperature begin to change, the point that represents the star on the H–R diagram moves away from the zero-age main sequence.

Calculations show that the temperature and density in the inner region slowly increase as helium accumulates in the center of a star. As the temperature gets hotter, each proton acquires more energy of motion on average; this means it is more likely to interact with other protons, and as a result, the rate of fusion also increases. For the proton-proton cycle described in The Sun: A Nuclear Powerhouse, the rate of fusion goes up roughly as the temperature to the fourth power.

If the rate of fusion goes up, the rate at which energy is being generated also increases, and the luminosity of the star gradually rises. Initially, however, these changes are small, and stars remain within the main-sequence band on the H–R diagram for most of their lifetimes.

Example 22.1: Star Temperature and Rate of Fusion

If a star's temperature were to double, by what factor would its rate of fusion increase?

Solution

Since the rate of fusion (like temperature) goes up to the fourth power, it would increase by a factor of 2^4, or 16 times.

Exercise 13.3.1

If the rate of fusion of a star increased 256 times, by what factor would the temperature increase?

Answer available at the end of this section.

Lifetimes on the Main Sequence

How many years a star remains in the main-sequence band depends on its mass. You might think that a more massive star, having more fuel, would last longer, but it's not that simple. The lifetime of a star in a particular stage of evolution depends on how much nuclear fuel it has and on *how quickly* it uses up that fuel. (In the same way, how long people can keep spending money depends not only on how much money they have but also on how quickly they spend it. This is why many lottery winners who go on spending sprees quickly wind up poor again.) In the case of stars, more massive ones use up their fuel much more quickly than stars of low mass.

The reason massive stars are such spendthrifts is that, as we saw above, the rate of fusion depends *very* strongly on the star's core temperature. And what determines how hot a star's central regions get? It is the *mass* of the star—the weight of the overlying layers determines how high the pressure in the core must be: higher mass requires higher pressure to balance it. Higher pressure, in turn, is produced by higher temperature. The higher the temperature in the central regions, the faster the star races through its storehouse of central hydrogen. Although massive stars have more fuel, they burn it so prodigiously that their lifetimes are much shorter than those of their low-mass counterparts. You can also understand now why the most massive main-sequence stars are also the most luminous. Like new rock stars with their first platinum album, they spend their resources at an astounding rate.

The main-sequence lifetimes of stars of different masses are listed in Table 22.1. This table shows that the most massive stars spend only a few million years on the main sequence. A star of 1 solar mass remains there for roughly 10 billion years, while a star of about 0.4 solar mass has a main-sequence lifetime of some 200 billion years, which is longer than the current age of the universe. (Bear in mind, however, that every star spends *most* of its total lifetime on the main sequence. Stars devote an average of 90% of their lives to peacefully fusing hydrogen into helium.)

Table 22.1: Lifetimes of Main-Sequence Stars

Spectral Type	Surface Temperature (K)	Mass (Mass of Sun = 1)	Lifetime on Main Sequence (years)
O5	54,000	40	1 million
B0	29,200	16	10 million
A0	9600	3.3	500 million
F0	7350	1.7	2.7 billion
G0	6050	1.1	9 billion
K0	5240	0.8	14 billion
M0	3750	0.4	200 billion

These results are not merely of academic interest. Human beings developed on a planet around a G-type star. This means that the Sun's stable main-sequence lifetime is so long that it afforded life on Earth plenty of time to evolve. When searching for intelligent life like our own on planets around other stars, it would be a pretty big waste of time to search around O- or B-type stars. These stars remain stable for such a short time that the development of creatures complicated enough to take astronomy courses is very unlikely.

From Main-Sequence Star to Red Giant

Eventually, all the hydrogen in a star's core, where it is hot enough for fusion reactions, is used up. The core then contains only helium, "contaminated" by whatever small percentage of heavier elements the star had to begin with. The helium in the core can be thought of as the accumulated "ash" from the nuclear "burning" of hydrogen during the main-sequence stage.

Energy can no longer be generated by hydrogen fusion in the stellar core because the hydrogen is all gone and, as we will see, the fusion of helium requires much higher temperatures. Since the central temperature is not yet high enough to fuse helium, there is no nuclear energy source to supply heat to the central region of the star. The long period of stability now ends, gravity again takes over, and the core begins to contract. Once more, the star's energy is partially supplied by gravitational energy, in the way described by Kelvin and Helmholtz (see Sources of Sunshine: Thermal and Gravitational Energy). As the star's core shrinks, the energy of the inward-falling material is converted to heat.

The heat generated in this way, like all heat, flows outward to where it is a bit cooler. In the process, the heat raises the temperature of a layer of hydrogen that spent the whole long main-sequence time just outside the core. Like an understudy waiting in the wings of a hit Broadway show for a chance at fame and glory, this hydrogen was almost (but not quite) hot enough to undergo fusion and take part in the main action that sustains the star. Now, the additional heat produced by the shrinking core puts this hydrogen "over the limit," and a shell of hydrogen nuclei just outside the core becomes hot enough for hydrogen fusion to begin.

New energy produced by fusion of this hydrogen now pours outward from this shell and begins to heat up layers of the star farther out, causing them to expand. Meanwhile, the helium core continues to contract, producing more heat right around it. This leads to more fusion in the shell of fresh hydrogen outside the core (Figure 22.2). The additional fusion produces still more energy, which also flows out into the upper layer of the star.

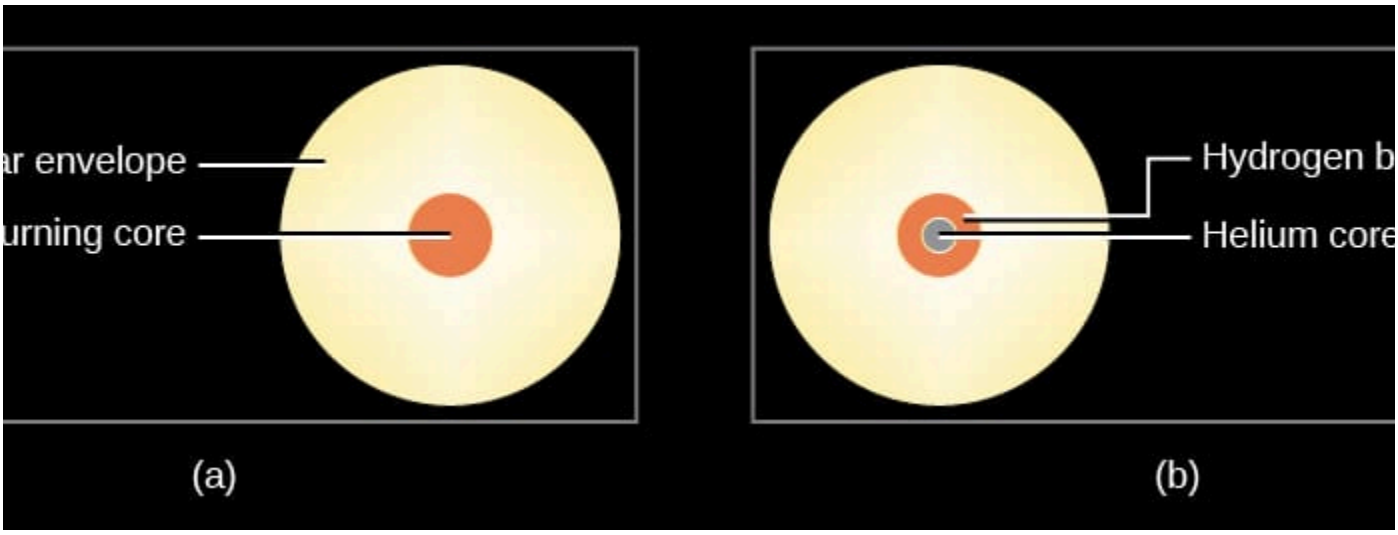


Figure 22.2 : Star Layers during and after the Main Sequence. (a) During the main sequence, a star has a core where fusion takes place and a much larger envelope that is too cold for fusion. (b) When the hydrogen in the core is exhausted (made of helium, not hydrogen), the core is compressed by gravity and heats up. The additional heat starts hydrogen fusion in a layer just outside the core. Note that these parts of the Sun are not drawn to scale.

Most stars actually generate more energy each second when they are fusing hydrogen in the shell surrounding the helium core than they did when hydrogen fusion was confined to the central part of the star; thus, they increase in luminosity. With all the new energy pouring outward, the outer layers of the star begin to expand, and the star eventually grows and grows until it reaches enormous proportions (Figure 22.3).

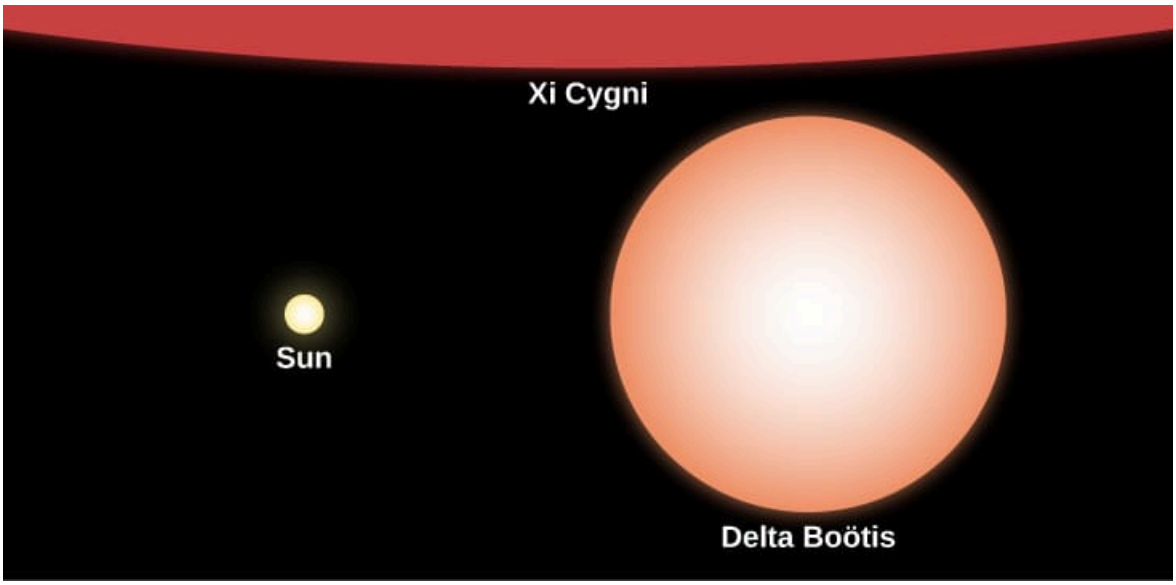


Figure 22.3 : Relative Sizes of Stars. This image compares the size of the Sun to that of Delta Boötis, a giant star, and Xi Cygni, a supergiant. Note that Xi Cygni is so large in comparison to the other two stars that only a small portion of it is visible at the top of the frame.

When you take the lid off a pot of boiling water, the steam can expand and it cools down. In the same way, the expansion of a star's outer layers causes the temperature at the surface to decrease. As it cools, the star's overall color becomes redder. (We saw in Radiation and Spectra that a red color corresponds to cooler temperature.)

So the star becomes simultaneously more luminous and cooler. On the H–R diagram, the star therefore leaves the main-sequence band and moves upward (brighter) and to the right (cooler surface temperature). Over time, massive stars become red supergiants, and lower-mass stars like the Sun become red giants. (We first discussed such giant stars in The Stars: A Celestial Census; here we see how such "swollen" stars originate.) You might also say that these stars have "split personalities": their cores are contracting while their outer layers are expanding. (Note that red giant stars do not actually look deep red; their colors are more like orange or orange-red.)

Just how different are these red giants and supergiants from a main-sequence star? Table 22.2 compares the Sun with the red supergiant Betelgeuse, which is visible above Orion's belt as the bright red star that marks the hunter's armpit. Relative to the Sun, this supergiant has a much larger radius, a much lower average density, a cooler surface, and a much hotter core.

Table 22.2: Comparing a Supergiant with the Sun

Property	Sun	Betelgeuse
Mass $(2 \times 10^{33}\text{g})$	1	16
Radius (km)	700,000	500,000,000
Surface temperature (K)	5,800	3,600
Core temperature (K)	15,000,000	160,000,000
Luminosity $(4 \times 10^{26}\text{W})$	1	46,000
Average density (g/cm^3)	1.4	1.3×10^{-7}
Age (millions of years)	4,500	10

Red giants can become so large that if we were to replace the Sun with one of them, its outer atmosphere would extend to the orbit of Mars or even beyond (Figure 22.4). This is the next stage in the life of a star as it moves (to continue our analogy to human lives) from its long period of "youth" and "adulthood" to "old age." By considering the relative ages of the Sun and Betelgeuse, we can also see that the idea that "bigger stars die faster" is indeed true here. Betelgeuse is a mere 10 million years old, which is relatively young compared with our Sun's 4.5 billion years, but it is already nearing its death throes as a red supergiant.

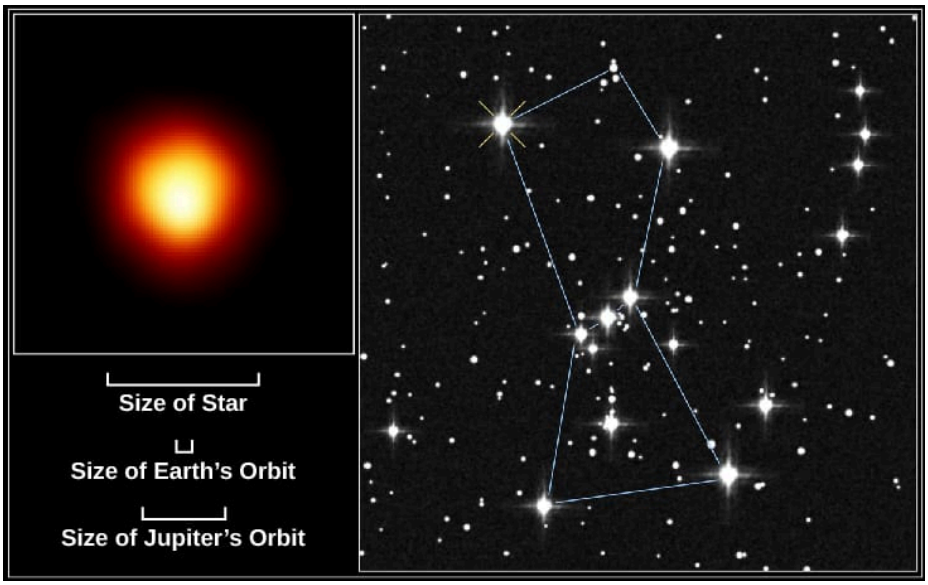


Figure 22.4 : Betelgeuse. Betelgeuse is in the constellation Orion, the hunter; in the right image, it is marked with a yellow "X" near the top left. In the left image, we see it in ultraviolet with the Hubble Space Telescope, in the first direct image ever made of the surface of another star. As shown by the scale at the bottom, Betelgeuse has an extended atmosphere so large that, if it were at the center of our solar system, it would stretch past the orbit of Jupiter. (credit: Modification of work by Andrea Dupree (Harvard-Smithsonian CfA), Ronald Gilliland (STScI), NASA and ESA)

Models for Evolution to the Giant Stage

As we discussed earlier, astronomers can construct computer models of stars with different masses and compositions to see how stars change throughout their lives. Figure 22.5, which is based on theoretical calculations by University of Illinois astronomer Icko Iben, shows an H–R diagram with several tracks of evolution from the main sequence to the giant stage. Tracks are shown for stars with different masses (from 0.5 to 15 times the mass of our Sun) and with chemical compositions similar to that of the Sun. The red line is the initial or zero-age main sequence. The numbers along the tracks indicate the time, in years, required for each star to reach those points in their evolution after leaving the main sequence. Once again, you can see that the more massive a star is, the more quickly it goes through each stage in its life.

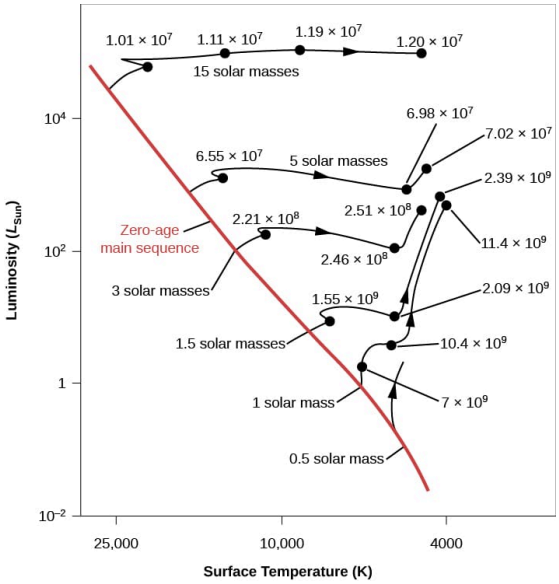


Figure 22.5 : Evolutionary Tracks of Stars of Different Masses. The solid black lines show the predicted evolution from the main sequence through the red giant or supergiant stage on the H–R diagram. Each track is labeled with the mass of the star it is describing. The numbers show how many years each star takes to become a giant. The red line is the zero-age main sequence. While theorists debate the exact number of years shown here, our main point should be clear. The more massive the star, the shorter time it takes for each stage in its life.

Note that the most massive star in this diagram has a mass similar to that of Betelgeuse, and so its evolutionary track shows approximately the history of Betelgeuse. The track for a 1-solar-mass star shows that the Sun is still in the main-sequence phase of evolution, since it is only about 4.5 billion years old. It will be billions of years before the Sun begins its own “climb” away from the main sequence—the expansion of its outer layers that will make it a red giant.

Link to Learning

Use the Star in a Box simulation to explore the evolution of stars of different masses. Select the star’s mass, and then hit play to see how its luminosity, temperature, and size change during its lifetime.

Answers

Exercise 13.3.1

The temperature would increase by a factor of $256^{0.25}$ (that is, the 4th root of 256), or 4 times.

This page titled 13.3: Evolution from the Main Sequence to Red Giants is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 22.2: Evolution from the Main Sequence to Red Giants by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

13.4: Further Evolution of Stars

Learning Objectives

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

- Explain what happens in a star's core when all of the hydrogen has been used up
- Define "planetary nebulae" and discuss their origin
- Discuss the creation of new chemical elements during the late stages of stellar evolution

The "life story" we have related so far applies to almost all stars: each starts as a contracting protostar, then lives most of its life as a stable main-sequence star, and eventually moves off the main sequence toward the red-giant region.

As we have seen, the pace at which each star goes through these stages depends on its mass, with more massive stars evolving more quickly. But after this point, the life stories of stars of different masses diverge, with a wider range of possible behavior according to their masses, their compositions, and the presence of any nearby companion stars.

Because we have written this book for students taking their first astronomy course, we will recount a simplified version of what happens to stars as they move toward the final stages in their lives. We will (perhaps to your heartfelt relief) not delve into all the possible ways aging stars can behave and the strange things that happen when a star is orbited by a second star in a binary system. Instead, we will focus only on the key stages in the evolution of single stars and show how the evolution of high-mass stars differs from that of low-mass stars (such as our Sun).

Helium Fusion

Let's begin by considering stars with composition like that of the Sun and whose *initial* masses are comparatively low—no more than about twice the mass of our Sun. (Such mass may not seem too low, but stars with masses less than this all behave in a fairly similar fashion. We will see what happens to more massive stars in the next section.) Because there are many more low-mass stars than high-mass stars in the Milky Way, the vast majority of stars—including our Sun—follow the scenario we are about to relate. By the way, we carefully used the term *initial masses* of stars because, as we will see, stars can lose quite a bit of mass in the process of aging and dying.

Remember that red giants start out with a helium core where no energy generation is taking place, surrounded by a shell where hydrogen is undergoing fusion. The core, having no source of energy to oppose the inward pull of gravity, is shrinking and growing hotter. As time goes on, the temperature in the core can rise to much hotter values than it had in its main-sequence days. Once it reaches a temperature of 100 million K (but not before such point), three helium atoms can begin to fuse to form a single carbon nucleus. This process is called the triple-alpha process, so named because physicists call the nucleus of the helium atom an alpha particle.

When the triple-alpha process begins in low-mass (about 0.8 to 2.0 solar masses) stars, calculations show that the entire core is ignited in a quick burst of fusion called a helium flash. (More massive stars also ignite helium but more gradually and not with a flash.) As soon as the temperature at the center of the star becomes high enough to start the triple-alpha process, the extra energy released is transmitted quickly through the entire helium core, producing very rapid heating. The heating speeds up the nuclear reactions, which provide more heating, and which accelerates the nuclear reactions even more. We have runaway generation of energy, which reignites the entire helium core in a flash.

You might wonder why the next major step in nuclear fusion in stars involves three helium nuclei and not just two. Although it is a lot easier to get two helium nuclei to collide, the product of this collision is not stable and falls apart very quickly. It takes three helium nuclei coming together simultaneously to make a stable nuclear structure. Given that each helium nucleus has two positive protons and that such protons repel one another, you can begin to see the problem. It takes a temperature of 100 million K to slam three helium nuclei (six protons) together and make them stick. But when that happens, the star produces a carbon nucleus.

Astronomy Basics: Stars in Your Little Finger

Stop reading for a moment and look at your little finger. It's full of carbon atoms because carbon is a fundamental chemical building block for life on Earth. Each of those carbon atoms was once inside a red giant star and was fused from helium nuclei in the triple-alpha process. All the carbon on Earth—in you, in the charcoal you use for barbecuing, and in the diamonds you might exchange with a loved one—was "cooked up" by previous generations of stars. How the carbon atoms (and other elements) made their way from inside some of those stars to become part of Earth is something we will discuss in the next chapter. For now, we want to emphasize that our description of stellar evolution is, in a very real sense, the story of our own cosmic "roots"—the history of how our own atoms originated among the stars. We are made of "star-stuff."

Becoming a Giant Again

After the helium flash, the star, having survived the "energy crisis" that followed the end of the main-sequence stage and the exhaustion of the hydrogen fuel at its center, finds its balance again. As the star readjusts to the release of energy from the triple-alpha process in its core, its internal structure changes once more: its surface temperature increases and its overall luminosity decreases. The point that represents the star on the H–R diagram thus moves to a new position to the left of and somewhat below its place as a red giant (Figure 22.16). The star then continues to fuse the helium in its core for a while, returning to the kind of equilibrium between pressure and gravity that characterized the main-sequence stage. During this time, a newly formed carbon nucleus at the center of the star can sometimes be joined by another helium nucleus to produce a nucleus of oxygen—another building block of life.

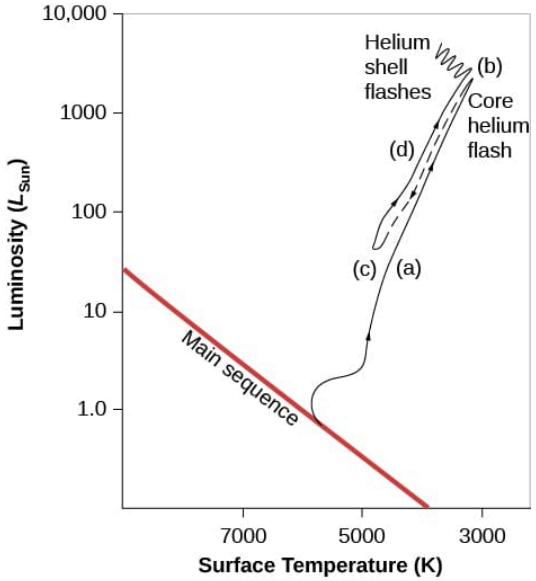


Figure 22.16 : Evolution of a Star Like the Sun on an H–R Diagram. Each stage in the star's life is labeled. (a) The star evolves from the main sequence to be a red giant, decreasing in surface temperature and increasing in luminosity. (b) A helium flash occurs, leading to a readjustment of the star's internal structure and to (c) a brief period of stability during which helium is fused to carbon and oxygen in the core (in the process the star becomes hotter and less luminous than it was as a red giant). (d) After the central helium is exhausted, the star becomes a giant again and moves to higher luminosity and lower temperature. By this time, however, the star has exhausted its inner resources and will soon begin to die. Where the evolutionary track becomes a dashed line, the changes are so rapid that they are difficult to model.

However, at a temperature of 100 million K, the inner core is converting its helium fuel to carbon (and a bit of oxygen) at a rapid rate. Thus, the new period of stability cannot last very long: it is far shorter than the main-sequence stage. Soon, all the helium hot enough for fusion will be used up, just like the hot hydrogen that was used up earlier in the star's evolution. Once again, the inner core will not be able to generate energy via fusion. Once more, gravity will take over, and the core will start to shrink again. We can think of stellar evolution as a story of a constant struggle against gravitational collapse. A star can avoid collapsing as long as it can tap energy sources, but once any particular fuel is used up, it starts to collapse again.

The star's situation is analogous to the end of the main-sequence stage (when the central hydrogen got used up), but the star now has a somewhat more complicated structure. Again, the star's core begins to collapse under its own weight. Heat released

by the shrinking of the carbon and oxygen core flows into a shell of helium just above the core. This helium, which had not been hot enough for fusion into carbon earlier, is heated just enough for fusion to begin and to generate a new flow of energy.

Farther out in the star, there is also a shell where fresh hydrogen has been heated enough to fuse helium. The star now has a multi-layered structure like an onion: a carbon-oxygen core, surrounded by a shell of helium fusion, a layer of helium, a shell of hydrogen fusion, and finally, the extended outer layers of the star (see Figure 22.17). As energy flows outward from the two fusion shells, once again the outer regions of the star begin to expand. Its brief period of stability is over; the star moves back to the red-giant domain on the H–R diagram for a short time (see Figure 22.16). But this is a brief and final burst of glory.

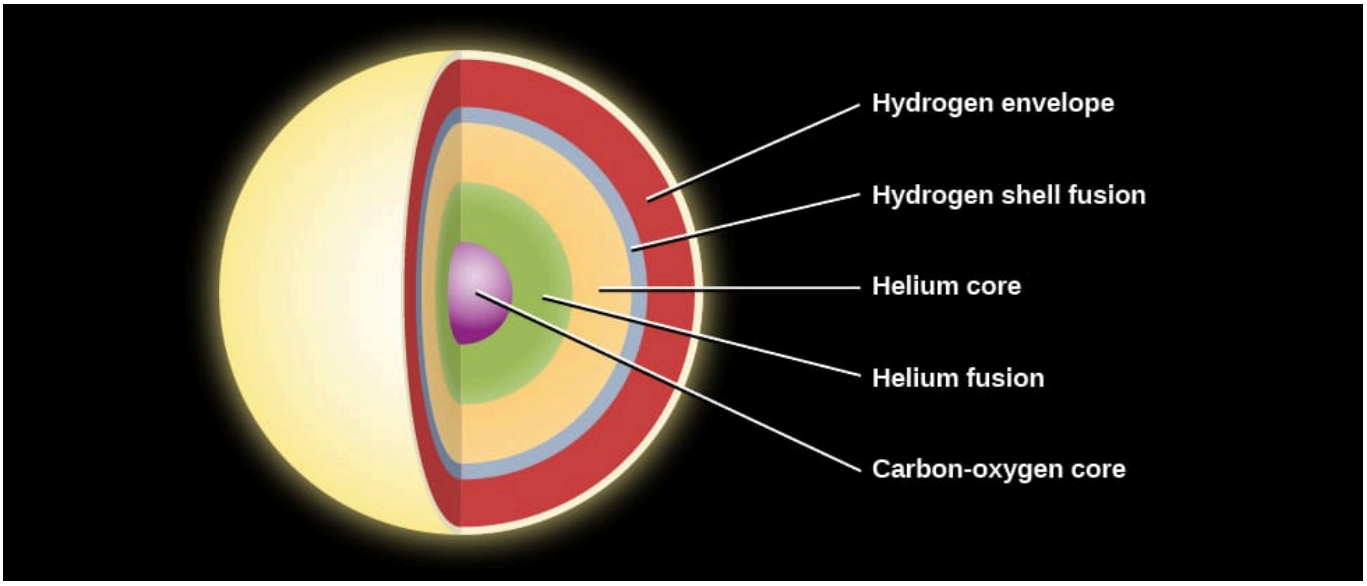


Figure 22.17 : Layers inside a Low-Mass Star before Death. Here we see the layers inside a star with an initial mass that is less than twice the mass of the Sun. These include, from the center outward, the carbon-oxygen core, a layer of helium hot enough to fuse, a layer of cooler helium, a layer of hydrogen hot enough to fuse, and then cooler hydrogen beyond.

Recall that the last time the star was in this predicament, helium fusion came to its rescue. The temperature at the star's center eventually became hot enough for the *product* of the previous step of fusion (helium) to become the *fuel* for the next step (helium fusing into carbon). But the step after the fusion of helium nuclei requires a temperature so hot that the kinds of lower-mass stars (less than 2 solar masses) we are discussing simply cannot compress their cores to reach it. No further types of fusion are possible for such a star.

In a star with a mass similar to that of the Sun, the formation of a carbon-oxygen core thus marks the end of the generation of nuclear energy at the center of the star. The star must now confront the fact that its death is near. We will discuss how stars like this end their lives in The Death of Stars, but in the meantime, Table 22.4 summarizes the stages discussed so far in the life of a star with the same mass as that of the Sun. One thing that gives us confidence in our calculations of stellar evolution is that when we make H–R diagrams of older clusters, we actually see stars in each of the stages that we have been discussing.

Table 22.4: The Evolution of a Star with the Sun's Mass

Stage	Time in This Stage (years)	Surface Temperature (K)	Luminosity (L_{Sun})	Diameter (Sun = 1)
Main sequence	11 billion	6000	1	1
Becomes red giant	1.3 billion	3100 at minimum	2300 at maximum	165
Helium fusion	100 million	4800	50	10
Giant again	20 million	3100	5200	180

Mass Loss from Red-Giant Stars and the Formation of Planetary Nebulae

When stars swell up to become red giants, they have very large radii and therefore a low escape velocity.[2] Radiation pressure,

stellar pulsations, and violent events like the helium flash can all drive atoms in the outer atmosphere away from the star, and cause it to lose a substantial fraction of its mass into space. Astronomers estimate that by the time a star like the Sun reaches the point of the helium flash, for example, it will have lost as much as 25% of its mass. And it can lose still more mass when it ascends the red-giant branch for the second time. As a result, aging stars are surrounded by one or more expanding shells of gas, each containing as much as 10–20% of the Sun's mass (or 0.1–0.2 M_{Sun}).

When nuclear energy generation in the carbon-oxygen core ceases, the star's core begins to shrink again and to heat up as it gets more and more compressed. (Remember that this compression will not be halted by another type of fusion in these low-mass stars.) The whole star follows along, shrinking and also becoming very hot—reaching surface temperatures as high as 100,000 K. Such hot stars are very strong sources of stellar winds and ultraviolet radiation, which sweep outward into the shells of material ejected when the star was a red giant. The winds and the ultraviolet radiation heat the shells, ionize them, and set them aglow (just as ultraviolet radiation from hot, young stars produces H II regions; see Between the Stars: Gas and Dust in Space).

The result is the creation of some of the most beautiful objects in the cosmos (see the gallery in Figure 22.18 and Figure 22.1). These objects were given an extremely misleading name when first found in the eighteenth century: planetary nebulae. The name is derived from the fact that a few planetary nebulae, when viewed through a small telescope, have a round shape bearing a superficial resemblance to planets. Actually, they have nothing to do with planets, but once names are put into regular use in astronomy, it is extremely difficult to change them. There are tens of thousands of planetary nebulae in our own Galaxy, although many are hidden from view because their light is absorbed by interstellar dust.

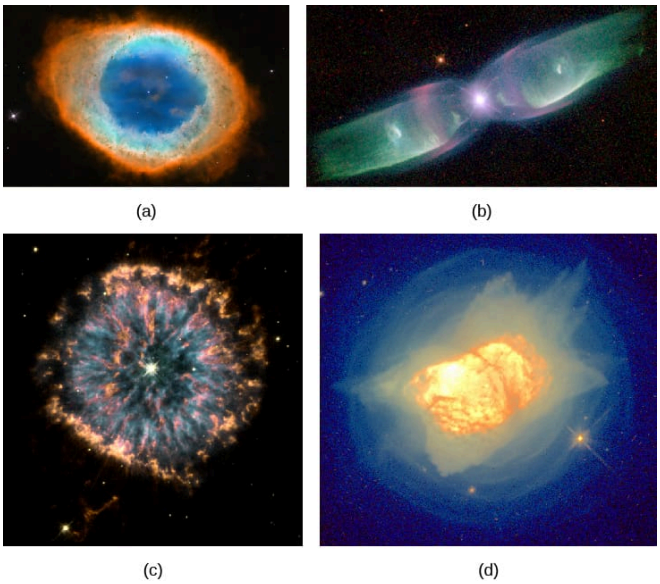


Figure 22.18 : Gallery of Planetary Nebulae. This series of beautiful images depicting some intriguing planetary nebulae highlights the capabilities of the Hubble Space Telescope. (a) Perhaps the best known planetary nebula is the Ring Nebula (M57), located about 2000 light-years away in the constellation of Lyra. The ring is about 1 light-year in diameter, and the central star has a temperature of about 120,000 °C. Careful study of this image has shown scientists that, instead of looking at a spherical shell around this dying star, we may be looking down the barrel of a tube or cone. The blue region shows emission from very hot helium, which is located very close to the star; the red region isolates emission from ionized nitrogen, which is radiated by the coolest gas farthest from the star; and the green region represents oxygen emission, which is produced at intermediate temperatures and is at an intermediate distance from the star. (b) This planetary nebula, M2-9, is an example of a butterfly nebula. The central star (which is part of a binary system) has ejected mass preferentially in two opposite directions. In other images, a disk, perpendicular to the two long streams of gas, can be seen around the two stars in the middle. The stellar outburst that resulted in the expulsion of matter occurred about 1200 years ago. Neutral oxygen is shown in red, once-ionized nitrogen in green, and twice-ionized oxygen in blue. The planetary nebula is about 2100 light-years away in the constellation of Ophiuchus. (c) In this image of the planetary nebula NGC 6751, the blue regions mark the hottest gas, which forms a ring around the central star. The orange and red regions show the locations of cooler gas. The origin of these cool streamers is not known, but their shapes indicate that they are affected by radiation and stellar winds from the hot star at the center. The temperature of the star is about 140,000 °C. The diameter of the nebula is about 600 times larger than the diameter of our solar system. The nebula is about 6500 light-years away in the constellation of Aquila. (d) This image of the planetary nebula NGC 7027 shows several stages of mass loss. The faint blue concentric shells surrounding the central region identify the mass that was shed slowly from the surface of the star when it became a red giant. Somewhat later, the remaining outer layers were ejected but not in a spherically symmetric way. The dense clouds formed by this late ejection produce the bright inner regions. The hot central star can be seen faintly near the center of the nebulosity. NGC 7027 is about 3000 light-years away in the direction of the constellation of Cygnus. (credit a: modification of work by NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration; credit b: modification of work by Bruce Balick (University of Washington), Vincent Icke (Leiden University, The Netherlands), Garrelt Mellema (Stockholm University), and NASA; credit c: modification of work by NASA, The Hubble Heritage Team (STScI/AURA); credit d: modification of work by H. Bond (STScI) and NASA)

As Figure 22.18 shows, sometimes a planetary nebula appears to be a simple ring. Others have faint shells surrounding the bright ring, which is evidence that there were multiple episodes of mass loss when the star was a red giant (see image (d) in Figure 22.18). In a few cases, we see two lobes of matter flowing in opposite directions. Many astronomers think that a considerable number of planetary nebulae basically consist of the same structure, but that the shape we see depends on the viewing angle (Figure 22.19). According to this idea, the dying star is surrounded by a very dense, doughnut-shaped disk of gas. (Theorists do not yet have a definite explanation for why the dying star should produce this ring, but many believe that binary stars, which are common, are involved.)

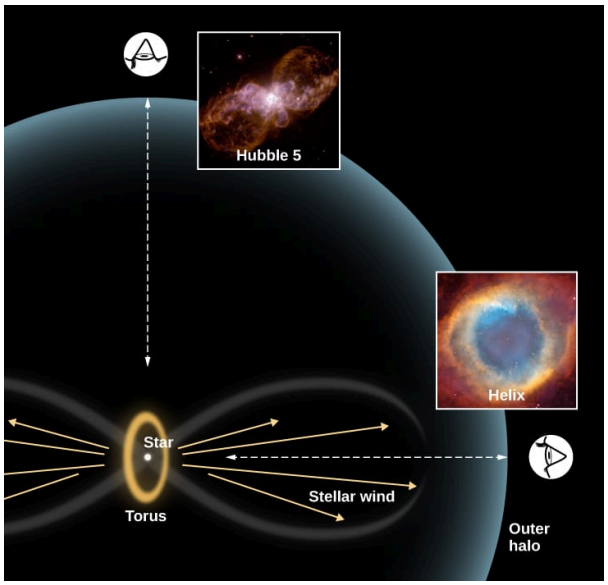


Figure 22.19 : Model to Explain the Different Shapes of Planetary Nebulae. The range of different shapes that we see among planetary nebulae may, in many cases, arise from the same geometric shape, but seen from a variety of viewing directions. The basic shape is a hot central star surrounded by a thick torus (or doughnut-shaped disk) of gas. The star's wind cannot flow out into space very easily in the direction of the torus, but can escape more freely in the two directions perpendicular to it. If we view the nebula along the direction of the flow (Helix Nebula), it will appear nearly circular (like looking directly down into an empty ice-cream cone). If we look along the equator of the torus, we see both outflows and a very elongated shape (Hubble 5). Current research on planetary nebulae focuses on the reasons for having a torus around the star in the first place. Many astronomers suggest that the basic cause may be that many of the central stars are actually close binary stars, rather than single stars. (credit "Hubble 5": modification of work by Bruce Balick (University of Washington), Vincent Icke (Leiden University, The Netherlands), Garrelt Mellema (Stockholm University), and NASA/ESA; credit "Helix": modification of work by NASA, ESA, C.R. O'Dell (Vanderbilt University), and M. Meixner, P. McCullough)

As the star continues to lose mass, any less dense gas that leaves the star cannot penetrate the torus, but the gas *can* flow outward in directions perpendicular to the disk. If we look perpendicular to the direction of outflow, we see the disk and both of the outward flows. If we look "down the barrel" and into the flows, we see a ring. At intermediate angles, we may see wonderfully complex structures. Compare the viewpoints in Figure 22.19 with the images in Figure 22.18.

Planetary nebula shells usually expand at speeds of 20–30 km/s, and a typical planetary nebula has a diameter of about 1 light-year. If we assume that the gas shell has expanded at a constant speed, we can calculate that the shells of all the planetary nebulae visible to us were ejected within the past 50,000 years at most. After this amount of time, the shells have expanded so much that they are too thin and tenuous to be seen. That's a pretty short time that each planetary nebula can be observed (when compared to the whole lifetime of the star). Given the number of such nebulae we nevertheless see, we must conclude that a large fraction of all stars evolve through the planetary nebula phase. Since we saw that low-mass stars are much more common than high-mass stars, this confirms our view of planetary nebulae as sort of "last gasp" of low-mass star evolution.

Cosmic Recycling

The loss of mass by dying stars is a key step in the gigantic cosmic recycling scheme we discussed in Between the Stars: Gas and Dust in Space. Remember that stars form from vast clouds of gas and dust. As they end their lives, stars return part of their gas to the galactic reservoirs of raw material. Eventually, some of the expelled material from aging stars will participate in the formation of new star systems.

However, the atoms returned to the Galaxy by an aging star are not necessarily the same ones it received initially. The star, after all, has fused hydrogen and helium to form new elements over the course of its life. And during the red-giant stage, material from the star's central regions is dredged up and mixed with its outer layers, which can cause further nuclear reactions and the creation of still more new elements. As a result, the winds that blow outward from such stars include atoms that were "newly minted" inside the stars' cores. (As we will see, this mechanism is even more effective for high-mass stars, but it does work for stars with masses like that of the Sun.) In this way, the raw material of the Galaxy is not only resupplied but also receives infusions of new elements. You might say this cosmic recycling plan allows the universe to get more "interesting" all the time.

Making Connections: The Red Giant Sun and the Fate of Earth

How will the evolution of the Sun affect conditions on Earth in the future? Although the Sun has appeared reasonably steady in size and luminosity over recorded human history, that brief span means nothing compared with the timescales we have been discussing. Let's examine the long-term prospects for our planet.

The Sun took its place on the zero-age main sequence approximately 4.5 billion years ago. At that time, it emitted only about 70% of the energy that it radiates today. One might expect that Earth would have been a lot colder than it is now, with the oceans frozen solid. But if this were the case, it would be hard to explain why simple life forms existed when Earth was less than a billion years old. Scientists now think that the explanation may be that much more carbon dioxide was present in Earth's atmosphere when it was young, and that a much stronger greenhouse effect kept Earth warm. (In the greenhouse effect, gases like carbon dioxide or water vapor allow the Sun's light to come in but do not allow the infrared radiation from the ground to escape back into space, so the temperature near Earth's surface increases.)

Carbon dioxide in Earth's atmosphere has steadily declined as the Sun has increased in luminosity. As the brighter Sun increases the temperature of Earth, rocks weather faster and react with carbon dioxide, removing it from the atmosphere. The warmer Sun and the weaker greenhouse effect have kept Earth at a nearly constant temperature for most of its life. This remarkable coincidence, which has resulted in fairly stable climatic conditions, has been the key in the development of complex life-forms on our planet.

As a result of changes caused by the buildup of helium in its core, the Sun will continue to increase in luminosity as it grows older, and more and more radiation will reach Earth. For a while, the amount of carbon dioxide will continue to decrease. (Note that this effect counteracts increases in carbon dioxide from human activities, but on a much-too-slow timescale to undo the changes in climate that are likely to occur in the next 100 years.)

Eventually, the heating of Earth will melt the polar caps and increase the evaporation of the oceans. Water vapor is also an efficient greenhouse gas and will more than compensate for the decrease in carbon dioxide. Sooner or later (atmospheric models are not yet good enough to say exactly when, but estimates range from 500 million to 2 billion years), the increased water vapor will cause a runaway greenhouse effect.

About 1 billion years from now, Earth will lose its water vapor. In the upper atmosphere, sunlight will break down water vapor into hydrogen, and the fast-moving hydrogen atoms will escape into outer space. Like Humpty Dumpty, the water molecules cannot be put back together again. Earth will start to resemble the Venus of today, and temperatures will become much too high for life as we know it.

All of this will happen before the Sun even becomes a red giant. Then the bad news really starts. The Sun, as it expands, will swallow Mercury and Venus, and friction with our star's outer atmosphere will make these planets spiral inward until they are completely vaporized. It is not completely clear whether Earth will escape a similar fate. As described in this chapter, the Sun will lose some of its mass as it becomes a red giant. The gravitational pull of the Sun decreases when it loses mass. The result would be that the diameter of Earth's orbit would increase (remember Kepler's third law). However, recent calculations also show that forces due to the tides raised on the Sun by Earth will act in the opposite direction, causing Earth's orbit to shrink. Thus, many astrophysicists conclude that Earth will be vaporized along with Mercury and Venus. Whether or not this dire prediction is true, there is little doubt that all life on Earth will surely be incinerated. But don't lose any sleep over this—we are talking about events that will occur billions of years from now.

What then are the prospects for preserving Earth life as we know it? The first strategy you might think of would be to move humanity to a more distant and cooler planet. However, calculations indicate that there are long periods of time (several hundred million years) when no planet is habitable. For example, Earth becomes far too warm for life long before Mars warms up enough.

A better alternative may be to move the entire Earth progressively farther from the Sun. The idea is to use gravity in the same way NASA has used it to send spacecraft to distant planets. When a spacecraft flies near a planet, the planet's motion can be used to speed up the spacecraft, slow it down, or redirect it. Calculations show that if we were to redirect an asteroid so that it follows just the right orbit between Earth and Jupiter, it could transfer orbital energy from Jupiter to Earth and move Earth slowly outward, pulling us away from the expanding Sun on each flyby. Since we have hundreds of millions of years to change Earth's orbit, the effect of each flyby need not be large. (Of course, the people directing the asteroid had better get the orbit exactly right and not cause the asteroid to hit Earth.)

> It may seem crazy to think about projects to move an entire planet to a different orbit. But remember that we are talking about the distant future. If, by some miracle, human beings are able to get along for all that time and don't blow ourselves to bits, our technology is likely to be far more sophisticated than it is today. It may also be that if humans survive for hundreds of millions of years, we may spread to planets or habitats around other stars. Indeed, Earth, by then, might be a museum world to which youngsters from other planets return to learn about the origin of our species. It is also possible that evolution will by then have changed us in ways that allow us to survive in very different environments. Wouldn't it be exciting to see how the story of the human race turns out after all those billions of years?

Footnotes

- 2Recall that the force of gravity depends not only on the mass doing the pulling, but also on our distance from the center of gravity. As a red giant star gets a lot bigger, a point on the surface of the star is now farther from the center, and thus has less gravity. That's why the speed needed to escape the star goes down.

This page titled 13.4: Further Evolution of Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 22.5: Further Evolution of Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

13.5: The Evolution of More Massive Stars

Learning Objectives

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

- Explain how and why massive stars evolve much more rapidly than lower-mass stars like our Sun
- Discuss the origin of the elements heavier than carbon within stars

If what we have described so far were the whole story of the evolution of stars and elements, we would have a big problem on our hands. We will see in later chapters that in our best models of the first few minutes of the universe, everything starts with the two simplest elements—hydrogen and helium (plus a tiny bit of lithium). All the predictions of the models imply that no heavier elements were produced at the beginning of the universe. Yet when we look around us on Earth, we see lots of other elements besides hydrogen and helium. These elements must have been made (fused) somewhere in the universe, *and the only place hot enough to make them is inside stars*. One of the fundamental discoveries of twentieth-century astronomy is that the stars are the source of most of the chemical richness that characterizes our world and our lives.

We have already seen that carbon, oxygen, and some other elements are made inside lower-mass stars during the red-giant stage. But where do the heavier elements we know and love (such as the silicon and iron inside Earth, and the gold and silver in our jewelry) come from? The kinds of stars we have been discussing so far never get hot enough at their centers to make these elements. It turns out that these heavier elements are formed later in the lives *more massive* stars, as well as during explosions that can mark the end of the lives of some stars.

Making New Elements in Massive Stars

Massive stars evolve in much the same way that the Sun does (but always more quickly)—up to the formation of a carbon-oxygen core. One difference is that for stars with more than about twice the mass of the Sun, helium begins fusion more gradually, rather than with a sudden flash. Also, when more massive stars become red giants, they become so bright and large that we call them *supergiants*. Such stars can expand until their outer regions become as large as the orbit of Jupiter, which is precisely what the Hubble Space Telescope has shown for the star Betelgeuse (see Figure 22.4). They also lose mass very effectively, producing dramatic winds and outbursts as they age. Figure 22.20 shows a wonderful image of the very massive star Eta Carinae, with a great deal of ejected material clearly visible.

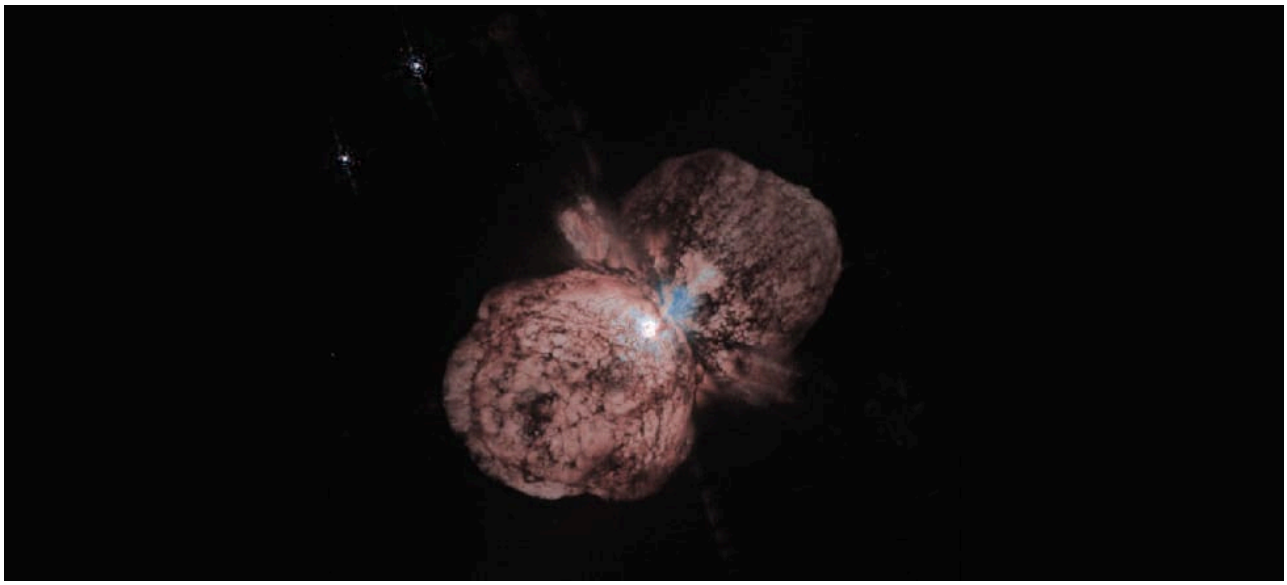


Figure 22.20 : Eta Carinae. With a mass at least 100 times that of the Sun, the hot supergiant Eta Carinae is one of the most massive stars known. This Hubble Space Telescope image records the two giant lobes and equatorial disk of material it has ejected in the course of its evolution. The pink outer region is material ejected in an outburst seen in 1843, the largest of such mass loss event that any star is known to have survived. Moving away from the star at a speed of about 1000 km/s, the material is rich in nitrogen and other elements formed in the interior of the star. The inner blue-white region is the material ejected at lower speeds and is thus still closer to the star. It appears blue-white because it contains dust and reflects the light of Eta Carinae, whose luminosity is 4 million times that of our Sun. (credit: modification of work by Jon Morse (University of Colorado) & NASA)

But the crucial way that massive stars diverge from the story we have outlined is that they can start additional kinds of fusion in their centers and in the shells surrounding their central regions. The outer layers of a star with a mass greater than about 8 solar masses have a weight that is enough to compress the carbon-oxygen core until it becomes hot enough to ignite fusion of carbon nuclei. Carbon can fuse into still more oxygen, and at still higher temperatures, oxygen and then neon, magnesium, and finally silicon can build even heavier elements. Iron is, however, the endpoint of this process. The fusion of iron atoms produces products that are *more* massive than the nuclei that are being fused and therefore the process *require*s energy, as opposed to releasing energy, which all fusion reactions up to this point have done. This required energy comes at the expense of the star itself, which is now on the brink of death (Figure 22.21). What happens next will be described in the chapter on The Death of Stars.

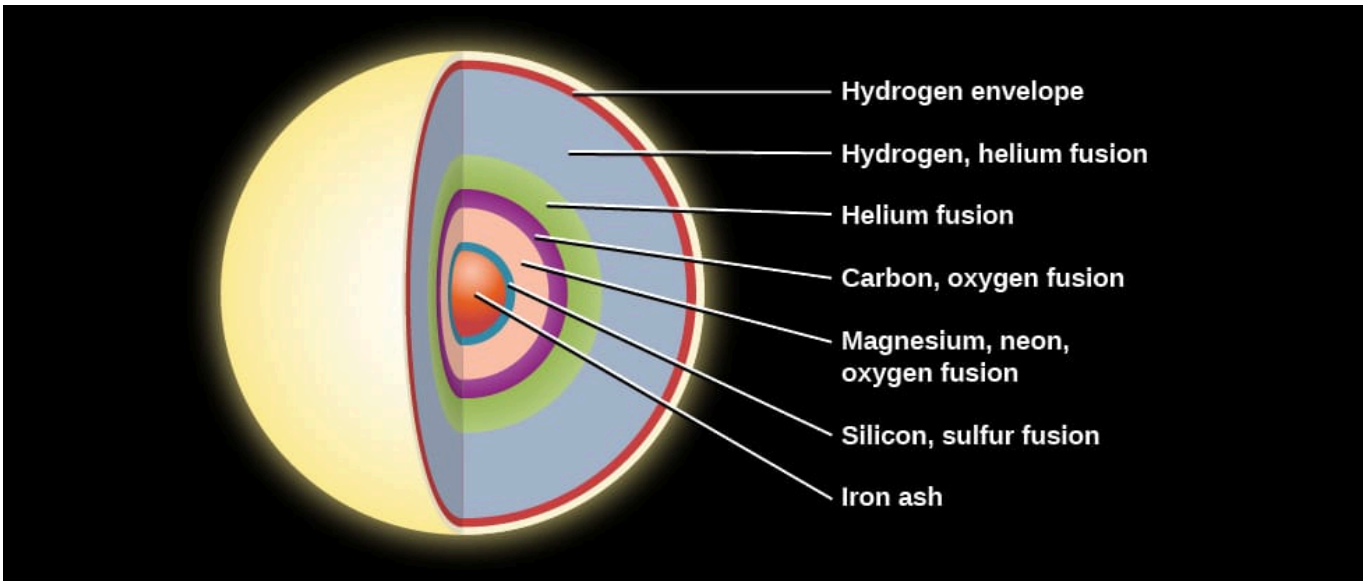


Figure 22.21 : Interior Structure of a Massive Star Just before It Exhausts Its Nuclear Fuel. High-mass stars can fuse elements heavier than carbon. As a massive star nears the end of its evolution, its interior resembles an onion. Hydrogen fusion is taking place in an outer shell, and progressively heavier elements are undergoing fusion in the higher-temperature layers closer to the center. All of these fusion reactions generate energy and enable the star to continue shining. Iron is different. The fusion of iron requires energy, and when iron is finally created in the core, the star has only minutes to live.

Physicists have now found nuclear pathways whereby virtually all chemical elements of atomic weights up to that of iron can be built up by this nucleosynthesis (the making of new atomic nuclei) in the centers of the more massive red giant stars. This still leaves the question of where elements *heavier* than iron come from. We will see in the next chapter that when massive stars finally exhaust their nuclear fuel, they most often die in a spectacular explosion—a supernova. Heavier elements can be synthesized in the stunning violence of such explosions.

Not only can we explain in this way where the elements that make up our world and others come from, but our theories of nucleosynthesis inside stars are even able to predict the *relative abundances* with which the elements occur in nature. The way stars build up elements during various nuclear reactions really can explain why some elements (oxygen, carbon, and iron) are common and others are quite rare (gold, silver, and uranium).

Elements in Globular Clusters and Open Clusters Are Not the Same

The fact that the elements are made in stars over time explains an important difference between globular and open clusters. Hydrogen and helium, which are the most abundant elements in stars in the solar neighborhood, are also the most abundant constituents of stars in both kinds of clusters. However, the abundances of the elements *heavier* than helium are very different.

In the Sun and most of its neighboring stars, the combined abundance (by mass) of the elements heavier than hydrogen and helium is 1–4% of the star's mass. Spectra show that most open-cluster stars also have 1–4% of their matter in the form of heavy elements. Globular clusters, however, are a different story. The heavy-element abundance of stars in typical globular clusters is found to be only 1/10 to 1/100 that of the Sun. A few very old stars not in clusters have been discovered with even lower abundances of heavy elements.

The differences in chemical composition are a direct consequence of the formation of a cluster of stars. The very first generation of stars initially contained only hydrogen and helium. We have seen that these stars, in order to generate energy, created heavier elements in their interiors. In the last stages of their lives, they ejected matter, now enriched in heavy elements, into the reservoirs of raw material between the stars. Such matter was then incorporated into a new generation of stars.

This means that the relative abundance of the heavy elements must be less and less as we look further into the past. We saw that the globular clusters are much older than the open clusters. Since globular-cluster stars formed much earlier (that is, they are an earlier generation of stars) than those in open clusters, they have only a relatively small abundance of elements heavier than hydrogen and helium.

As time passes, the proportion of heavier elements in the “raw material” that makes new stars and planets increases. This means that the first generation of stars that formed in our Galaxy would not have been accompanied by a planet like Earth, full of silicon, iron, and many other heavy elements. Earth (and the astronomy students who live on it) was possible only after generations of stars had a chance to make and recycle their heavier elements.

Now the search is on for true *first*-generation stars, made only of hydrogen and helium. Theories predict that such stars should be very massive, live fast, and die quickly. They should have lived and died long ago. The place to look for them is in very distant galaxies that formed when the universe was only a few hundred million years old, but whose light is only arriving at Earth now.

Approaching Death

Compared with the main-sequence lifetimes of stars, the events that characterize the last stages of stellar evolution pass very quickly (especially for massive stars). As the star’s luminosity increases, its rate of nuclear fuel consumption goes up rapidly—just at that point in its life when its fuel supply is beginning to run down.

After the prime fuel—hydrogen—is exhausted in a star’s core, we saw that other sources of nuclear energy are available to the star in the fusion of, first, helium, and then of other more complex elements. But the energy yield of these reactions is much less than that of the fusion of hydrogen to helium. And to trigger these reactions, the central temperature must be higher than that required for the fusion of hydrogen to helium, leading to even more rapid consumption of fuel. Clearly this is a losing game, and very quickly the star reaches its end. As it does so, however, some remarkable things can happen, as we will see in The Death of Stars.

This page titled 13.5: The Evolution of More Massive Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 22.6: The Evolution of More Massive Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14: WHITE DWARFS, NEUTRON STARS, AND BLACK HOLES

14.1: The Death of Low-Mass Stars
14.2: Evolution of Massive Stars- An Explosive Finish
14.3: Supernova Observations
14.4: Pulsars and the Discovery of Neutron Stars
14.5: The Evolution of Binary Star Systems
14.6: Black Holes
14.7: Evidence for Black Holes

This page titled 14: White Dwarfs, Neutron Stars, and Black Holes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

14.1: The Death of Low-Mass Stars

Learning Objectives

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

- Describe the physical characteristics of degenerate matter and explain how the mass and radius of degenerate stars are related
- Plot the future evolution of a white dwarf and show how its observable features will change over time
- Distinguish which stars will become white dwarfs

Let's begin with those stars whose final mass just before death is less than about 1.4 times the mass of the Sun (M_{Sun}). (We will explain why this mass is the crucial dividing line in a moment.) Note that most stars in the universe fall into this category. The number of stars decreases as mass increases; really massive stars are rare (see The Stars: A Celestial Census). This is similar to the music business where only a few musicians ever become superstars. Furthermore, many stars with an initial mass much greater than 1.4 M_{Sun} will be reduced to that level by the time they die. For example, we now know that stars that start out with masses of at least 8.0 M_{Sun} (and possibly as much as 10 M_{Sun}) manage to lose enough mass during their lives to fit into this category (an accomplishment anyone who has ever attempted to lose weight would surely envy).

A Star in Crisis

In the last chapter, we left the life story of a star with a mass like the Sun's just after it had climbed up to the red-giant region of the H–R diagram for a second time and had shed some of its outer layers to form a planetary nebula. Recall that during this time, the *core* of the star was undergoing an "energy crisis." Earlier in its life, during a brief stable period, helium in the core had gotten hot enough to fuse into carbon (and oxygen). But after this helium was exhausted, the star's core had once more found itself without a source of pressure to balance gravity and so had begun to contract.

This collapse is the final event in the life of the core. Because the star's mass is relatively low, it cannot push its core temperature high enough to begin another round of fusion (in the same way larger-mass stars can). The core continues to shrink until it reaches a density equal to nearly a million times the density of water! That is 200,000 times greater than the average density of Earth. At this extreme density, a new and different way for matter to behave kicks in and helps the star achieve a final state of equilibrium. In the process, what remains of the star becomes one of the strange *white dwarfs* that we met in The Stars: A Celestial Census.

Degenerate Stars

Because white dwarfs are far denser than any substance on Earth, the matter inside them behaves in a very unusual way—unlike anything we know from everyday experience. At this high density, gravity is incredibly strong and tries to shrink the star still further, but all the *electrons* resist being pushed closer together and set up a powerful pressure inside the core. This pressure is the result of the fundamental rules that govern the behavior of electrons (the quantum physics you were introduced to in The Sun: A Nuclear Powerhouse). According to these rules (known to physicists as the Pauli exclusion principle), which have been verified in studies of atoms in the laboratory, no two electrons can be in the same place at the same time doing the same thing. We specify the *place* of an electron by its position in space, and we specify what it is doing by its motion and the way it is spinning.

The temperature in the interior of a star is always so high that the atoms are stripped of virtually all their electrons. For most of a star's life, the density of matter is also relatively low, and the electrons in the star are moving rapidly. This means that no two of them will be in the same place moving in exactly the same way at the same time. But this all changes when a star exhausts its store of nuclear energy and begins its final collapse.

As the star's core contracts, electrons are squeezed closer and closer together. Eventually, a star like the Sun becomes so dense that further contraction would in fact require two or more electrons to violate the rule against occupying the same place and moving in the same way. Such a dense gas is said to be degenerate (a term coined by physicists and not related to the electron's moral character). The electrons in a degenerate gas resist further crowding with tremendous pressure. (It's as if the electrons

said, “You can press inward all you want, but there is simply no room for any other electrons to squeeze in here without violating the rules of our existence.”)

The degenerate electrons do not require an input of heat to maintain the pressure they exert, and so a star with this kind of structure, if nothing disturbs it, can last essentially forever. (Note that the repulsive force between degenerate electrons is different from, and much stronger than, the normal electrical repulsion between charges that have the same sign.)

The electrons in a degenerate gas do move about, as do particles in any gas, but not with a lot of freedom. A particular electron cannot change position or momentum until another electron in an adjacent stage gets out of the way. The situation is much like that in the parking lot after a big football game. Vehicles are closely packed, and a given car cannot move until the one in front of it moves, leaving an empty space to be filled.

Of course, the dying star also has atomic nuclei in it, not just electrons, but it turns out that the nuclei must be squeezed to much higher densities before their quantum nature becomes apparent. As a result, in white dwarfs, the nuclei do not exhibit degeneracy pressure. Hence, in the white dwarf stage of stellar evolution, it is the degeneracy pressure of the electrons, and not of the nuclei, that halts the collapse of the core.

White Dwarfs

White dwarfs, then, are stable, compact objects with electron-degenerate cores that cannot contract any further. Calculations showing that white dwarfs are the likely end state of low-mass stars were first carried out by the Indian-American astrophysicist Subrahmanyan Chandrasekhar. He was able to show how much a star will shrink before the degenerate electrons halt its further contraction and hence what its final diameter will be (Figure 23.2).

When Chandrasekhar made his calculation about white dwarfs, he found something very surprising: the radius of a white dwarf shrinks as the mass in the star increases (the larger the mass, the more tightly packed the electrons can become, resulting in a smaller radius). According to the best theoretical models, a white dwarf with a mass of about 1.4 M_{Sun} or larger would have a radius of zero. What the calculations are telling us is that even the force of degenerate electrons cannot stop the collapse of a star with more mass than this. The maximum mass that a star can end its life with and still become a white dwarf—1.4 M_{Sun}—is called the Chandrasekhar limit. Stars with end-of-life masses that exceed this limit have a different kind of end in store—one that we will explore in the next section.

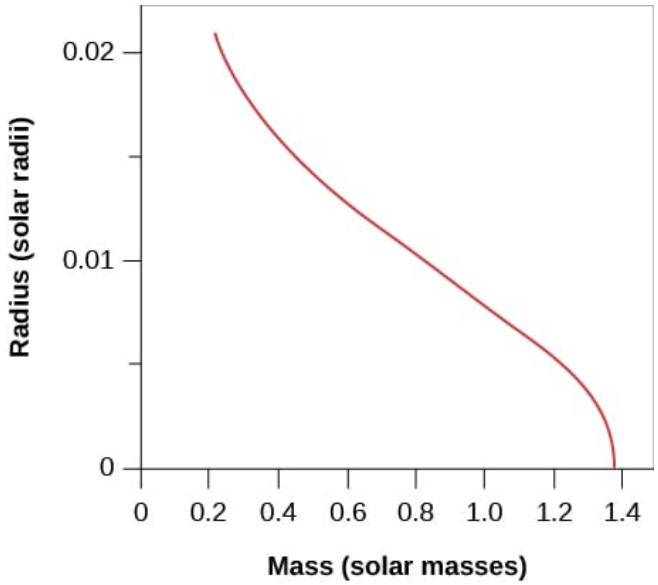


Figure 23.2 : Relating Masses and Radii of White Dwarfs. Models of white-dwarf structure predict that as the mass of the star increases (toward the right), its radius gets smaller and smaller.

Voyagers in Astronomy: Subrahmanyan Chandrasekhar

Born in 1910 in Lahore, India, Subrahmanyan Chandrasekhar (known as Chandra to his friends and colleagues) grew up in a home that encouraged scholarship and an interest in science (Figure 23.3). His uncle, C. V. Raman, was a physicist who won the 1930 Nobel Prize. A precocious student, Chandra tried to read as much as he could about the latest ideas in physics and astronomy, although obtaining technical books was not easy in India at the time. He finished college at age 19 and won a scholarship to study in England. It was during the long boat voyage to get to graduate school that he first began doing calculations about the structure of white dwarf stars.

Chandra developed his ideas during and after his studies as a graduate student, showing—as we have discussed—that white dwarfs with masses greater than 1.4 times the mass of the Sun cannot exist and that the theory predicts the existence of other kinds of stellar corpses. He wrote later that he felt very shy and lonely during this period, isolated from students, afraid to assert himself, and sometimes waiting for hours to speak with some of the famous professors he had read about in India. His calculations soon brought him into conflict with certain distinguished astronomers, including Sir Arthur Eddington, who publicly ridiculed Chandra's ideas. At a number of meetings of astronomers, such leaders in the field as Henry Norris Russell refused to give Chandra the opportunity to defend his ideas, while allowing his more senior critics lots of time to criticize them.

Yet Chandra persevered, writing books and articles elucidating his theories, which turned out not only to be correct, but to lay the foundation for much of our modern understanding of the death of stars. In 1983, he received the Nobel Prize in physics for this early work.

In 1937, Chandra came to the United States and joined the faculty at the University of Chicago, where he remained for the rest of his life. There he devoted himself to research and teaching, making major contributions to many fields of astronomy, from our understanding of the motions of stars through the Galaxy to the behavior of the bizarre objects called black holes (see Black Holes and Curved Spacetime). In 1999, NASA named its sophisticated orbiting X-ray telescope (designed in part to explore such stellar corpses) the Chandra X-ray Observatory.



Figure 23.3 : S. Chandrasekhar (1910–1995). Chandra's research provided the basis for much of what we now know about stellar corpses. (credit: modification of work by American Institute of Physics)

Chandra spent a great deal of time with his graduate students, supervising the research of more than 50 PhDs during his life. He took his teaching responsibilities very seriously: during the 1940s, while based at the Yerkes Observatory, he willingly drove the more than 100-mile trip to the university each week to teach a class of only a few students.

Chandra also had a deep devotion to music, art, and philosophy, writing articles and books about the relationship between the humanities and science. He once wrote that "one can learn science the way one enjoys music or art. . . . Heisenberg had a marvelous phrase 'shuddering before the beautiful'. . . that is the kind of feeling I have."

Link to Learning

Using the Hubble Space Telescope, astronomers were able to detect images of faint white dwarf stars and other “stellar corpses” in the M4 star cluster, located about 7200 light-years away.

The Ultimate Fate of White Dwarfs

If the birth of a main-sequence star is defined by the onset of fusion reactions, then we must consider the end of all fusion reactions to be the time of a star’s death. As the core is stabilized by degeneracy pressure, a last shudder of fusion passes through the outside of the star, consuming the little hydrogen still remaining. Now the star is a true white dwarf: nuclear fusion in its interior has ceased. Figure 23.4 shows the path of a star like the Sun on the H–R diagram during its final stages.

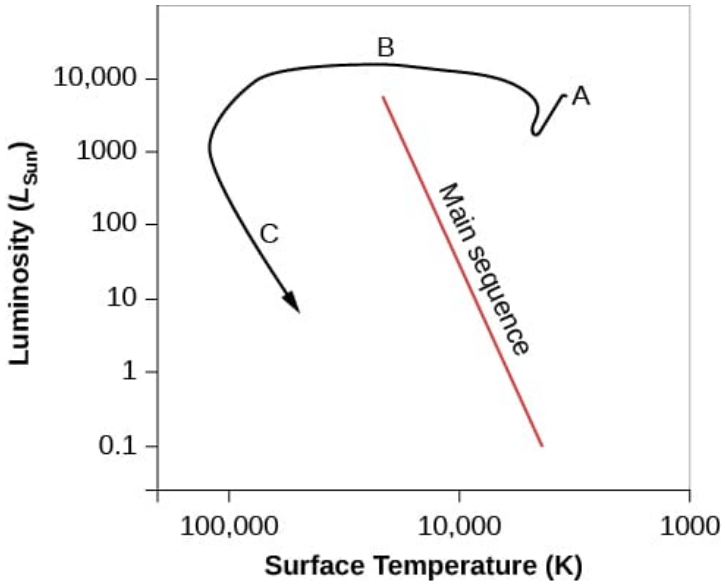


Figure 23.4 : Evolutionary Track for a Star Like the Sun. This diagram shows the changes in luminosity and surface temperature for a star with a mass like the Sun’s as it nears the end of its life. After the star becomes a giant again (point A on the diagram), it will lose more and more mass as its core begins to collapse. The mass loss will expose the hot inner core, which will appear at the center of a planetary nebula. In this stage, the star moves across the diagram to the left as it becomes hotter and hotter during its collapse (point B). At first, the luminosity remains nearly constant, but as the star begins to cool off, it becomes less and less bright (point C). It is now a white dwarf and will continue to cool slowly for billions of years until all of its remaining store of energy is radiated away. (This assumes the Sun will lose between 46–50% of its mass during the giant stages, based upon various theoretical models).

Since a stable white dwarf can no longer contract or produce energy through fusion, its only energy source is the heat represented by the motions of the atomic nuclei in its interior. The light it emits comes from this internal stored heat, which is substantial. Gradually, however, the white dwarf radiates away all its heat into space. After many billions of years, the nuclei will be moving much more slowly, and the white dwarf will no longer shine (Figure 23.5). It will then be a black dwarf—a cold stellar corpse with the mass of a star and the size of a planet. It will be composed mostly of carbon, oxygen, and neon, the products of the most advanced fusion reactions of which the star was capable.

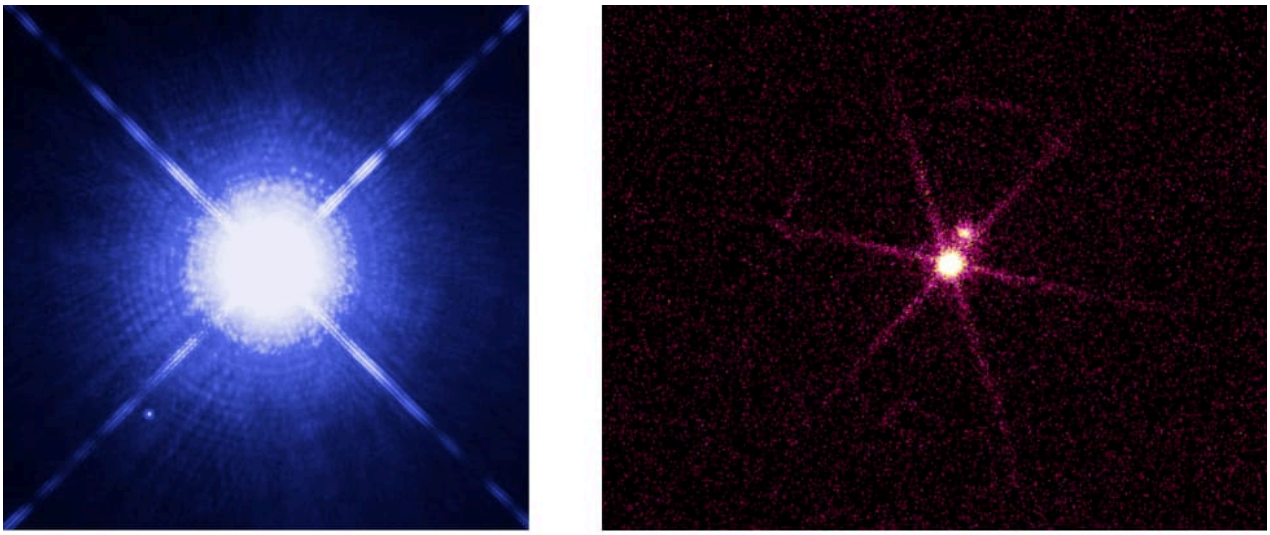


Figure 23.5 : Visible Light and X-Ray Images of the Sirius Star System. (a) This image taken by the Hubble Space Telescope shows Sirius A (the large bright star), and its companion star, the white dwarf known as Sirius B (the tiny, faint star at the lower left). Sirius A and B are 8.6 light-years from Earth and are our fifth-closest star system. Note that the image has intentionally been overexposed to allow us to see Sirius B. (b) The same system is shown in X-ray taken with the Chandra Space Telescope. Note that Sirius A is fainter in X-rays than the hot white dwarf that is Sirius B. (credit a: modification of work by NASA, ESA, H. Bond, M. Barstow(University of Leicester); credit b: modification of work by NASA/SAO/CXC)

We have one final surprise as we leave our low-mass star in the stellar graveyard. Calculations show that as a degenerate star cools, the atoms inside it in essence "solidify" into a giant, highly compact lattice (organized rows of atoms, just like in a crystal). When carbon is compressed and crystallized in this way, it becomes a giant *diamond-like* star. A white dwarf star might make the most impressive engagement present you could ever see, although any attempt to mine the diamond-like material inside would crush an ardent lover instantly!

Link to Learning

Learn about a recent "diamond star" find, a cold, white dwarf star detected in 2014, which is considered the coldest and dimmest found to date, at the website of the National Radio Astronomy Observatory.

Evidence That Stars Can Shed a Lot of Mass as They Evolve

Whether or not a star will become a white dwarf depends on how much mass is lost in the red-giant and earlier phases of evolution. All stars that have masses below the Chandrasekhar limit when they run out of fuel will become white dwarfs, no matter what mass they were born with. But which stars shed enough mass to reach this limit?

One strategy for answering this question is to look in young, open clusters (which were discussed in Star Clusters). The basic idea is to search for young clusters that contain one or more white dwarf stars. Remember that more massive stars go through all stages of their evolution more rapidly than less massive ones. Suppose we find a cluster that has a white dwarf member and also contains stars on the main sequence that have 6 times the mass of the Sun. This means that only those stars with masses greater than 6 M_{Sun} have had time to exhaust their supply of nuclear energy and complete their evolution to the white dwarf stage. The star that turned into the white dwarf must therefore have had a main-sequence mass of more than 6 M_{Sun}, since stars with lower masses have not yet had time to use up their stores of nuclear energy. The star that became the white dwarf must, therefore, have gotten rid of at least 4.6 M_{Sun} so that its mass at the time nuclear energy generation ceased could be less than 1.4 M_{Sun}.

Astronomers continue to search for suitable clusters to make this test, and the evidence so far suggests that stars with masses up to about 8 M_{Sun} can shed enough mass to end their lives as white dwarfs. Stars like the Sun will probably lose about 45% of their initial mass and become white dwarfs with masses less than 1.4 M_{Sun}.

This page titled 14.1: The Death of Low-Mass Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 23.2: The Death of Low-Mass Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14.2: Evolution of Massive Stars- An Explosive Finish

Learning Objectives

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

- Describe the interior of a massive star before a supernova
- Explain the steps of a core collapse and explosion
- List the hazards associated with nearby supernovae

Thanks to mass loss, then, stars with starting masses of at least 8 M_{Sun} (and perhaps even more) probably end their lives as white dwarfs. But we know stars can have masses as large as 150 (or more) M_{Sun}. They have a different kind of death in store for them. As we will see, these stars die with a bang.

Nuclear Fusion of Heavy Elements

After the helium in its core is exhausted (see The Evolution of More Massive Stars), the evolution of a massive star takes a significantly different course from that of lower-mass stars. In a massive star, the weight of the outer layers is sufficient to force the carbon core to contract until it becomes hot enough to fuse carbon into oxygen, neon, and magnesium. This cycle of contraction, heating, and the ignition of another nuclear fuel repeats several more times. After each of the possible nuclear fuels is exhausted, the core contracts again until it reaches a new temperature high enough to fuse still-heavier nuclei. The products of carbon fusion can be further converted into silicon, sulfur, calcium, and argon. And these elements, when heated to a still-higher temperature, can combine to produce iron. Massive stars go through these stages very, very quickly. In really massive stars, some fusion stages toward the very end can take only months or even days! This is a far cry from the millions of years they spend in the main-sequence stage.

At this stage of its evolution, a massive star resembles an onion with an iron core. As we get farther from the center, we find shells of decreasing temperature in which nuclear reactions involve nuclei of progressively lower mass—silicon and sulfur, oxygen, neon, carbon, helium, and finally, hydrogen (Figure 23.6).

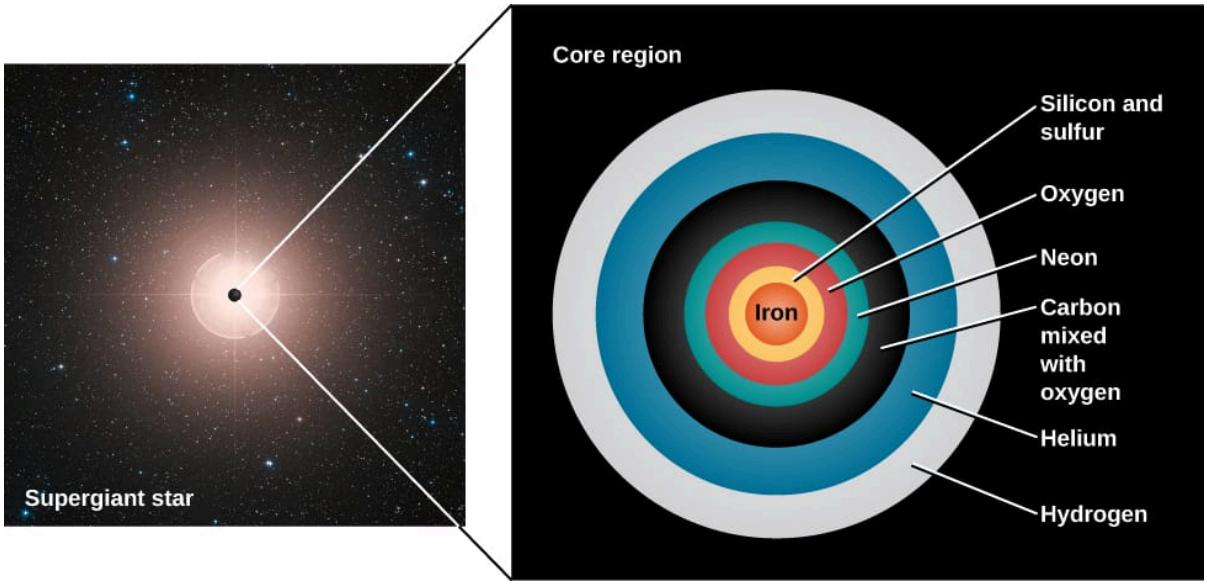


Figure 23.6 : Structure of an Old Massive Star. Just before its final gravitational collapse, the core of a massive star resembles an onion. The iron core is surrounded by layers of silicon and sulfur, oxygen, neon, carbon mixed with some oxygen, helium, and finally hydrogen. Outside the core, the composition is mainly hydrogen and helium. (Note that this diagram is not precisely to scale but is just meant to convey the general idea of what such a star would be like.) (credit: modification of work by ESO, Digitized Sky Survey)

But there is a limit to how long this process of building up elements by fusion can go on. The fusion of silicon into iron turns out to be the last step in the sequence of nonexplosive element production. Up to this point, each fusion reaction has *produced* energy because the nucleus of each fusion product has been a bit more stable than the nuclei that formed it. As discussed in The Sun: A Nuclear Powerhouse, light nuclei give up some of their binding energy in the process of fusing into more tightly bound, heavier nuclei. It is this released energy that maintains the outward pressure in the core so that the star does not collapse. But of all the nuclei known, iron is the most tightly bound and thus the most stable.

You might think of the situation like this: all smaller nuclei want to "grow up" to be like iron, and they are willing to pay (*produce* energy) to move toward that goal. But iron is a mature nucleus with good self-esteem, perfectly content being iron; it requires payment (must *absorb* energy) to change its stable nuclear structure. This is the exact opposite of what has happened in each nuclear reaction so far: instead of *providing* energy to balance the inward pull of gravity, any nuclear reactions involving iron would *remove* some energy from the core of the star.

Unable to generate energy, the star now faces catastrophe.

Collapse into a Ball of Neutrons

When nuclear reactions stop, the core of a massive star is supported by degenerate electrons, just as a white dwarf is. For stars that begin their evolution with masses of at least 10 M_{Sun}, this core is likely made mainly of iron. (For stars with initial masses in the range 8 to 10 M_{Sun}, the core is likely made of oxygen, neon, and magnesium, because the star never gets hot enough to form elements as heavy as iron. The exact composition of the cores of stars in this mass range is very difficult to determine because of the complex physical characteristics in the cores, particularly at the very high densities and temperatures involved.) We will focus on the more massive iron cores in our discussion.

While no energy is being generated within the white dwarf core of the star, fusion still occurs in the shells that surround the core. As the shells finish their fusion reactions and stop producing energy, the ashes of the last reaction fall onto the white dwarf core, increasing its mass. As Figure 23.2 shows, a higher mass means a smaller core. The core can contract because even a degenerate gas is still mostly empty space. Electrons and atomic nuclei are, after all, extremely small. The electrons and nuclei in a stellar core may be crowded compared to the air in your room, but there is still lots of space between them.

The electrons at first resist being crowded closer together, and so the core shrinks only a small amount. Ultimately, however, the iron core reaches a mass so large that even degenerate electrons can no longer support it. When the density reaches 4×10^{11} g/cm^3 (400 billion times the density of water), some electrons are actually squeezed into the atomic nuclei, where they combine with protons to form neutrons and neutrinos. This transformation is not something that is familiar from everyday life, but becomes very important as such a massive star core collapses.

Some of the electrons are now gone, so the core can no longer resist the crushing mass of the star's overlying layers. The core begins to shrink rapidly. More and more electrons are now pushed into the atomic nuclei, which ultimately become so saturated with neutrons that they cannot hold onto them.

At this point, the neutrons are squeezed out of the nuclei and can exert a new force. As is true for electrons, it turns out that the neutrons strongly resist being in the same place and moving in the same way. The force that can be exerted by such *degenerate neutrons* is much greater than that produced by degenerate electrons, so unless the core is too massive, they can ultimately stop the collapse.

This means the collapsing core can reach a stable state as a crushed ball made mainly of neutrons, which astronomers call a neutron star. We don't have an exact number (a "Chandrasekhar limit") for the maximum mass of a neutron star, but calculations tell us that the upper mass limit of a body made of neutrons might only be about 3 M_{Sun}. So if the mass of the core were greater than this, then even neutron degeneracy would not be able to stop the core from collapsing further. The dying star must end up as something even more extremely compressed, which until recently was believed to be only one possible type of object—the state of ultimate compaction known as a black hole (which is the subject of our next chapter). This is because no force was believed to exist that could stop a collapse beyond the neutron star stage.

Collapse and Explosion

When the collapse of a high-mass star's core is stopped by degenerate neutrons, the core is saved from further destruction, but

it turns out that the rest of the star is literally blown apart. Here's how it happens.

The collapse that takes place when electrons are absorbed into the nuclei is very rapid. In less than a second, a core with a mass of about 1 M_{Sun}, which originally was approximately the size of Earth, collapses to a diameter of less than 20 kilometers. The speed with which material falls inward reaches one-fourth the speed of light. The collapse halts only when the density of the core exceeds the density of an atomic nucleus (which is the densest form of matter we know). A typical neutron star is so compressed that to duplicate its density, we would have to squeeze all the people in the world into a single sugar cube! This would give us one sugar cube's worth (one cubic centimeter's worth) of a neutron star.

The neutron degenerate core strongly resists further compression, abruptly halting the collapse. The shock of the sudden jolt initiates a shock wave that starts to propagate outward. However, this shock alone is not enough to create a star explosion. The energy produced by the outflowing matter is quickly absorbed by atomic nuclei in the dense, overlying layers of gas, where it breaks up the nuclei into individual neutrons and protons.

Our understanding of nuclear processes indicates (as we mentioned above) that each time an electron and a proton in the star's core merge to make a neutron, the merger releases a *neutrino*. These ghostly subatomic particles, introduced in The Sun: A Nuclear Powerhouse, carry away some of the nuclear energy. It is their presence that launches the final disastrous explosion of the star. The total energy contained in the neutrinos is huge. In the initial second of the star's explosion, the power carried by the neutrinos (10^{46} watts) is greater than the power put out by all the stars in over a billion galaxies.

While neutrinos ordinarily do not interact very much with ordinary matter (we earlier accused them of being downright antisocial), matter near the center of a collapsing star is so dense that the neutrinos do interact with it to some degree. They deposit some of this energy in the layers of the star just outside the core. This huge, sudden input of energy reverses the infall of these layers and drives them explosively outward. Most of the mass of the star (apart from that which went into the neutron star in the core) is then ejected outward into space. As we saw earlier, such an explosion requires a star of at least 8 M_{Sun}, and the neutron star can have a mass of at most 3 M_{Sun}. Consequently, at least five times the mass of our Sun is ejected into space in each such explosive event!

The resulting explosion is called a supernova (Figure 23.7). When these explosions happen close by, they can be among the most spectacular celestial events, as we will discuss in the next section. (Actually, there are at least two different types of supernova explosions: the kind we have been describing, which is the collapse of a massive star, is called, for historical reasons, a type II supernova. We will describe how the types differ later in this chapter).

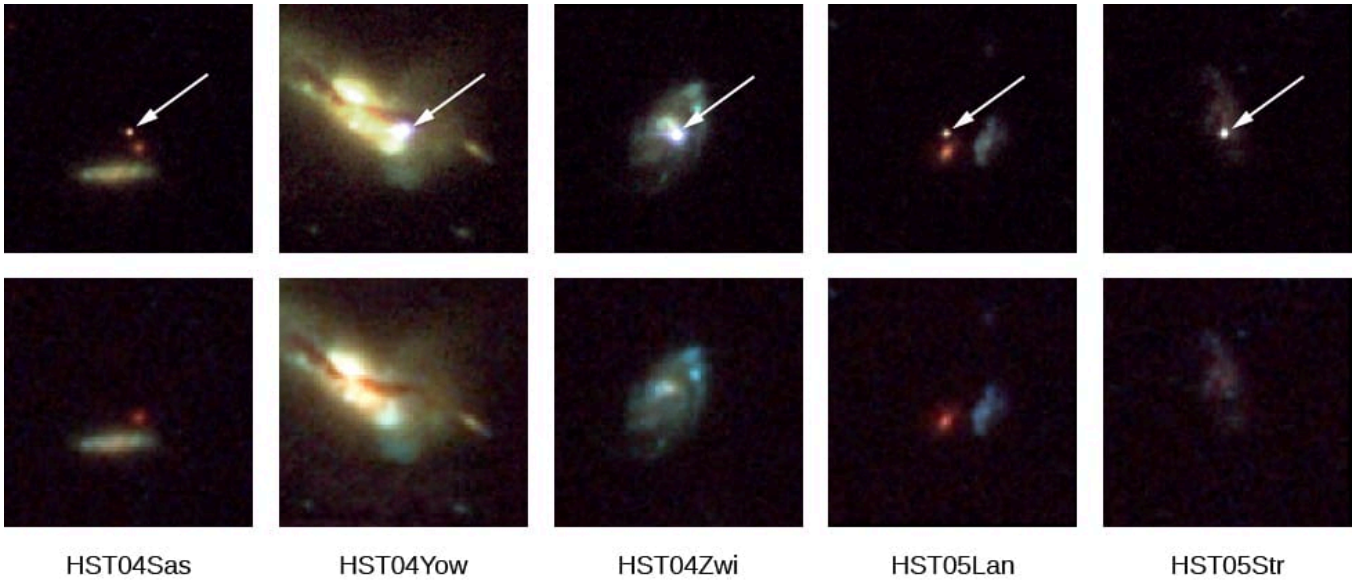


Figure 23.7 : Five Supernova Explosions in Other Galaxies. The arrows in the top row of images point to the supernovae. The bottom row shows the host galaxies before or after the stars exploded. Each of these supernovae exploded between 3.5 and 10 billion years ago. Note that the supernovae when they first explode can be as bright as an entire galaxy. (credit: modification of work by NASA, ESA, and A. Riess (STScI))

Table 23.1 summarizes the discussion so far about what happens to stars and substellar objects of different initial masses at the ends of their lives. Like so much of our scientific understanding, this list represents a progress report: it is the best we can do with our present models and observations. The mass limits corresponding to various outcomes may change somewhat as models

are improved. There is much we do not yet understand about the details of what happens when stars die.

Table 23.1: The Ultimate Fate of Stars and Substellar Objects with Different Masses

Initial Mass (Mass of Sun = 1)[1]	Final State at the End of Its Life
< 0.01	Planet
0.01 to 0.08	Brown dwarf
0.08 to 0.25	White dwarf made mostly of helium
0.25 to 8	White dwarf made mostly of carbon and oxygen
8 to 10	White dwarf made of oxygen, neon, and magnesium
10 to 40	Supernova explosion that leaves a neutron star
> 40	Supernova explosion that leaves a black hole

The Supernova Giveth and the Supernova Taketh Away

After the supernova explosion, the life of a massive star comes to an end. But the death of each massive star is an important event in the history of its galaxy. The elements built up by fusion during the star's life are now "recycled" into space by the explosion, making them available to enrich the gas and dust that form new stars and planets. Because these heavy elements ejected by supernovae are critical for the formation of planets and the origin of life, it's fair to say that without mass loss from supernovae and planetary nebulae, neither the authors nor the readers of this book would exist.[2]

But the supernova explosion has one more creative contribution to make, one we alluded to in Stars from Adolescence to Old Age when we asked where the atoms in your jewelry came from. The supernova explosion produces a flood of energetic neutrons that barrel through the expanding material. These neutrons can be absorbed by iron and other nuclei where they can turn into protons. Thus, they can build up elements that are more massive than iron, possibly including such terrestrial favorites as gold, silver and uranium. Supernovae (and, as we will shortly see, the explosive mergers of neutron stars) are the only candidates we have for places where such heavier atoms can be made. Next time you wear some gold jewelry (or give some to your sweetheart), bear in mind that those gold atoms were forged long ago in these kinds of celestial explosions!

When supernovae explode, these elements (as well as the ones the star made during more stable times) are ejected into the existing gas between the stars and mixed with it. Thus, supernovae play a crucial role in enriching their galaxy with heavier elements, allowing, among other things, the chemical elements that make up earthlike planets and the building blocks of life to become more common as time goes on (Figure 23.8).

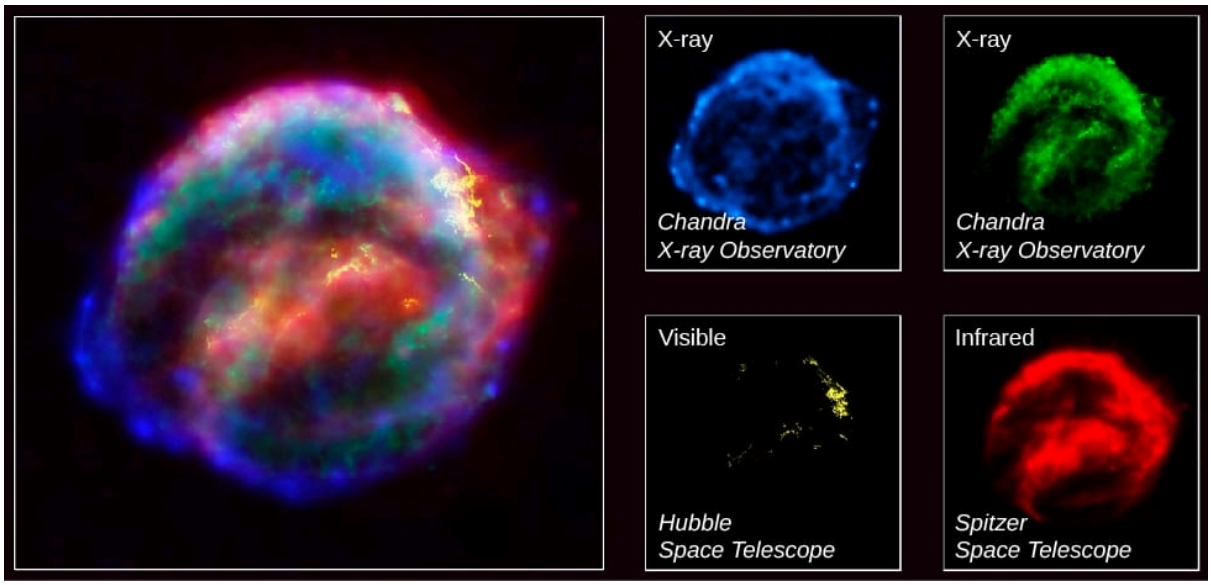


Figure 23.8 : Kepler Supernova Remant. This image shows the expanding remains of a supernova explosion, which was first seen about 400 years ago by sky watchers, including the famous astronomer Johannes Kepler. The bubble-shaped shroud of gas and dust is now 14 light-years wide and is expanding at 2,000 kilometers per second (4 million miles per hour). The remnant emits energy at wavelengths from X-rays (shown in blue and green) to visible light (yellow) and into the infrared (red). The expanding shell is rich in iron, which was produced in the star that exploded. The main image combines the individual single-color images seen at the bottom into one multi-wavelength picture. (credit: modification of work by NASA, ESA, R. Sankrit and W. Blair (Johns Hopkins University))

Supernovae are also thought to be the source of many of the high-energy *cosmic ray* particles discussed in Cosmic Rays. Trapped by the magnetic field of the Galaxy, the particles from exploded stars continue to circulate around the vast spiral of the Milky Way. Scientists speculate that high-speed cosmic rays hitting the genetic material of Earth organisms over billions of years may have contributed to the steady *mutations*—subtle changes in the genetic code—that drive the evolution of life on our planet. In all the ways we have mentioned, supernovae have played a part in the development of new generations of stars, planets, and life.

But supernovae also have a dark side. Suppose a life form has the misfortune to develop around a star that happens to lie near a massive star destined to become a supernova. Such life forms may find themselves snuffed out when the harsh radiation and high-energy particles from the neighboring star's explosion reach their world. If, as some astronomers speculate, life can develop on many planets around long-lived (lower-mass) stars, then the suitability of that life's *own star* and planet may not be all that matters for its long-term evolution and survival. Life may well have formed around a number of pleasantly stable stars only to be wiped out because a massive nearby star suddenly went supernova. Just as children born in a war zone may find themselves the unjust victims of their violent neighborhood, life too close to a star that goes supernova may fall prey to having been born in the wrong place at the wrong time.

What is a safe distance to be from a supernova explosion? A lot depends on the violence of the particular explosion, what type of supernova it is (see The Evolution of Binary Star Systems), and what level of destruction we are willing to accept. Calculations suggest that a supernova less than 50 light-years away from us would certainly end all life on Earth, and that even one 100 light-years away would have drastic consequences for the radiation levels here. One minor extinction of sea creatures about 2 million years ago on Earth may actually have been caused by a supernova at a distance of about 120 light-years.

The good news is that there are at present no massive stars that promise to become supernovae within 50 light-years of the Sun. (This is in part because the kinds of massive stars that become supernovae are overall quite rare.) The massive star closest to us, Spica (in the constellation of Virgo), is about 260 light-years away, probably a safe distance, even if it were to explode as a supernova in the near future.

Example 23.1: Extreme Gravity

In this section, you were introduced to some very dense objects. How would those objects' gravity affect you? Recall that the force of gravity, F, between two bodies is calculated as

$$F = \frac{GM_1M_2}{R^2}$$

where G is the gravitational constant, $6.67 \times 10^{-11}\text{Nm}^2/\text{kg}^2, M_1 \text{and} M_2$ are the masses of the two bodies, and R is their separation. Also, from Newton's second law,

$$F = M \times a$$

where a is the acceleration of a body with mass M.

So let's consider the situation of a mass—say, you—standing on a body, such as Earth or a white dwarf (where we assume you will be wearing a heat-proof space suit). You are M_1 and the body you are standing on is M_2. The distance between you and the center of gravity of the body on which you stand is its radius, R. The force exerted on you is

$$F = M_1 \times a = \text{GM}_1M_2/R^2$$

Solving for a, the acceleration of gravity on that world, we get

$$g = \frac{(G \times M)}{R^2}$$

Note that we have replaced the general symbol for acceleration, a, with the symbol scientists use for the acceleration of gravity, g.

Say that a particular white dwarf has the mass of the Sun $(2 \times 10^{30}\text{kg})$ but the radius of Earth $(6.4 \times 10^6\text{m})$. What is the acceleration of gravity at the surface of the white dwarf?

Solution

The acceleration of gravity at the surface of the white dwarf is

$$g(\text{white dwarf}) = \frac{(G \times M_{\text{Sun}})}{{R_{\text{Earth}}}^2} = \frac{\left(6.67 \times 10^{-11}\,\text{m}^2/\text{kg}\,\text{s}^2 \times 2 \times 10^{30}\,\text{kg}\right)}{\left(6.4 \times 10^6\,\text{m}\right)^2} = 3.26 \times 10^6\,\text{m/s}^2$$

Compare this to g on the surface of Earth, which is 9.8 m/s^2.

Exercise 14.2.1

What is the acceleration of gravity at the surface if the white dwarf has the twice the mass of the Sun and is only half the radius of Earth?

Answer available at the end of this section.

Footnotes

- 1Stars in the mass ranges 0.25–8 and 8–10 may later produce a type of supernova different from the one we have discussed so far. These are discussed in The Evolution of Binary Star Systems.
- 2To tell the full story, there is another way to make and recycle these heavier elements which we will discuss in The Milky Way Galaxy. That is the merger of two neutron stars, which also causes a huge explosion.

Answers

Exercise 14.2.1

$$g(\text{white dwarf}) = \frac{(G \times 2M_{\text{Sun}})}{(0.5R_{\text{Earth}})^2} = \frac{\left(6.67 \times 10^{-11}\text{m}^2/\text{kg}\,\text{s}^2 \times 4 \times 10^{30}\,\text{kg}\right)}{\left(3.2 \times 10^6\right)^2} = 2.61 \times 10^7\,\text{m/s}^2$$

This page titled 14.2: Evolution of Massive Stars- An Explosive Finish is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 23.3: Evolution of Massive Stars- An Explosive Finish by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14.3: Supernova Observations

Learning Objectives

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

- Describe the observed features of SN 1987A both before and after the supernova
- Explain how observations of various parts of the SN 1987A event helped confirm theories about supernovae

Supernovae were discovered long before astronomers realized that these spectacular cataclysms mark the death of stars (see Making Connections: Supernovae in History). The word *nova* means “new” in Latin; before telescopes, when a star too dim to be seen with the unaided eye suddenly flared up in a brilliant explosion, observers concluded it must be a brand-new star. Twentieth-century astronomers reclassified the explosions with the greatest luminosity as *super*novae.

From historical records of such explosions, from studies of the remnants of supernovae in our Galaxy, and from analyses of supernovae in other galaxies, we estimate that, on average, one supernova explosion occurs somewhere in the Milky Way Galaxy every 25 to 100 years. Unfortunately, however, no supernova explosion has been observable in our Galaxy since the invention of the telescope. Either we have been exceptionally unlucky or, more likely, recent explosions have taken place in parts of the Galaxy where interstellar dust blocks light from reaching us.

Making Connections: Supernovae in History

Although many supernova explosions in our own Galaxy have gone unnoticed, a few were so spectacular that they were clearly seen and recorded by sky watchers and historians at the time. We can use these records, going back two millennia, to help us pinpoint where the exploding stars were and thus where to look for their remnants today.

The most dramatic supernova was observed in the year 1006. It appeared in May as a brilliant point of light visible during the daytime, perhaps 100 times brighter than the planet Venus. It was bright enough to cast shadows on the ground during the night and was recorded with awe and fear by observers all over Europe and Asia. No one had seen anything like it before; Chinese astronomers, noting that it was a temporary spectacle, called it a “guest star.”

Astronomers David Clark and Richard Stephenson have scoured records from around the world to find more than 20 reports of the 1006 supernova (SN 1006) (Figure 23.9). This has allowed them to determine with some accuracy where in the sky the explosion occurred. They place it in the modern constellation of Lupus; at roughly the position they have determined, we find a supernova remnant, now quite faint. From the way its filaments are expanding, it indeed appears to be about 1000 years old.



Figure 23.9 : Supernova 1006 Remnant. This composite view of SN 1006 from the Chandra X-Ray Observatory shows the X-rays coming from the remnant in blue, visible light in white-yellow, and radio emission in red. (credit: modification of work by NASA, ESA, Zolt Levay(STScI))

Another guest star, now known as SN 1054, was clearly recorded in Chinese records in July 1054. The remnant of that star is one of the most famous and best-studied objects in the sky, called the Crab Nebula (Figure 23.14). It is a marvelously complex object, which has been key to understanding the death of massive stars. When its explosion was first seen, we estimate that it was about as bright as the planet Jupiter: nowhere near as dazzling as the 1006 event but still quite dramatic to anyone who kept track of objects in the sky. Another fainter supernova was seen in 1181.

The next supernova became visible in November 1572 and, being brighter than the planet Venus, was quickly spotted by a number of observers, including the young Tycho Brahe (see Orbits and Gravity). His careful measurements of the star over a year and a half showed that it was not a comet or something in Earth's atmosphere since it did not move relative to the stars. He correctly deduced that it must be a phenomenon belonging to the realm of the stars, not of the solar system. The remnant of Tycho's Supernova (as it is now called) can still be detected in many different bands of the electromagnetic spectrum.

Not to be outdone, Johannes Kepler, Tycho Brahe's scientific heir, found his own supernova in 1604, now known as Kepler's Supernova (Figure 23.8). Fainter than Tycho's, it nevertheless remained visible for about a year. Kepler wrote a book about his observations that was read by many with an interest in the heavens, including Galileo.

No supernova has been spotted in our Galaxy for the past 300 years. Since the explosion of a visible supernova is a chance event, there is no way to say when the next one might occur. Around the world, dozens of professional and amateur astronomers keep a sharp lookout for "new" stars that appear overnight, hoping to be the first to spot the next guest star in our sky and make a little history themselves.

At their maximum brightness, the most luminous supernovae have about 10 billion times the luminosity of the Sun. For a brief time, a supernova may outshine the entire galaxy in which it appears. After maximum brightness, the star's light fades and disappears from telescopic visibility within a few months or years. At the time of their outbursts, supernovae eject material at typical velocities of 10,000 kilometers per second (and speeds twice that have been observed). A speed of 20,000 kilometers per second corresponds to about 45 million miles per hour, truly an indication of great cosmic violence.

Supernovae are classified according to the appearance of their spectra, but in this chapter, we will focus on the two main causes of supernovae. Type Ia supernovae are ignited when a lot of material is dumped on degenerate white dwarfs (Figure 23.10); these supernovae will be discussed later in this chapter. For now, we will continue our story about the death of massive stars and focus on type II supernovae, which are produced when the core of a massive star collapses.

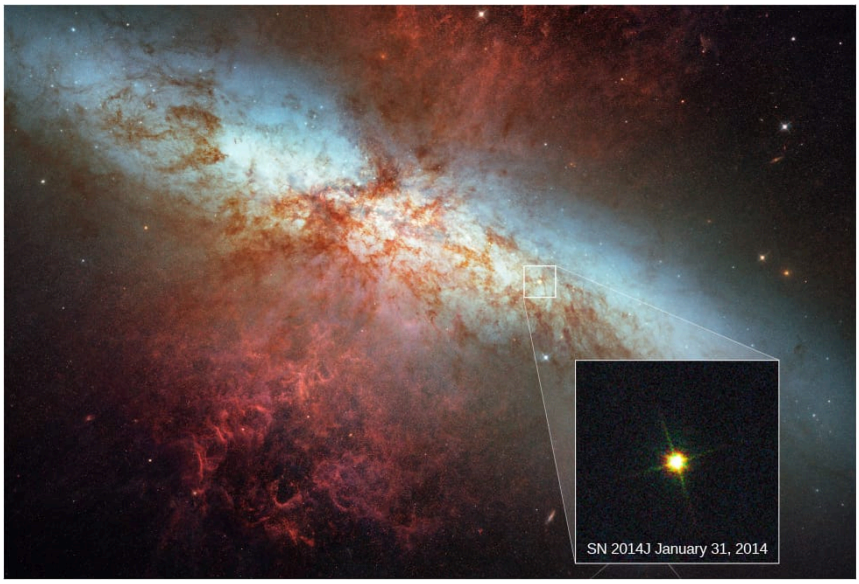


Figure 23.10 : Supernova 2014J. This image of supernova 2014J, located in Messier 82 (M82), which is also known as the Cigar galaxy, was taken by the Hubble Space Telescope and is superposed on a mosaic image of the galaxy also taken with Hubble. The supernova event is indicated by the box and the inset. This explosion was produced by a type Ia supernova, which is theorized to be triggered in binary systems consisting of a white dwarf and another star—and could be a second white dwarf, a star like our Sun, or a giant star. This type of supernova will be discussed later in this chapter. At a distance of approximately 11.5 million light-years from Earth, this is the closest supernova of type Ia discovered in the past few decades. In the image, you can see reddish plumes of hydrogen coming from the central region of the galaxy, where a considerable number of young stars are being born. (credit: modification of work by NASA, ESA, A. Goobar (Stockholm University), and the Hubble Heritage Team (STScI/AURA))

Supernova 1987A

Our most detailed information about what happens when a type II supernova occurs comes from an event that was observed in 1987. Before dawn on February 24, Ian Shelton, a Canadian astronomer working at an observatory in Chile, pulled a photographic plate from the developer. Two nights earlier, he had begun a survey of the Large Magellanic Cloud, a small galaxy that is one of the Milky Way's nearest neighbors in space. Where he expected to see only faint stars, he saw a large bright spot. Concerned that his photograph was flawed, Shelton went outside to look at the Large Magellanic Cloud . . . and saw that a new object had indeed appeared in the sky (see Figure 23.11). He soon realized that he had discovered a supernova, one that could be seen with the unaided eye even though it was about 160,000 light-years away.



Figure 23.11 : Hubble Space Telescope Image of SN 1987A. The supernova remnant with its inner and outer red rings of material is located in the Large Magellanic Cloud. This image is a composite of several images taken in 1994, 1996, and 1997—about a decade after supernova 1987A was first observed. (credit: modification of work by the Hubble Heritage Team (AURA/STScI/NASA/ESA))

Now known as SN 1987A, since it was the first supernova discovered in 1987, this brilliant newcomer to the southern sky gave astronomers their first opportunity to study the death of a relatively nearby star with modern instruments. It was also the first time astronomers had observed a star *before* it became a supernova. The star that blew up had been included in earlier surveys of the Large Magellanic Cloud, and as a result, we know the star was a blue supergiant just before the explosion.

By combining theory and observations at many different wavelengths, astronomers have reconstructed the life story of the star that became SN 1987A. Formed about 10 million years ago, it originally had a mass of about 20 M_{Sun}. For 90% of its life, it lived quietly on the main sequence, converting hydrogen into helium. At this time, its luminosity was about 60,000 times that of the Sun (L_{Sun}), and its spectral type was O. When the hydrogen in the center of the star was exhausted, the core contracted and ultimately became hot enough to fuse helium. By this time, the star was a red supergiant, emitting about 100,000 times more energy than the Sun. While in this stage, the star lost some of its mass.

This lost material has actually been detected by observations with the Hubble Space Telescope (Figure 23.12). The gas driven out into space by the subsequent supernova explosion is currently colliding with the material the star left behind when it was a red giant. As the two collide, we see a glowing ring.

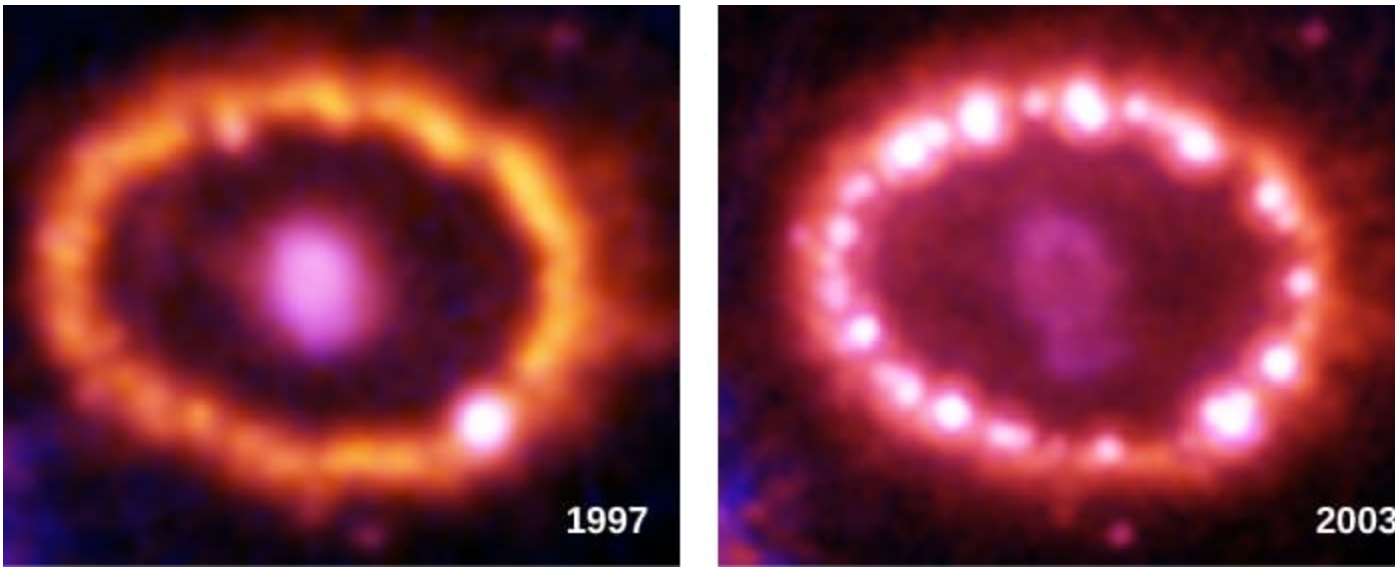


Figure 23.12 : Ring around Supernova 1987A. These two images show a ring of gas expelled by a red giant star about 30,000 years before the star exploded and was observed as Supernova 1987A. The supernova, which has been artificially dimmed, is located at the center of the ring. The left-hand image was taken in 1997 and the right-hand image in 2003. Note that the number of bright spots has increased from 1 to more than 15 over this time interval. These spots occur where high-speed gas ejected by the supernova and moving at millions of miles per hour has reached the ring and blasted into it. The collision has heated the gas in the ring and caused it to glow more brightly. The fact that we see individual spots suggests that material ejected by the supernova is first hitting narrow, inward-projecting columns of gas in the clumpy ring. The hot spots are the first signs of a dramatic and violent collision between the new and old material that will continue over the next few years. By studying these bright spots, astronomers can determine the composition of the ring and hence learn about the nuclear processes that build heavy elements inside massive stars. (credit: modification of work by NASA, P. Challis, R. Kirshner (Harvard-Smithsonian Center for Astrophysics) and B. Sugerman (STScI))

Helium fusion lasted only about 1 million years. When the helium was exhausted at the center of the star, the core contracted again, the radius of the surface also decreased, and the star became a blue supergiant with a luminosity still about equal to 100,000 L_{Sun}. This is what it still looked like on the outside when, after brief periods of further fusion, it reached the iron crisis we discussed earlier and exploded.

Some key stages of evolution of the star that became SN 1987A, including the ones following helium exhaustion, are listed in Table 23.2. While we don't expect you to remember these numbers, note the patterns in the table: each stage of evolution happens more quickly than the preceding one, the temperature and pressure in the core increase, and progressively heavier elements are the source of fusion energy. Once iron was created, the collapse began. It was a catastrophic collapse, lasting only a few tenths of a second; the speed of infall in the outer portion of the iron core reached 70,000 kilometers per second, about one-fourth the speed of light.

Table 23.2: Evolution of the Star That Exploded as SN 1987A

Phase	Central Temperature (K)	Central Density (g/cm^3)	Time Spent in This Phase
Hydrogen fusion	40×10^6	5	8×10^6 years
Helium fusion	190×10^6	970	10^6 years
Carbon fusion	870×10^6	170,000	2000 years
Neon fusion	1.6×10^9	3.0×10^6	6 months
Oxygen fusion	2.0×10^9	5.6×10^6	1 year
Silicon fusion	3.3×10^9	4.3×10^7	Days
Core collapse	200×10^9	2×10^{14}	Tenths of a second

In the meantime, as the core was experiencing its last catastrophe, the outer shells of neon, oxygen, carbon, helium, and hydrogen in the star did not yet know about the collapse. Information about the physical movement of different layers travels through a star at the speed of sound and cannot reach the surface in the few tenths of a second required for the core collapse to occur. Thus,

the surface layers of our star hung briefly suspended, much like a cartoon character who dashes off the edge of a cliff and hangs momentarily in space before realizing that he is no longer held up by anything.

The collapse of the core continued until the densities rose to several times that of an atomic nucleus. The resistance to further collapse then became so great that the core rebounded. Infalling material ran into the “brick wall” of the rebounding core and was thrown outward with a great shock wave. Neutrinos poured out of the core, helping the shock wave blow the star apart. The shock reached the surface of the star a few hours later, and the star began to brighten into the supernova Ian Shelton observed in 1987.

The Synthesis of Heavy Elements

The variations in the brightness of SN 1987A in the days and months after its discovery, which are shown in Figure 23.13, helped confirm our ideas about heavy element production. In a single day, the star soared in brightness by a factor of about 1000 and became just visible without a telescope. The star then continued to increase slowly in brightness until it was about the same apparent brightness as the stars in the Little Dipper. Up until about day 40 after the outburst, the energy being radiated away was produced by the explosion itself. But then SN 1987A did not continue to fade away, as we might have expected the light from the explosion to do. Instead, SN 1987A remained bright as energy from newly created radioactive elements came into play.

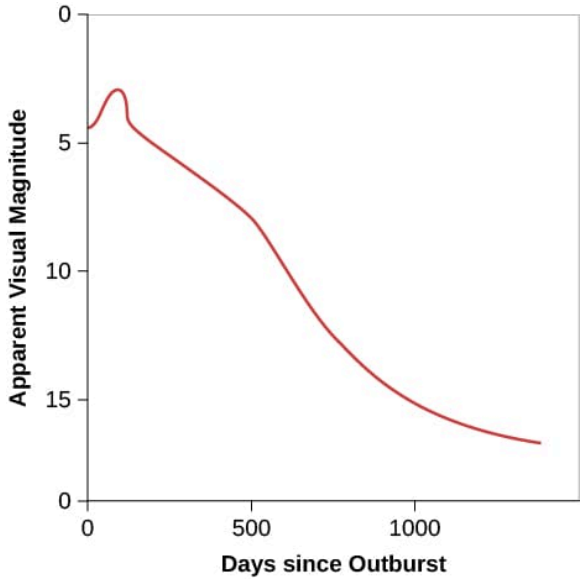


Figure 23.13 : Change in the Brightness of SN 1987A over Time. Note how the rate of decline of the supernova’s light slowed between days 40 and 500. During this time, the brightness was mainly due to the energy emitted by newly formed (and quickly decaying) radioactive elements. Remember that magnitudes are a backward measure of brightness: the larger the magnitude, the dimmer the object looks.

One of the elements formed in a supernova explosion is radioactive nickel, with an atomic mass of 56 (that is, the total number of protons plus neutrons in its nucleus is 56). Nickel-56 is unstable and changes spontaneously (with a half-life of about 6 days) to cobalt-56. (Recall that a half-life is the time it takes for half the nuclei in a sample to undergo radioactive decay.) Cobalt-56 in turn decays with a half-life of about 77 days to iron-56, which is stable. Energetic gamma rays are emitted when these radioactive nuclei decay. Those gamma rays then serve as a new source of energy for the expanding layers of the supernova. The gamma rays are absorbed in the overlying gas and re-emitted at visible wavelengths, keeping the remains of the star bright.

As you can see in Figure 23.13, astronomers did observe brightening due to radioactive nuclei in the first few months following the supernova’s outburst and then saw the extra light die away as more and more of the radioactive nuclei decayed to stable iron. The gamma-ray heating was responsible for virtually all of the radiation detected from SN 1987A after day 40. Some gamma rays also escaped directly without being absorbed. These were detected by Earth-orbiting telescopes at the wavelengths expected for the decay of radioactive nickel and cobalt, clearly confirming our understanding that new elements were indeed formed in the crucible of the supernova.

Neutrinos from SN 1987A

If there had been any human observers in the Large Magellanic Cloud about 160,000 years ago, the explosion we call SN 1987A would have been a brilliant spectacle in their skies. Yet we know that less than 1/10 of 1% of the energy of the explosion appeared as visible light. About 1% of the energy was required to destroy the star, and the rest was carried away by neutrinos. The overall energy in these neutrinos was truly astounding. In the initial second of the event, as we noted earlier in our general discussion of supernovae, their total luminosity exceeded the luminosity of all the stars in over a billion galaxies. And the supernova generated this energy in a volume less than 50 kilometers in diameter! Supernovae are one of the most violent events in the universe, and their *light* turns out to be only the tip of the iceberg in revealing how much energy they produce.

In 1987, the neutrinos from SN 1987A were detected by two instruments—which might be called “neutrino telescopes”—almost a full day before Shelton’s observations. (This is because the neutrinos get out of the exploding star more easily than light does, and also because you don’t need to wait until nightfall to catch a “glimpse” of them.) Both neutrino telescopes, one in a deep mine in Japan and the other under Lake Erie, consist of several thousand tons of purified water surrounded by several hundred light-sensitive detectors. Incoming neutrinos interact with the water to produce positrons and electrons, which move rapidly through the water and emit deep blue light.

Altogether, 19 neutrinos were detected. Since the neutrino telescopes were in the Northern Hemisphere and the supernova occurred in the Southern Hemisphere, the detected neutrinos had already passed through Earth and were on their way back out into space when they were captured.

Only a few neutrinos were detected because the probability that they will interact with ordinary matter is very, very low. It is estimated that the supernova actually released 10^{58} neutrinos. A tiny fraction of these, about 30 billion, eventually passed through each square centimeter of Earth’s surface. About a million people actually experienced a neutrino interaction within their bodies as a result of the supernova. This interaction happened to only a single nucleus in each person and thus had absolutely no biological effect; it went completely unnoticed by everyone concerned.

Since the neutrinos come directly from the heart of the supernova, their energies provided a measure of the temperature of the core as the star was exploding. The central temperature was about 200 billion K, a stunning figure to which no earthly analog can bring much meaning. With neutrino telescopes, we are peering into the final moment in the life stories of massive stars and observing conditions beyond all human experience. Yet we are also seeing the unmistakable hints of our own origins.

This page titled 14.3: Supernova Observations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 23.4: Supernova Observations by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14.4: Pulsars and the Discovery of Neutron Stars

Learning Objectives

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

- Explain the research method that led to the discovery of neutron stars, located hundreds or thousands of light-years away
- Describe the features of a neutron star that allow it to be detected as a pulsar
- List the observational evidence that links pulsars and neutron stars to supernovae

After a type II supernova explosion fades away, all that is left behind is either a neutron star or something even more strange, a black hole. We will describe the properties of black holes in Black Holes and Curved Spacetime, but for now, we want to examine how the neutron stars we discussed earlier might become observable.

Neutron stars are the densest objects in the universe; the force of gravity at their surface is 10^{11} times greater than what we experience at Earth's surface. The interior of a neutron star is composed of about 95% neutrons, with a small number of protons and electrons mixed in. In effect, a neutron star is a giant atomic nucleus, with a mass about 10^{57} times the mass of a proton. Its diameter is more like the size of a small town or an asteroid than a star. (Table 23.3 compares the properties of neutron stars and white dwarfs.) Because it is so small, a neutron star probably strikes you as the object least likely to be observed from thousands of light-years away. Yet neutron stars do manage to signal their presence across vast gulfs of space.

Table 23.3: Properties of a Typical White Dwarf and a Neutron Star

Property	White Dwarf	Neutron Star
Mass (Sun = 1)	0.6 (always <1.4)	Always >1.4 and <3
Radius	7000 km	10 km
Density	8×10^5 g/cm^3	10^{14} g/cm^3

The Discovery of Neutron Stars

In 1967, Jocelyn Bell, a research student at Cambridge University, was studying distant radio sources with a special detector that had been designed and built by her advisor Antony Hewish to find rapid variations in radio signals. The project computers spewed out reams of paper showing where the telescope had surveyed the sky, and it was the job of Hewish's graduate students to go through it all, searching for interesting phenomena. In September 1967, Bell discovered what she called "a bit of scruff"—a strange radio signal unlike anything seen before.

What Bell had found, in the constellation of Vulpecula, was a source of rapid, sharp, intense, and extremely regular pulses of radio radiation. Like the regular ticking of a clock, the pulses arrived precisely every 1.33728 seconds. Such exactness first led the scientists to speculate that perhaps they had found signals from an intelligent civilization. Radio astronomers even half-jokingly dubbed the source "LGM" for "little green men." Soon, however, three similar sources were discovered in widely separated directions in the sky.

When it became apparent that this type of radio source was fairly common, astronomers concluded that they were highly unlikely to be signals from other civilizations. By today, nearly 3000 such sources have been discovered; they are now called pulsars, short for "pulsating radio sources."

The pulse periods of different pulsars range from a little longer than 1/1000 of a second to nearly 10 seconds. At first, the pulsars seemed particularly mysterious because nothing could be seen at their location on visible-light photographs. But then a pulsar was discovered right in the center of the Crab Nebula, a cloud of gas produced by SN 1054, a supernova that was recorded by the Chinese in 1054 (Figure 23.14). The energy from the Crab Nebula pulsar arrives in sharp bursts that occur 30 times each second—with a regularity that would be the envy of a Swiss watchmaker. In addition to pulses of radio energy, we can observe

pulses of visible light and X-rays from the Crab Nebula. The fact that the pulsar was just in the region of the supernova remnant where we expect the leftover neutron star to be immediately alerted astronomers that pulsars might be connected with these elusive “corpses” of massive stars.



Figure 23.14 : Crab Nebula. This image shows X-ray emissions from the Crab Nebula, which is about 6500 light-years away. The pulsar is the bright spot at the center of the concentric rings. Data taken over about a year show that particles stream away from the inner ring at about half the speed of light. The jet that is perpendicular to this ring is a stream of matter and antimatter electrons also moving at half the speed of light. (credit: modification of work by NASA/CXC/SAO)

The Crab Nebula is a fascinating object. The whole nebula glows with radiation at many wavelengths, and its overall energy output is more than 100,000 times that of the Sun—not a bad trick for the remnant of a supernova that exploded almost a thousand years ago. Astronomers soon began to look for a connection between the pulsar and the large energy output of the surrounding nebula.

Link to Learning

Read the transcript of an interesting interview with Jocelyn Bell (Burnell) to learn about her life and work (this is part of a project at the American Institute of Physics to record interviews with pathbreaking scientists while they are still alive).

A Spinning Lighthouse Model

By applying a combination of theory and observation, astronomers eventually concluded that pulsars must be *spinning neutron stars*. According to this model, a neutron star is something like a lighthouse on a rocky coast (Figure 23.15). To warn ships in all directions and yet not cost too much to operate, the light in a modern lighthouse turns, sweeping its beam across the dark sea. From the vantage point of a ship, you see a pulse of light each time the beam points in your direction. In the same way, radiation from a small region on a neutron star sweeps across the oceans of space, giving us a pulse of radiation each time the beam points toward Earth.



Figure 23.15 : Lighthouse. A lighthouse in California warns ships on the ocean not to approach too close to the dangerous shoreline. The lighted section at the top rotates so that its beam can cover all directions. (credit: Anita Ritenour)

Neutron stars are ideal candidates for such a job because the collapse has made them so small that they can turn very rapidly. Recall the principle of the *conservation of angular momentum* from Newton's Great Synthesis: if an object gets smaller, it can spin more rapidly. Even if the parent star was rotating very slowly when it was on the main sequence, its rotation had to speed up as it collapsed to form a neutron star. With a diameter of only 10 to 20 kilometers, a neutron star can complete one full spin in only a fraction of a second. This is just the sort of time period we observe between pulsar pulses.

Any magnetic field that existed in the original star will be highly compressed when the core collapses to a neutron star. At the surface of the neutron star, in the outer layer consisting of ordinary matter (and not just pure neutrons), protons and electrons are caught up in this spinning field and accelerated nearly to the speed of light. In only two places—the north and south magnetic poles—can the trapped particles escape the strong hold of the magnetic field (Figure 23.16). The same effect can be seen (in reverse) on Earth, where charged particles from space are *kept out* by our planet's magnetic field everywhere except near the poles. As a result, Earth's auroras (caused when charged particles hit the atmosphere at high speed) are seen mainly near the poles.

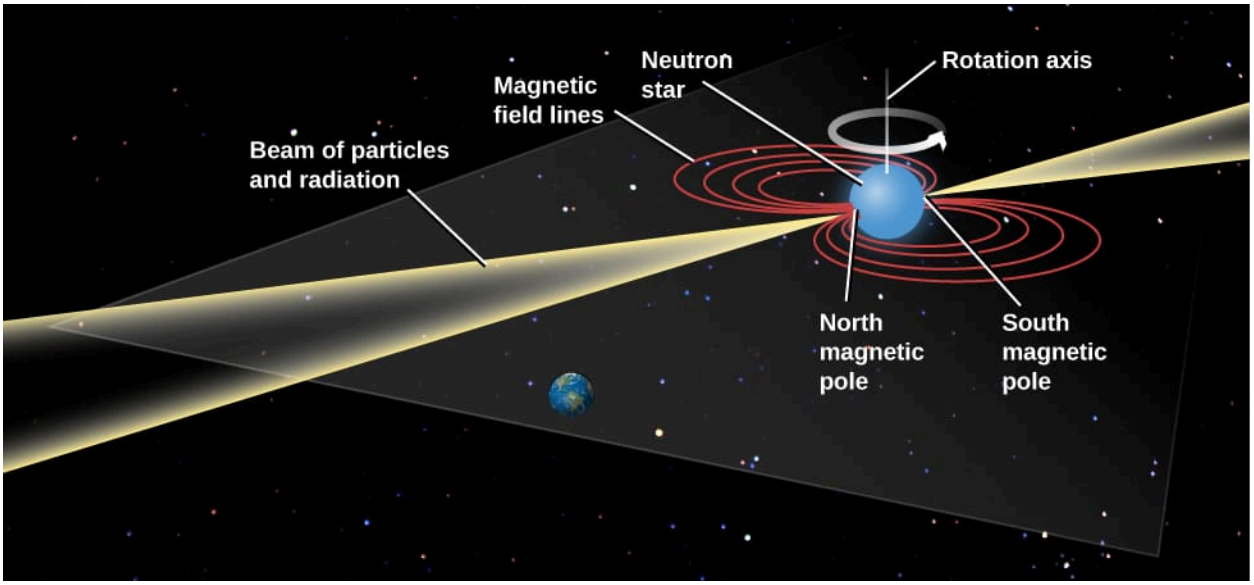


Figure 23.16 : Model of a Pulsar. A diagram showing how beams of radiation at the magnetic poles of a neutron star can give rise to pulses of emission as the star rotates. As each beam sweeps over Earth, like a lighthouse beam sweeping over a distant ship, we see a short pulse of radiation. This model requires that the magnetic poles be located in different places from the rotation poles. (credit "stars": modification of work by Tony Hisgett)

Note that in a neutron star, the magnetic north and south poles do not have to be anywhere close to the north and south poles defined by the star's rotation. In the same way, we discussed in the chapter on The Giant Planets that the magnetic poles on the planets Uranus and Neptune are not lined up with the poles of the planet's spin. Figure 23.16 shows the poles of the magnetic field perpendicular to the poles of rotation, but the two kinds of poles could make any angle.

In fact, the misalignment of the rotational axis with the magnetic axis plays a crucial role in the generation of the observed pulses in this model. At the two magnetic poles, the particles from the neutron star are focused into a narrow beam and come streaming out of the whirling magnetic region at enormous speeds. They emit energy over a broad range of the electromagnetic spectrum. The radiation itself is also confined to a narrow beam, which explains why the pulsar acts like a lighthouse. As the rotation carries first one and then the other magnetic pole of the star into our view, we see a pulse of radiation each time.

Tests of the Model

This explanation of pulsars in terms of beams of radiation from highly magnetic and rapidly spinning neutron stars is a very clever idea. But what evidence do we have that it is the correct model? First, we can measure the masses of some pulsars, and they do turn out be in the range of 1.4 to 1.8 times that of the Sun—just what theorists predict for neutron stars. The masses are found using Kepler's law for those few pulsars that are members of binary star systems.

But there is an even-better confirming argument, which brings us back to the Crab Nebula and its vast energy output. When the high-energy charged particles from the neutron star pulsar hit the slower-moving material from the supernova, they energize this material and cause it to "glow" at many different wavelengths—just what we observe from the Crab Nebula. The pulsar beams are a power source that "light up" the nebula long after the initial explosion of the star that made it.

Who "pays the bills" for all the energy we see coming out of a remnant like the Crab Nebula? After all, when energy emerges from one place, it must be depleted in another. The ultimate energy source in our model is the rotation of the neutron star, which propels charged particles outward and spins its magnetic field at enormous speeds. As its rotational energy is used to excite the Crab Nebula year after year, the pulsar inside the nebula slows down. As it slows, the pulses come a little less often; more time elapses before the slower neutron star brings its beam back around.

Several decades of careful observations have now shown that the Crab Nebula pulsar is not a perfectly regular clock as we originally thought: instead, it is gradually slowing down. Having measured how much the pulsar is slowing down, we can calculate how much rotation energy the neutron star is losing. Remember that it is very densely packed and spins amazingly quickly. Even a tiny slowing down can mean an immense loss of energy.

To the satisfaction of astronomers, the rotational energy lost by the pulsar turns out to be the same as the amount of energy emerging from the nebula surrounding it. In other words, the slowing down of a rotating neutron star can explain precisely why the Crab Nebula is glowing with the amount of energy we observe.

The Evolution of Pulsars

From observations of the pulsars discovered so far, astronomers have concluded that one new pulsar is born somewhere in the Galaxy every 25 to 100 years, the same rate at which supernovae are estimated to occur. Calculations suggest that the typical lifetime of a pulsar is about 10 million years; after that, the neutron star no longer rotates fast enough to produce significant beams of particles and energy, and is no longer observable. We estimate that there are about 100 million neutron stars in our Galaxy, most of them rotating too slowly to come to our notice.

The Crab pulsar is rather young (only about 960 years old) and has a short period, whereas other, older pulsars have already slowed to longer periods. Pulsars thousands of years old have lost too much energy to emit appreciably in the visible and X-ray wavelengths, and they are observed only as radio pulsars; their periods are a second or longer.

There is one other reason we can see only a fraction of the pulsars in the Galaxy. Consider our lighthouse model again. On Earth, all ships approach on the same plane—the surface of the ocean—so the lighthouse can be built to sweep its beam over that surface. But in space, objects can be anywhere in three dimensions. As a given pulsar's beam sweeps over a circle in space, there is absolutely no guarantee that this circle will include the direction of Earth. In fact, if you think about it, many more circles in space will *not* include Earth than will include it. Thus, we estimate that we are unable to observe a large number of neutron stars because their pulsar beams miss us entirely.

At the same time, it turns out that only a few of the pulsars discovered so far are embedded in the visible clouds of gas that mark the remnant of a supernova. This might at first seem mysterious, since we know that supernovae give rise to neutron stars and we should expect each pulsar to have begun its life in a supernova explosion. But the lifetime of a pulsar turns out to be about 100 times longer than the length of time required for the expanding gas of a supernova remnant to disperse into interstellar space. Thus, most pulsars are found with no other trace left of the explosion that produced them.

In addition, some pulsars are ejected by a supernova explosion that is not the same in all directions. If the supernova explosion is stronger on one side, it can kick the pulsar entirely out of the supernova remnant (some astronomers call this “getting a birth kick”). We know such kicks happen because we see a number of young supernova remnants in nearby galaxies where the pulsar is to one side of the remnant and racing away at several hundred miles per second (Figure 23.17).



Figure 23.17 : Speeding Pulsar. This intriguing image (which combines X-ray, visible, and radio observations) shows the jet trailing behind a pulsar (at bottom right, lined up between the two bright stars). With a length of 37 light-years, the jet trail (seen in purple) is the longest ever observed from an object in the Milky Way. (There is also a mysterious shorter, comet-like tail that is almost perpendicular to the purple jet.) Moving at a speed between 2.5 and 5 million miles per hour, the pulsar is traveling away from the core of the supernova remnant where it originated. (credit: X-ray: NASA/CXC/ISDC/L.Pavan et al, Radio: CSIRO/ATNF/ATCA Optical: 2MASS/UMass/IPAC-Caltech/NASA/NSF)

Making Connections: Touched by a Neutron Star

On December 27, 2004, Earth was bathed with a stream of X-ray and gamma-ray radiation from a neutron star known as SGR 1806-20. What made this event so remarkable was that, despite the distance of the source, its tidal wave of radiation had measurable effects on Earth’s atmosphere. The apparent brightness of this gamma-ray flare was greater than any historical star explosion.

The primary effect of the radiation was on a layer high in Earth’s atmosphere called the *ionosphere*. At night, the ionosphere is normally at a height of about 85 kilometers, but during the day, energy from the Sun ionizes more molecules and lowers the boundary of the ionosphere to a height of about 60 kilometers. The pulse of X-ray and gamma-ray radiation produced about the same level of ionization as the daytime Sun. It also caused some sensitive satellites above the atmosphere to shut down their electronics.

Measurements by telescopes in space indicate that SGR 1806-20 was a special type of fast-spinning neutron star called a *magnetar*. Astronomers Robert Duncan and Christopher Thomson gave them this name because their magnetic fields are stronger than that of any other type of astronomical source—in this case, about 800 trillion times stronger than the magnetic field of Earth.

A magnetar is thought to consist of a superdense core of neutrons surrounded by a rigid crust of atoms about a mile deep with a surface made of iron. The magnetar’s field is so strong that it creates huge stresses inside that can sometimes crack open the hard crust, causing a starquarke. The vibrating crust produces an enormous blast of radiation. An astronaut 0.1 light-year from this particular magnetar would have received a fatal does from the blast in less than a second.

Fortunately, we were far enough away from magnetar SGR 1806-20 to be safe. Could a magnetar ever present a real danger to Earth? To produce enough energy to disrupt the ozone layer, a magnetar would have to be located within the cloud of comets that surround the solar system, and we know no magnetars are that close. Nevertheless, it is a fascinating discovery that events on distant star corpses can have measurable effects on Earth.

This page titled 14.4: Pulsars and the Discovery of Neutron Stars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 23.5: Pulsars and the Discovery of Neutron Stars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14.5: The Evolution of Binary Star Systems

Learning Objectives

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

- Describe the kind of binary star system that leads to a nova event
- Describe the type of binary star system that leads to a type Ia supernovae event
- Indicate how type Ia supernovae differ from type II supernovae

The discussion of the life stories of stars presented so far has suffered from a bias—what we might call "single-star chauvinism." Because the human race developed around a star that goes through life alone, we tend to think of most stars in isolation. But as we saw in The Stars: A Celestial Census, it now appears that as many as half of all stars may develop in *binary* systems—those in which two stars are born in each other's gravitational embrace and go through life orbiting a common center of mass.

For these stars, the presence of a close-by companion can have a profound influence on their evolution. Under the right circumstances, stars can exchange material, especially during the stages when one of them swells up into a giant or supergiant, or has a strong wind. When this happens and the companion stars are sufficiently close, material can flow from one star to another, decreasing the mass of the donor and increasing the mass of the recipient. Such *mass transfer* can be especially dramatic when the recipient is a stellar remnant such as a white dwarf or a neutron star. While the detailed story of how such binary stars evolve is beyond the scope of our book, we do want to mention a few examples of how the stages of evolution described in this chapter may change when there are two stars in a system.

White Dwarf Explosions: The Mild Kind

Let's consider the following system of two stars: one has become a white dwarf and the other is gradually transferring material onto it. As fresh hydrogen from the outer layers of its companion accumulates on the surface of the hot white dwarf, it begins to build up a layer of hydrogen. As more and more hydrogen accumulates and heats up on the surface of the degenerate star, the new layer eventually reaches a temperature that causes fusion to begin in a sudden, explosive way, blasting much of the new material away.

In this way, the white dwarf quickly (but only briefly) becomes quite bright, hundreds or thousands of times its previous luminosity. To observers before the invention of the telescope, it seemed that a new star suddenly appeared, and they called it a nova.[3] Novae fade away in a few months to a few years.

Hundreds of novae have been observed, each occurring in a binary star system and each later showing a shell of expelled material. A number of stars have more than one nova episode, as more material from its neighboring star accumulates on the white dwarf and the whole process repeats. As long as the episodes do not increase the mass of the white dwarf beyond the Chandrasekhar limit (by transferring too much mass too quickly), the dense white dwarf itself remains pretty much unaffected by the explosions on its surface.

White Dwarf Explosions: The Violent Kind

If a white dwarf accumulates matter from a companion star at a much faster rate, it can be pushed over the Chandrasekhar limit. The evolution of such a binary system is shown in Figure 23.18. When its mass approaches the Chandrasekhar mass limit (exceeds 1.4 M_{Sun}), such an object can no longer support itself as a white dwarf, and it begins to contract. As it does so, it heats up, and new nuclear reactions can begin in the degenerate core. The star "simmers" for the next century or so, building up internal temperature. This simmering phase ends in less than a second, when an enormous amount of fusion (especially of carbon) takes place all at once, resulting in an explosion. The fusion energy produced during the final explosion is so great that it completely destroys the white dwarf. Gases are blown out into space at velocities of about 10,000 kilometers per second, and afterward, no trace of the white dwarf remains.

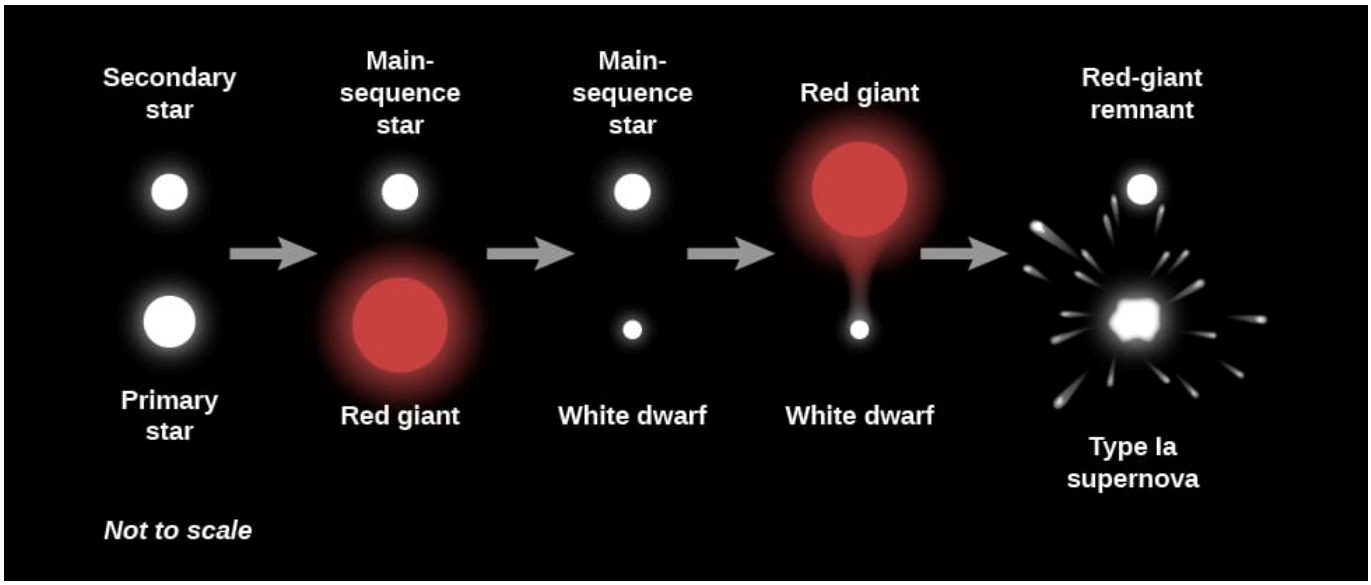


Figure 23.18 : Evolution of a Binary System. The more massive star evolves first to become a red giant and then a white dwarf. The white dwarf then begins to attract material from its companion, which in turn evolves to become a red giant. Eventually, the white dwarf acquires so much mass that it is pushed over the Chandrasekhar limit and becomes a type Ia supernova.

Such an explosion is also called a supernova, since, like the destruction of a high-mass star, it produces a huge amount of energy in a very short time. However, unlike the explosion of a high-mass star, which can leave behind a neutron star or black hole remnant, the white dwarf is completely destroyed in the process, leaving behind no remnant. We call these white dwarf explosions type Ia supernovae.

We distinguish type I supernovae from those of supernovae of type II originating from the death of massive stars discussed earlier by the absence of hydrogen in their observed spectra. Hydrogen is the most common element in the universe and is a major component of massive, evolved stars. However, as we learned earlier, hydrogen is absent from the white dwarf remnant, which is primarily composed of carbon and oxygen for masses comparable to the Chandrasekhar mass limit.

The “a” subdesignation of type Ia supernovae further refers to the presence of strong silicon absorption lines, which are absent from supernovae originating from the collapse of massive stars. Silicon is one of the products that results from the fusion of carbon and oxygen, which bears out the scenario we described above—that there is a sudden onset of the fusion of the carbon (and oxygen) of which the white dwarf was made.

Observational evidence now strongly indicates that SN 1006, Tycho’s Supernova, and Kepler’s Supernova (see Supernovae in History) were all type Ia supernovae. For instance, in contrast to the case of SN 1054, which yielded the spinning pulsar in the Crab Nebula, none of these historical supernovae shows any evidence of stellar remnants that have survived their explosions. Perhaps even more puzzling is that, so far, astronomers have not been able to identify the companion star feeding the white dwarf in any of these historical supernovae.

Consequently, in order to address the mystery of the absent companion stars and other outstanding puzzles, astronomers have recently begun to investigate alternative mechanisms of generating type Ia supernovae. All proposed mechanisms rely upon white dwarfs composed of carbon and oxygen, which are needed to meet the observed absence of hydrogen in the type Ia spectrum. And because any isolated white dwarf below the Chandrasekhar mass is stable, all proposed mechanisms invoke a binary companion to explode the white dwarf. The leading alternative mechanism scientists believe creates a type Ia supernova is the merger of two white dwarf stars in a binary system. The two white dwarfs may have unstable orbits, such that over time, they would slowly move closer together until they merge. If their combined mass is greater than the Chandrasekhar limit, the result could also be a type Ia supernova explosion.

Link to Learning

You can watch a short video about Supernova SN 2014J, a type Ia supernova discovered in the Messier 82 (M82) galaxy on January 21, 2014, as well as see brief animations of the two mechanisms by which such a supernova could form.

Type Ia supernovae are of great interest to astronomers in other areas of research. This type of supernova is brighter than supernovae produced by the collapse of a massive star. Thus, type Ia supernovae can be seen at very large distances, and they are found in all types of galaxies. The energy output from most type Ia supernovae is consistent, with little variation in their maximum luminosities, or in how their light output initially increases and then slowly decreases over time. These properties make type Ia supernovae extremely valuable "standard bulbs" for astronomers looking out at great distances—well beyond the limits of our own Galaxy. You'll learn more about their use in measuring distances to other galaxies in The Big Bang.

In contrast, type II supernovae are about 5 times less luminous than type Ia supernovae and are only seen in galaxies that have recent, massive star formation. Type II supernovae are also less consistent in their energy output during the explosion and can have a range a peak luminosity values.

Neutron Stars with Companions

Now let's look at an even-more mismatched pair of stars in action. It is possible that, under the right circumstances, a binary system can even survive the explosion of one of its members as a type II supernova. In that case, an ordinary star can eventually share a system with a neutron star. If material is then transferred from the "living" star to its "dead" (and highly compressed) companion, this material will be pulled in by the strong gravity of the neutron star. Such infalling gas will be compressed and heated to incredible temperatures. It will quickly become so hot that it will experience an explosive burst of fusion. The energies involved are so great that we would expect much of the radiation from the burst to emerge as X-rays. And indeed, high-energy observatories above Earth's atmosphere (see Astronomical Instruments) have recorded many objects that undergo just these types of X-ray *bursts*.

If the neutron star and its companion are positioned the right way, a significant amount of material can be transferred to the neutron star and can set it spinning faster (as spin energy is also transferred). The radius of the neutron star would also decrease as more mass was added. Astronomers have found pulsars in binary systems that are spinning at a rate of more than 500 times per second! (These are sometimes called millisecond pulsars since the pulses are separated by a few thousandths of a second.)

Such a rapid spin could not have come from the birth of the neutron star; it must have been externally caused. (Recall that the Crab Nebula pulsar, one of the youngest pulsars known, was spinning "only" 30 times per second.) Indeed, some of the fast pulsars are observed to be part of binary systems, while others may be alone only because they have "fully consumed" their former partner stars through the mass transfer process. (These have sometimes been called "black widow pulsars.")

Link to Learning

View this short video to see Dr. Scott Ransom, of the National Radio Astronomy Observatory, explain how millisecond pulsars come about, with some nice animations.

And if you thought that a neutron star interacting with a "normal" star was unusual, there are also binary systems that consist of two neutron stars. One such system has the stars in very close orbits to one another, so much that they continually alter each other's orbit. Another binary neutron star system includes two pulsars that are orbiting each other every 2 hours and 25 minutes. As we discussed earlier, pulsars radiate away their energy, and these two pulsars are slowly moving toward one another, such that in about 85 million years, they will actually merge (see Gravitational Wave Astronomy for our first observations of such a merger).

We have now reached the end of our description of the final stages of stars, yet one piece of the story remains to be filled in. We saw that stars whose core masses are less than 1.4 M_{Sun} at the time they run out of fuel end their lives as white dwarfs. Dying stars with core masses between 1.4 and about 3 M_{Sun} become neutron stars. But there are stars whose core masses are greater than 3 M_{Sun} when they exhaust their fuel supplies. What becomes of them? The truly bizarre result of the death of such massive stellar cores (called a black hole) is the subject of our next chapter. But first, we will look at an astronomical mystery that turned out to be related to the deaths of stars and was solved through clever sleuthing and a combination of observation and theory.

Footnotes

- 3We now know that this historical terminology is quite misleading since novae do not originate from new stars. In fact, quite to the contrary, novae originate from white dwarfs, which are actually the endpoint of stellar evolution for low-mass stars. But since the system of two stars was too faint to be visible to the naked eye, it did seem to people, before telescopes were

invented, that a star had appeared where nothing had been visible.

This page titled 14.5: The Evolution of Binary Star Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 23.6: The Evolution of Binary Star Systems by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14.6: Black Holes

Learning Objectives

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

- Explain the event horizon surrounding a black hole
- Discuss why the popular notion of black holes as great sucking monsters that can ingest material at great distances from them is erroneous
- Use the concept of warped spacetime near a black hole to track what happens to any object that might fall into a black hole
- Recognize why the concept of a singularity—with its infinite density and zero volume—presents major challenges to our understanding of matter

Let's now apply what we have learned about gravity and spacetime curvature to the issue we started with: the collapsing core in a very massive star. We saw that if the core's mass is greater than about 3 M_{Sun}, theory says that nothing can stop the core from collapsing forever. We will examine this situation from two perspectives: first from a pre-Einstein point of view, and then with the aid of general relativity.

Classical Collapse

Let's begin with a thought experiment. We want to know what speeds are required to escape from the gravitational pull of different objects. A rocket must be launched from the surface of Earth at a very high speed if it is to escape the pull of Earth's gravity. In fact, any object—rocket, ball, astronomy book—that is thrown into the air with a velocity less than 11 kilometers per second will soon fall back to Earth's surface. Only those objects launched with a speed greater than this *escape velocity* can get away from Earth.

The escape velocity from the surface of the Sun is higher yet—618 kilometers per second. Now imagine that we begin to compress the Sun, forcing it to shrink in diameter. Recall that the pull of gravity depends on both the mass that is pulling you and your distance from the center of gravity of that mass. If the Sun is compressed, its *mass* will remain the same, but the *distance* between a point on the Sun's surface and the center will get smaller and smaller. Thus, as we compress the star, the pull of gravity for an object on the shrinking surface will get stronger and stronger (Figure 24.12).

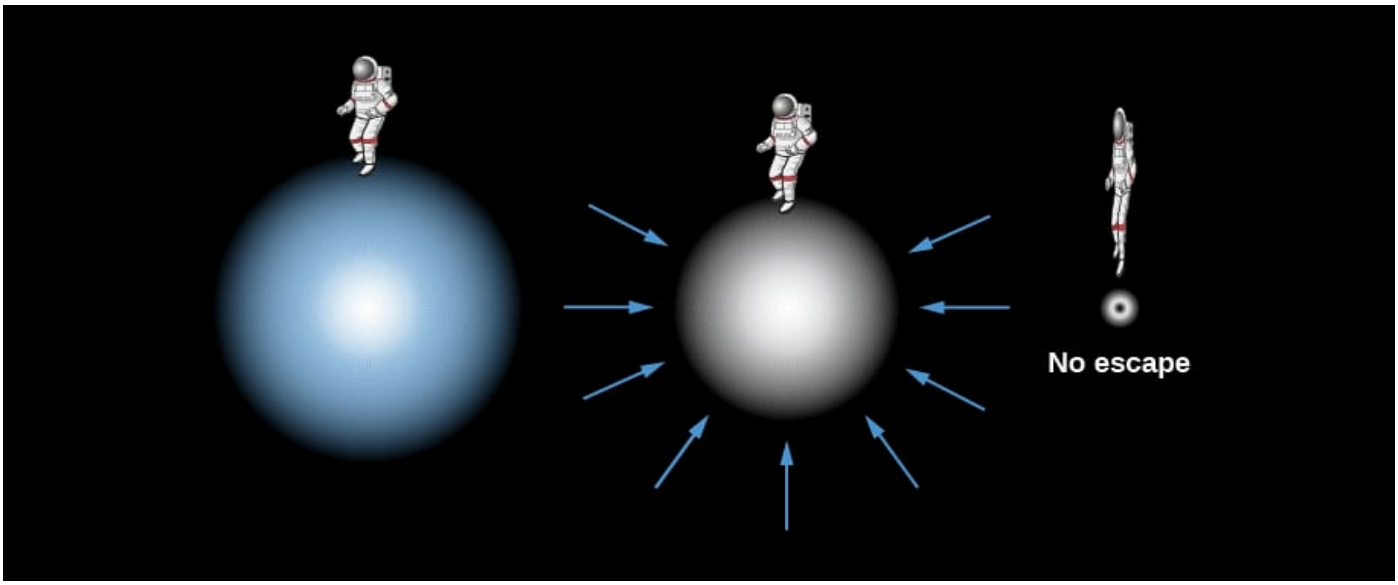


Figure 24.12 : Formation of a Black Hole. At left, an imaginary astronaut floats near the surface of a massive star-core about to collapse. As the same mass falls into a smaller sphere, the gravity at its surface goes up, making it harder for anything to escape from the stellar surface. Eventually the mass collapses into so small a sphere that the escape velocity exceeds the speed of light and nothing can get away. Note that the size of the astronaut has been exaggerated. In the last picture, the astronaut is just outside the sphere we will call the event horizon and is stretched and squeezed by the strong gravity.

When the shrinking Sun reaches the diameter of a neutron star (about 20 kilometers), the velocity required to escape its gravitational pull will be about half the speed of light. Suppose we continue to compress the Sun to a smaller and smaller diameter. (We saw this can't happen to a star like our Sun in the real world because of electron degeneracy, i.e., the mutual repulsion between tightly packed electrons; this is just a quick "thought experiment" to get our bearings).

Ultimately, as the Sun shrinks, the escape velocity near the surface would exceed the speed of light. If the speed you need to get away is faster than the fastest possible speed in the universe, then nothing, not even light, is able to escape. An object with such large escape velocity emits no light, and anything that falls into it can never return.

In modern terminology, we call an object from which light cannot escape a black hole, a name popularized by the America scientist John Wheeler starting in the late 1960s (Figure 24.13). The idea that such objects might exist is, however, not a new one. Cambridge professor and amateur astronomer John Michell wrote a paper in 1783 about the possibility that stars with escape velocities exceeding that of light might exist. And in 1796, the French mathematician Pierre-Simon, marquis de Laplace, made similar calculations using Newton's theory of gravity; he called the resulting objects "dark bodies."

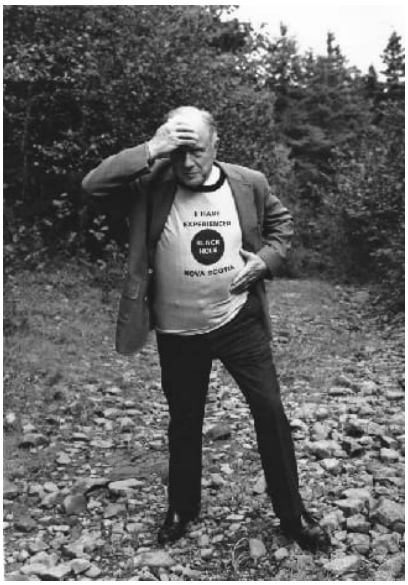


Figure 24.13 : John Wheeler (1911–2008). This brilliant physicist did much pioneering work in general relativity theory and popularized the term *black hole* starting in the late 1960s. (credit: modification of work by Roy Bishop)

While these early calculations provided strong hints that something strange should be expected if very massive objects collapse under their own gravity, we really need general relativity theory to give an adequate description of what happens in such a situation.

Collapse with Relativity

General relativity tells us that gravity is really a curvature of spacetime. As gravity increases (as in the collapsing Sun of our thought experiment), the curvature gets larger and larger. Eventually, if the Sun could shrink down to a diameter of about 6 kilometers, only light beams sent out perpendicular to the surface would escape. All others would fall back onto the star (Figure 24.14). If the Sun could then shrink just a little more, even that one remaining light beam would no longer be able to escape.



Figure 24.14 : Light Paths near a Massive Object. Suppose a person could stand on the surface of a normal star with a flashlight. The light leaving the flashlight travels in a straight line no matter where the flashlight is pointed. Now consider what happens if the star collapses so that it is just a little larger than a black hole. All the light paths, except the one straight up, curve back to the surface. When the star shrinks inside the event horizon and becomes a black hole, even a beam directed straight up returns.

Keep in mind that gravity is not pulling on the light. The concentration of matter has curved spacetime, and light (like the trained

ant of our earlier example) is “doing its best” to go in a straight line, yet is now confronted with a world in which straight lines that used to go outward have become curved paths that lead back in. The collapsing star is a *black hole* in this view, because the very concept of “out” has no geometrical meaning. The star has become trapped in its own little pocket of spacetime, from which there is no escape.

The star’s geometry cuts off communication with the rest of the universe at precisely the moment when, in our earlier picture, the escape velocity becomes equal to the speed of light. The size of the star at this moment defines a surface that we call the event horizon. It’s a wonderfully descriptive name: just as objects that sink below our horizon cannot be seen on Earth, so anything happening inside the event horizon can no longer interact with the rest of the universe.

Imagine a future spacecraft foolish enough to land on the surface of a massive star just as it begins to collapse in the way we have been describing. Perhaps the captain is asleep at the gravity meter, and before the crew can say “Albert Einstein,” they have collapsed with the star inside the event horizon. Frantically, they send an escape pod straight outward. But paths outward twist around to become paths inward, and the pod turns around and falls toward the center of the black hole. They send a radio message to their loved ones, bidding good-bye. But radio waves, like light, must travel through spacetime, and curved spacetime allows nothing to get out. Their final message remains unheard. Events inside the event horizon can never again affect events outside it.

The characteristics of an event horizon were first worked out by astronomer and mathematician Karl Schwarzschild (Figure 24.15). A member of the German army in World War I, he died in 1916 of an illness he contracted while doing artillery shell calculations on the Russian front. His paper on the theory of event horizons was among the last things he finished as he was dying; it was the first exact solution to Einstein’s equations of general relativity. The radius of the event horizon is called the *Schwarzschild radius* in his memory.



Figure 24.15 : Karl Schwarzschild (1873–1916). This German scientist was the first to demonstrate mathematically that a black hole is possible and to determine the size of a nonrotating black hole’s event horizon.

The event horizon is the boundary of the black hole; calculations show that it does not get smaller once the whole star has collapsed inside it. It is the region that separates the things trapped inside it from the rest of the universe. Anything coming from the outside is also trapped once it comes inside the event horizon. The horizon’s size turns out to depend only on the mass inside it. If the Sun, with its mass of 1 M_{Sun}, were to become a black hole (fortunately, it can’t—this is just a thought experiment), the Schwarzschild radius would be about 3 kilometers; thus, the entire black hole would be about one-third the size of a neutron star of that same mass. Feed the black hole some mass, and the horizon will grow—but not very much. Doubling the mass will make the black hole 6 kilometers in radius, still very tiny on the cosmic scale.

The event horizons of more massive black holes have larger radii. For example, if a globular cluster of 100,000 stars (solar masses) could collapse to a black hole, it would be 300,000 kilometers in radius, a little less than half the radius of the Sun. If the entire Galaxy could collapse to a black hole, it would be only about 10^{12} kilometers in radius—about a tenth of a light year. Smaller masses have correspondingly smaller horizons: for Earth to become a black hole, it would have to be compressed to a radius of only 1 centimeter—less than the size of a grape. A typical asteroid, if crushed to a small enough size to be a black hole,

would have the dimensions of an atomic nucleus.

Example 24.1: The Milky Way's Black Hole

The size of the event horizon of a black hole depends on the mass of the black hole. The greater the mass, the larger the radius of the event horizon. General relativity calculations show that the formula for the Schwarzschild radius (R_S) of the event horizon is

$$R_S = \frac{2GM}{c^2}$$

where *c* is the speed of light, *G* is the gravitational constant, and *M* is the mass of the black hole. Note that in this formula, 2, *G*, and *c* are all constant; only the mass changes from black hole to black hole.

As we will see in the chapter on The Milky Way Galaxy, astronomers have traced the paths of several stars near the center of our Galaxy and found that they seem to be orbiting an unseen object—dubbed Sgr A* (pronounced "Sagittarius A-star")—with a mass of about 4 million solar masses. What is the size of its Schwarzschild radius?

Solution

We can substitute data for *G*, *M*, and *c* (from Appendix E) directly into the equation:

$$\begin{aligned} R_S &= \tfrac{2GM}{c^2} = \tfrac{2(6.67\times10^{-11}\,\mathrm{N\cdot m^2/kg^2})(4\times10^{6})(1.99\times10^{30}\,\mathrm{kg})}{(3.00\times10^{8}\,\mathrm{m/s})^2} \\ &= 1.18\times10^{10}\,\mathrm{m} \end{aligned}$$

This distance is about one-fifth of the radius of Mercury's orbit around the Sun, yet the object contains 4 million solar masses and cannot be seen with our largest telescopes. You can see why astronomers are convinced this object is a black hole.

Exercise 14.6.1

What would be the size of a black hole that contained only as much mass as a typical pickup truck (about 3000 kg)? (Note that something with so little mass could never actually form a black hole, but it's interesting to think about the result.)

Answer available at the end of this section.

A Black Hole Myth

Much of the modern folklore about black holes is misleading. One idea you may have heard is that black holes go about sucking things up with their gravity. Actually, it is only very close to a black hole that the strange effects we have been discussing come into play. The gravitational attraction far away from a black hole is the same as that of the star that collapsed to form it.

Remember that the gravity of any star some distance away acts as if all its mass were concentrated at a point in the center, which we call the center of gravity. For real stars, we merely *imagine* that all mass is concentrated there; for black holes, all the mass *really is* concentrated at a point in the center.

So, if you are a star or distant planet orbiting around a star that becomes a black hole, your orbit may not be significantly affected by the collapse of the star (although it may be affected by any mass loss that precedes the collapse). If, on the other hand, you venture close to the event horizon, it would be very hard for you to resist the "pull" of the warped spacetime near the black hole. You have to get really close to the black hole to experience any significant effect.

If another star or a spaceship were to pass one or two solar radii from a black hole, Newton's laws would be adequate to describe what would happen to it. Only very near the event horizon of a black hole is the gravitation so strong that Newton's laws break down. The black hole remnant of a massive star coming into our neighborhood would be far, far safer to us than its earlier incarnation as a brilliant, hot star.

Making Connections: Gravity and Time Machines

Time machines are one of the favorite devices of science fiction. Such a device would allow you to move through time at a different pace or in a different direction from everyone else. General relativity suggests that it is possible, in theory, to construct a time machine using gravity that could take you into the future.

Let's imagine a place where gravity is terribly strong, such as near a black hole. General relativity predicts that the stronger the gravity, the slower the pace of time (as seen by a distant observer). So, imagine a future astronaut, with a fast and strongly built spaceship, who volunteers to go on a mission to such a high-gravity environment. The astronaut leaves in the year 2222, just after graduating from college at age 22. She takes, let's say, exactly 10 years to get to the black hole. Once there, she orbits some distance from it, taking care not to get pulled in.

She is now in a high-gravity realm where time passes much more slowly than it does on Earth. This isn't just an effect on the mechanism of her clocks—*time itself* is running slowly. That means that every way she has of measuring time will give the same slowed-down reading when compared to time passing on Earth. Her heart will beat more slowly, her hair will grow more slowly, her antique wristwatch will tick more slowly, and so on. She is not aware of this slowing down because all her readings of time, whether made by her own bodily functions or with mechanical equipment, are measuring the same—slower—time. Meanwhile, back on Earth, time passes as it always does.

Our astronaut now emerges from the region of the black hole, her mission of exploration finished, and returns to Earth. Before leaving, she carefully notes that (according to her timepieces) she spent about 2 weeks around the black hole. She then takes exactly 10 years to return to Earth. Her calculations tell her that since she was 22 when she left the Earth, she will be 42 plus 2 weeks when she returns. So, the year on Earth, she figures, should be 2242, and her classmates should now be approaching their midlife crises.

But our astronaut should have paid more attention in her astronomy class! Because time slowed down near the black hole, much less time passed for her than for the people on Earth. While her clocks measured 2 weeks spent near the black hole, more than 2000 weeks (depending on how close she got) could well have passed on Earth. That's equal to 40 years, meaning her classmates will be senior citizens in their 80s when she (a mere 42-year-old) returns. On Earth it will be not 2242, but 2282—and she will say that she has arrived *in the future*.

Is this scenario real? Well, it has a few practical challenges: we don't think any black holes are close enough for us to reach in 10 years, and we don't think any spaceship or human can survive near a black hole. But the key point about the slowing down of time is a natural consequence of Einstein's general theory of relativity, and we saw that its predictions have been confirmed by experiment after experiment.

Such developments in the understanding of science also become inspiration for science fiction writers. Recently, the film *Interstellar* featured the protagonist traveling close to a massive black hole; the resulting delay in his aging relative to his earthbound family is a key part of the plot.

Science fiction novels, such as *Gateway* by Frederik Pohl and *A World out of Time* by Larry Niven, also make use of the slowing down of time near black holes as major turning points in the story. For a list of science fiction stories based on good astronomy, you can go to bit.ly/astroscifi.

A Trip into a Black Hole

The fact that scientists cannot see inside black holes has not kept them from trying to calculate what they are like. One of the first things these calculations showed was that the formation of a black hole obliterates nearly all information about the star that collapsed to form it. Physicists like to say "black holes have no hair," meaning that nothing sticks out of a black hole to give us clues about what kind of star produced it or what material has fallen inside. The only information a black hole can reveal about itself is its mass, its spin (rotation), and whether it has any electrical charge.

What happens to the collapsing star-core that made the black hole? Our best calculations predict that the material will continue to collapse under its own weight, forming an infinitely *squozen* point—a place of zero volume and infinite density—to which we give the name singularity. At the singularity, spacetime ceases to exist. The laws of physics as we know them break down. We do not yet have the physical understanding or the mathematical tools to describe the singularity itself, or even if singularities

actually occur. From the outside, however, the entire structure of a basic black hole (one that is not rotating) can be described as a singularity surrounded by an event horizon. Compared to humans, black holes are really very simple objects.

Scientists have also calculated what would happen if an astronaut were to fall into a black hole. Let's take up an observing position a long, safe distance away from the event horizon and watch this astronaut fall toward it. At first he falls away from us, moving ever faster, just as though he were approaching any massive star. However, as he nears the event horizon of the black hole, things change. The strong gravitational field around the black hole will make his clocks run more slowly, when seen from our outside perspective.

If, as he approaches the event horizon, he sends out a signal once per second according to his clock, we will see the spacing between his signals grow longer and longer until it becomes infinitely long when he reaches the event horizon. (Recalling our discussion of gravitational redshift, we could say that if the infalling astronaut uses a blue light to send his signals every second, we will see the light get redder and redder until its wavelength is nearly infinite.) As the spacing between clock ticks approaches infinity, it will appear to us that the astronaut is slowly coming to a stop, frozen in time at the event horizon.

In the same way, all matter falling into a black hole will also appear to an outside observer to stop at the event horizon, frozen in place and taking an infinite time to fall through it. But don't think that matter falling into a black hole will therefore be easily visible at the event horizon. The tremendous redshift will make it very difficult to observe any radiation from the "frozen" victims of the black hole.

This, however, is only how we, located far away from the black hole, see things. To the astronaut, his time goes at its normal rate and he falls right on through the event horizon into the black hole. (Remember, this horizon is not a physical barrier, but only a region in space where the curvature of spacetime makes escape impossible.)

You may have trouble with the idea that you (watching from far away) and the astronaut (falling in) have such different ideas about what has happened. This is the reason Einstein's ideas about space and time are called theories of *relativity*. What each observer measures about the world depends on (is relative to) his or her frame of reference. The observer in strong gravity measures time and space differently from the one sitting in weaker gravity. When Einstein proposed these ideas, many scientists also had difficulty with the idea that two such different views of the same event could be correct, each in its own "world," and they tried to find a mistake in the calculations. There were no mistakes: we and the astronaut really would see him fall into a black hole very differently.

For the astronaut, there is no turning back. Once inside the event horizon, the astronaut, along with any signals from his radio transmitter, will remain hidden forever from the universe outside. He will, however, not have a long time (from his perspective) to feel sorry for himself as he approaches the black hole. Suppose he is falling feet first. The force of gravity that the singularity exerts on his feet is greater than on his head, so he will be stretched slightly. Because the singularity is a point, the left side of his body will be pulled slightly toward the right, and the right slightly toward the left, bringing each side closer to the singularity. The astronaut will therefore be slightly squeezed in one direction and stretched in the other. Some scientists like to call this process of stretching and narrowing *spaghettification*. The point at which the astronaut becomes so stretched that he perishes depends on the size of the black hole. For black holes with masses billions of times the mass of the Sun, such as those found at the centers of galaxies, the spaghettification becomes significant only after the astronaut passes through the event horizon. For black holes with masses of a few solar masses, the astronaut will be stretched and ripped apart even before he reaches the event horizon.

Earth exerts similar *tidal forces* on an astronaut performing a spacewalk. In the case of Earth, the tidal forces are so small that they pose no threat to the health and safety of the astronaut. Not so in the case of a black hole. Sooner or later, as the astronaut approaches the black hole, the tidal forces will become so great that the astronaut will be ripped apart, eventually reduced to a collection of individual atoms that will continue their inexorable fall into the singularity.

Link to Learning

From the previous discussion, you will probably agree that jumping into a black hole is definitely a once-in-a-lifetime experience! You can see an engaging explanation of death by black hole by Neil deGrasse Tyson, where he explains the effect of tidal forces on the human body until it dies by spaghettification.

An overview of black holes is given in this Discovery Channel video excerpt.

Answers

Exercise 14.6.1

Substituting the data into our equation gives

$R_S = \frac{2GM}{c^2} = \frac{2(6.67 \times 10^{-11}\,\mathrm{N \cdot m^2/kg^2})(3000\,\mathrm{kg})}{(3.00 \times 10^8\,\mathrm{m/s})^2} = 4.4 \times 10^{-24}\,\mathrm{m}.$

For comparison, the size of a proton is usually considered to be about 8×10^{-16} m, which would be about ten million times larger.

This page titled 14.6: Black Holes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 24.6: Black Holes by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

14.7: Evidence for Black Holes

Learning Objectives

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

- Describe what to look for when seeking and confirming the presence of a stellar black hole
- Explain how a black hole is inherently black yet can be associated with luminous matter
- Differentiate between stellar black holes and the black holes in the centers of galaxies

Theory tells us what black holes are like. But do they actually exist? And how do we go about looking for something that is many light years away, only about a few dozen kilometers across (if a stellar black hole), and completely black? It turns out that the trick is not to look for the black hole itself but instead to look for what it does to a nearby companion star.

As we saw, when very massive stars collapse, they leave behind their gravitational influence. What if a member of a double-star system becomes a black hole, and its companion manages to survive the death of the massive star? While the black hole disappears from our view, we may be able to deduce its presence from the things it does to its companion.

Requirements for a Black Hole

So, here is a prescription for finding a black hole: start by looking for a star whose motion (determined from the Doppler shift of its spectral lines) shows it to be a member of a binary star system. If both stars are visible, neither can be a black hole, so focus your attention on just those systems where only one star of the pair is visible, even with our most sensitive telescopes.

Being invisible is not enough, however, because a relatively faint star might be hard to see next to the glare of a brilliant companion or if it is shrouded by dust. And even if the star really is invisible, it could be a neutron star. Therefore, we must also have evidence that the unseen star has a mass too high to be a neutron star and that it is a collapsed object—an extremely small stellar remnant.

We can use Kepler's law (see Orbits and Gravity) and our knowledge of the visible star to measure the mass of the invisible member of the pair. If the mass is greater than about 3 M_{Sun}, then we are likely seeing (or, more precisely, not seeing) a black hole—as long as we can make sure the object really is a collapsed star.

If matter falls toward a compact object of high gravity, the material is accelerated to high speed. Near the event horizon of a black hole, matter is moving at velocities that approach the speed of light. As the atoms whirl chaotically toward the event horizon, they rub against each other; internal friction can heat them to temperatures of 100 million K or more. Such hot matter emits radiation in the form of flickering X-rays. The last part of our prescription, then, is to look for a source of X-rays associated with the binary system. Since X-rays do not penetrate Earth's atmosphere, such sources must be found using X-ray telescopes in space.

In our example, the infalling gas that produces the X-ray emission comes from the black hole's companion star. As we saw in The Death of Stars, stars in close binary systems can exchange mass, especially as one of the members expands into a red giant. Suppose that one star in a double-star system has evolved to a black hole and that the second star begins to expand. If the two stars are not too far apart, the outer layers of the expanding star may reach the point where the black hole exerts more gravitational force on them than do the inner layers of the red giant to which the atmosphere belongs. The outer atmosphere then passes through the point of no return between the stars and falls toward the black hole.

The mutual revolution of the giant star and the black hole causes the material falling toward the black hole to spiral around it rather than flow directly into it. The infalling gas whirls around the black hole in a pancake of matter called an accretion disk. It is within the inner part of this disk that matter is revolving about the black hole so fast that internal friction heats it up to X-ray–emitting temperatures (see Figure 24.1).

Another way to form an accretion disk in a binary star system is to have a powerful stellar wind come from the black hole's companion. Such winds are a characteristic of several stages in a star's life. Some of the ejected gas in the wind will then flow close enough to the black hole to be captured by it into the disk (Figure 24.16).

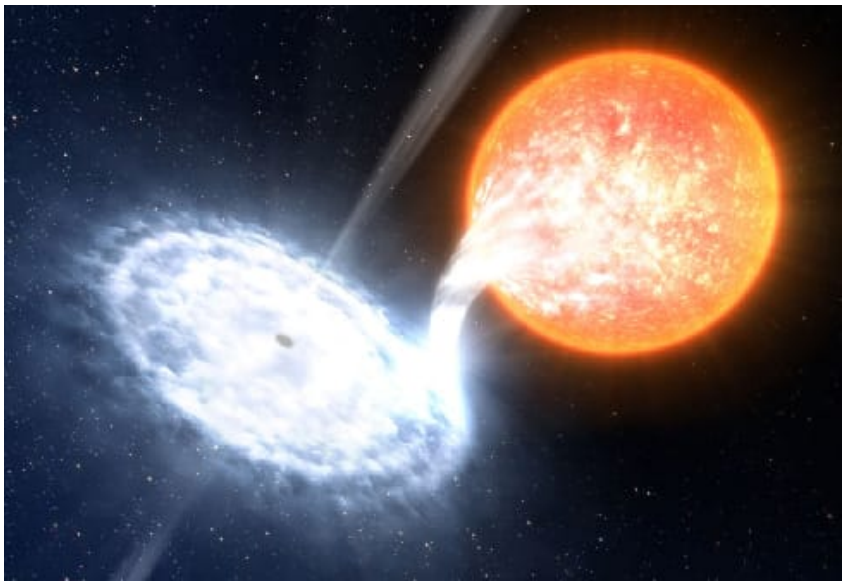


Figure 24.16 : Binary Black Hole. This artist's rendition shows a black hole and star (red). As matter streams from the star, it forms a disk around the black hole. Some of the swirling material close to the black hole is pushed outward perpendicular to the disk in two narrow jets. (credit: modification of work by ESO/L. Calçada)

We should point out that, as often happens, the measurements we have been discussing are not quite as simple as they are described in introductory textbooks. In real life, Kepler's law allows us to calculate only the combined mass of the two stars in the binary system. We must learn more about the visible star of the pair and its history to ascertain the distance to the binary pair, the true size of the visible star's orbit, and how the orbit of the two stars is tilted toward Earth, something we can rarely measure. And neutron stars can also have accretion disks that produce X-rays, so astronomers must study the properties of these X-rays carefully when trying to determine what kind of object is at the center of the disk. Nevertheless, a number of systems that clearly contain black holes have now been found.

The Discovery of Stellar-Mass Black Holes

Because X-rays are such important tracers of black holes that are having some of their stellar companions for lunch, the search for black holes had to await the launch of sophisticated X-ray telescopes into space. These instruments must have the resolution to locate the X-ray sources accurately and thereby enable us to match them to the positions of binary star systems.

The first black hole binary system to be discovered is called Cygnus X-1 (see Figure 24.1). The visible star in this binary system is spectral type O. Measurements of the Doppler shifts of the O star's spectral lines show that it has an unseen companion. The X-rays flickering from it strongly indicate that the companion is a small collapsed object. The mass of the invisible collapsed companion is about 15 times that of the Sun. The companion is therefore too massive to be either a white dwarf or a neutron star.

A number of other binary systems also meet all the conditions for containing a black hole. Table 24.1 lists the characteristics of some of the best examples.

Table 24.1: Some Black Hole Candidates in Binary Star Systems

Name/Catalog Designation[2]	Companion Star Spectral Type	Orbital Period (days)	Black Hole Mass Estimates (M_{Sun})
LMC X-1	O giant	3.9	10.9
Cygnus X-1	O supergiant	5.6	15
XTE J1819.3-254 (V4641 Sgr)	B giant	2.8	6–7
LMC X-3	B main sequence	1.7	7
4U1543-475 (IL Lup)	A main sequence	1.1	9

Name/Catalog Designation[2]	Companion Star Spectral Type	Orbital Period (days)	Black Hole Mass Estimates (M_{Sun})
GRO J1655-40 (V1033 Sco)	F subgiant	2.6	7
GRS 1915+105	K giant	33.5	14
GS2023+338 (V404 Cyg)	K giant	6.5	12
XTE J1550-564	K giant	1.5	11
A0620-00 (V616 Mon)	K main sequence	0.33	9–13
H1705-250 (Nova Oph 1977)	K main sequence	0.52	5–7
GRS1124-683 (Nova Mus 1991)	K main sequence	0.43	7
GS2000+25 (QZ Vul)	K main sequence	0.35	5–10
GRS1009-45 (Nova Vel 1993)	K dwarf	0.29	8–9
XTE J1118+480	K dwarf	0.17	7
XTE J1859+226	K dwarf	0.38	5.4
GRO J0422+32	M dwarf	0.21	4

Feeding a Black Hole

After an isolated star, or even one in a binary star system, becomes a black hole, it probably won't be able to grow much larger. Out in the suburban regions of the Milky Way Galaxy where we live (see The Milky Way Galaxy), stars and star systems are much too far apart for other stars to provide "food" to a hungry black hole. After all, material must approach very close to the event horizon before the gravity is any different from that of the star before it became the black hole.

But, as will see, the central regions of galaxies are quite different from their outer parts. Here, stars and raw material can be quite crowded together, and they can interact much more frequently with each other. Therefore, black holes in the centers of galaxies may have a much better opportunity to find mass close enough to their event horizons to pull in. Black holes are not particular about what they "eat": they are happy to consume other stars, asteroids, gas, dust, and even other black holes. (If two black holes merge, you just get a black hole with more mass and a larger event horizon.)

As a result, black holes in crowded regions can grow, eventually swallowing thousands or even millions of times the mass of the Sun. Ground-based observations have provided compelling evidence that there is a black hole in the center of our own Galaxy with a mass of about 4 million times the mass of the Sun (we'll discuss this further in the chapter on The Milky Way Galaxy). Observations with the Hubble Space Telescope have shown dramatic evidence for the existence of black holes in the centers of many other galaxies. These black holes can contain more than a billion solar masses. The feeding frenzy of such supermassive black holes may be responsible for some of the most energetic phenomena in the universe (see Active Galaxies, Quasars, and Supermassive Black Holes). And evidence from more recent X-ray observations is also starting to indicate the existence of "middle-weight" black holes, whose masses are dozens to thousands of times the mass of the Sun. The crowded inner regions of the globular clusters we described in Stars from Adolescence to Old Age may be just the right breeding grounds for such intermediate-mass black holes.

Over the past decades, many observations, especially with the Hubble Space Telescope and with X-ray satellites, have been made that can be explained only if black holes really do exist. Furthermore, the observational tests of Einstein's general theory of relativity have convinced even the most skeptical scientists that his picture of warped or curved spacetime is indeed our best description of the effects of gravity near these black holes.

Footnotes

- 2As you can tell, there is no standard way of naming these candidates. The chain of numbers is the location of the source in right ascension and declination (the longitude and latitude system of the sky); some of the letters preceding the numbers refer to objects (e.g., LMC) and constellations (e.g., Cygnus), while other letters refer to the satellite that discovered the

candidate—A for Ariel, G for Ginga, and so on. The notations in parentheses are those used by astronomers who study binary star system or novae.

This page titled 14.7: Evidence for Black Holes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 24.7: Evidence for Black Holes by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15: THE MILKY WAY, OUR ISLAND UNIVERSE

15.1: The Architecture of the Galaxy
15.2: The Interstellar Medium
15.3: Interstellar Gas
15.4: Cosmic Dust
15.5: The Life Cycle of Cosmic Material
15.6: Spiral Structure
15.7: The Mass of the Galaxy
15.8: Stellar Populations in the Galaxy
15.9: The Formation of the Galaxy
15.10: The Center of the Galaxy

This page titled 15: The Milky Way, Our Island Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

15.1: The Architecture of the Galaxy

Learning Objectives

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

- Explain why William and Caroline Herschel concluded that the Milky Way has a flattened structure centered on the Sun and solar system
- Describe the challenges of determining the Galaxy's structure from our vantage point within it
- Identify the main components of the Galaxy

The Milky Way Galaxy surrounds us, and you might think it is easy to study because it is so close. However, the very fact that we are embedded within it presents a difficult challenge. Suppose you were given the task of mapping New York City. You could do a much better job from a helicopter flying over the city than you could if you were standing in Times Square. Similarly, it would be easier to map our Galaxy if we could only get a little way outside it, but instead we are trapped inside and way out in its suburbs—far from the galactic equivalent of Times Square.

Herschel Measures the Galaxy

In 1785, William Herschel (Figure 25.2) made the first important discovery about the architecture of the Milky Way Galaxy. Using a large reflecting telescope that he had built, William and his sister Caroline counted stars in different directions of the sky. They found that most of the stars they could see lay in a flattened structure encircling the sky, and that the numbers of stars were about the same in any direction around this structure. Herschel therefore concluded that the stellar system to which the Sun belongs has the shape of a disk or wheel (he might have called it a Frisbee except Frisbees hadn't been invented yet), and that the Sun must be near the hub of the wheel (Figure 25.3).



Figure 25.2 : William Herschel (1738–1822) and Caroline Herschel (1750–1848). William Herschel was a German musician who emigrated to England and took up astronomy in his spare time. He discovered the planet Uranus, built several large telescopes, and made measurements of the Sun's place in the Galaxy, the Sun's motion through space, and the comparative brightnesses of stars. This painting shows William and his sister Caroline polishing a telescope lens. (credit: modification of work by the Wellcome Library)

To understand why Herschel reached this conclusion, imagine that you are a member of a band standing in formation during halftime at a football game. If you count the band members you see in different directions and get about the same number each time, you can conclude that the band has arranged itself in a circular pattern with you at the center. Since you see no band members above you or underground, you know that the circle made by the band is much flatter than it is wide.

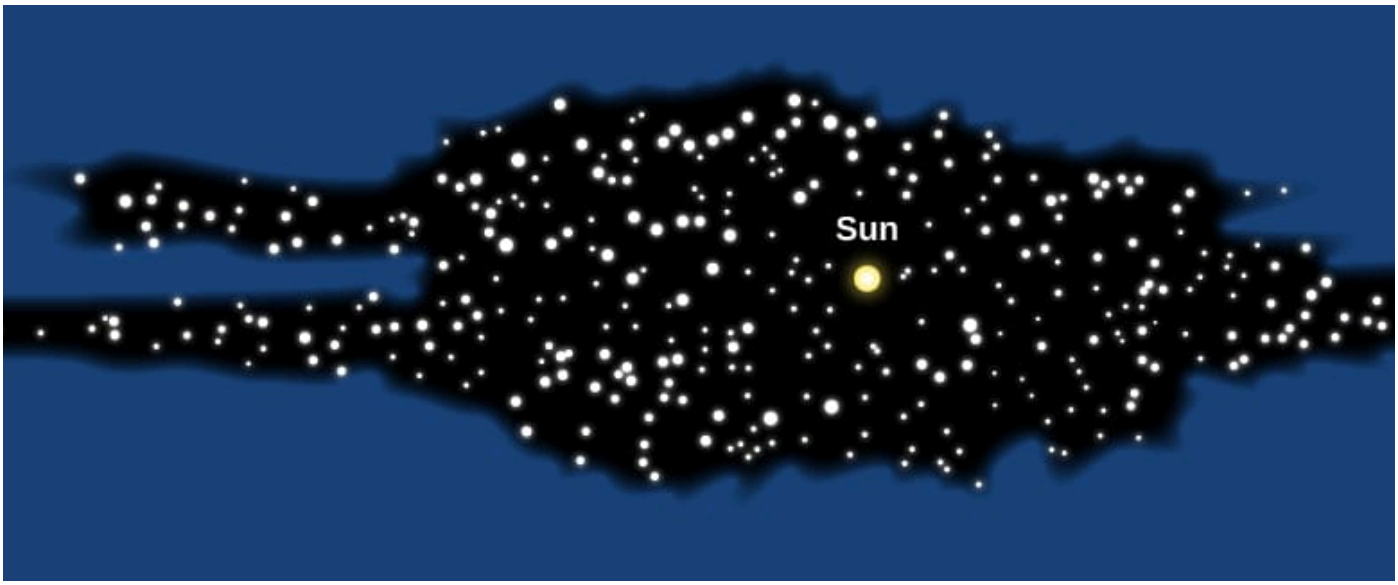


Figure 25.3 : Herschel's Diagram of the Milky Way. Herschel constructed this cross section of the Galaxy by counting stars in various directions.

We now know that Herschel was right about the shape of our system, but wrong about where the Sun lies within the disk. As we saw in Between the Stars: Gas and Dust in Space, we live in a dusty Galaxy. Because interstellar dust absorbs the light from stars, Herschel could see only those stars within about 6000 light-years of the Sun. Today we know that this is a very small section of the entire 100,000-light-year-diameter disk of stars that makes up the Galaxy.

Voyagers in Astronomy: Harlow Shapley: Mapmaker to the Stars

Until the early 1900s, astronomers generally accepted Herschel's conclusion that the Sun is near the center of the Galaxy. The discovery of the Galaxy's true size and our actual location came about largely through the efforts of Harlow Shapley. In 1917, he was studying RR Lyrae variable stars in globular clusters. By comparing the known intrinsic luminosity of these stars to how bright they appeared, Shapley could calculate how far away they are. (Recall that it is distance that makes the stars look dimmer than they would be "up close," and that the brightness fades as the distance squared.) Knowing the distance to any star in a cluster then tells us the distance to the cluster itself.

Globular clusters can be found in regions that are free of interstellar dust and so can be seen at very large distances. When Shapley used the distances and directions of 93 globular clusters to map out their positions in space, he found that the clusters are distributed in a spherical volume, which has its center not at the Sun but at a distant point along the Milky Way in the direction of Sagittarius. Shapley then made the bold assumption, verified by many other observations since then, that the point on which the system of globular clusters is centered is also the center of the entire Galaxy (Figure 25.4).

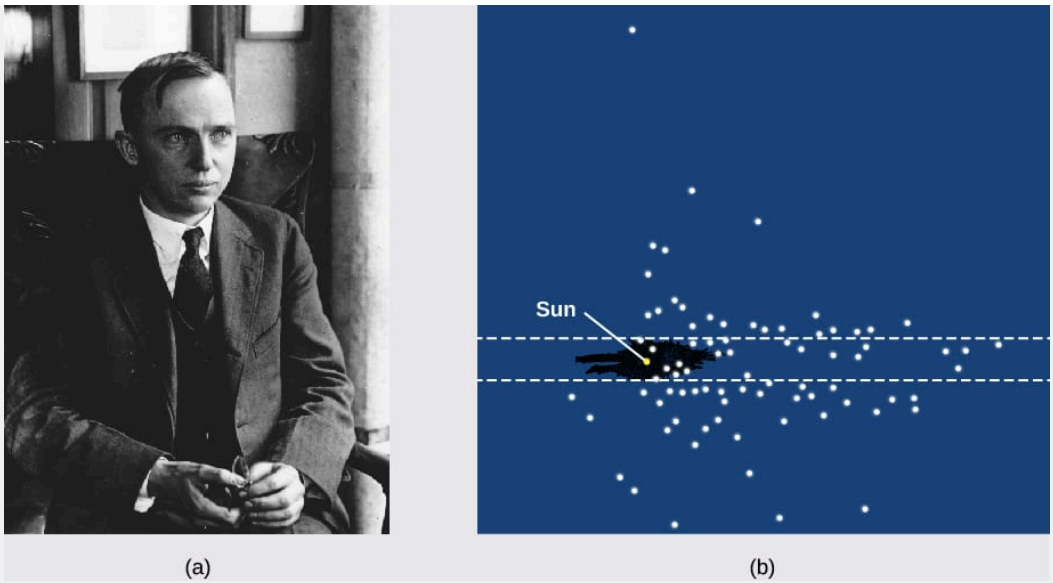


Figure 25.4 : Harlow Shapley and His Diagram of the Milky Way. (a) Shapley poses for a formal portrait. (b) His diagram shows the location of globular clusters, with the position of the Sun also marked. The black area shows Herschel's old diagram, centered on the Sun, approximately to scale.

Shapley's work showed once and for all that our star has no special place in the Galaxy. We are in a nondescript region of the Milky Way, only one of 200 to 400 billion stars that circle the distant center of our Galaxy.

Born in 1885 on a farm in Missouri, Harlow Shapley at first dropped out of school with the equivalent of only a fifth-grade education. He studied at home and at age 16 got a job as a newspaper reporter covering crime stories. Frustrated by the lack of opportunities for someone who had not finished high school, Shapley went back and completed a six-year high-school program in only two years, graduating as class valedictorian.

In 1907, at age 22, he went to the University of Missouri, intent on studying journalism, but found that the school of journalism would not open for a year. Leafing through the college catalog (or so he told the story later), he chanced to see "Astronomy" among the subjects beginning with "A." Recalling his boyhood interest in the stars, he decided to study astronomy for the next year (and the rest, as the saying goes, is history).

Upon graduation Shapley received a fellowship for graduate study at Princeton and began to work with the brilliant Henry Norris Russell (see the Henry Norris Russell feature box). For his PhD thesis, Shapley made major contributions to the methods of analyzing the behavior of eclipsing binary stars. He was also able to show that cepheid variable stars are not binary systems, as some people thought at the time, but individual stars that pulsate with striking regularity.

Impressed with Shapley's work, George Ellery Hale offered him a position at the Mount Wilson Observatory, where the young man took advantage of the clear mountain air and the 60-inch reflector to do his pioneering study of variable stars in globular clusters.

Shapley subsequently accepted the directorship of the Harvard College Observatory, and over the next 30 years, he and his collaborators made contributions to many fields of astronomy, including the study of neighboring galaxies, the discovery of dwarf galaxies, a survey of the distribution of galaxies in the universe, and much more. He wrote a series of nontechnical books and articles and became known as one of the most effective popularizers of astronomy. Shapley enjoyed giving lectures around the country, including at many smaller colleges where students and faculty rarely got to interact with scientists of his caliber.

During World War II, Shapley helped rescue many scientists and their families from Eastern Europe; later, he helped found UNESCO, the United Nations Educational, Scientific, and Cultural Organization. He wrote a pamphlet called *Science from Shipboard* for men and women in the armed services who had to spend many weeks on board transport ships to Europe. And during the difficult period of the 1950s, when congressional committees began their "witch hunts" for communist

sympathizers (including such liberal leaders as Shapley), he spoke out forcefully and fearlessly in defense of the freedom of thought and expression. A man of many interests, he was fascinated by the behavior of ants, and wrote scientific papers about them as well as about galaxies.

By the time he died in 1972, Shapley was acknowledged as one of the pivotal figures of modern astronomy, a "twentieth-century Copernicus" who mapped the Milky Way and showed us our place in the Galaxy.

Link to Learning

To find more information about Shapley's life and work, see the entry for him on the Bruce Medalists website. (This site features the winners of the Bruce Medal of the Astronomical Society of the Pacific, one of the highest honors in astronomy; the list is a who's who of some of the greatest astronomers of the last twelve decades.)

Disks and Haloes

With modern instruments, astronomers can now penetrate the "smog" of the Milky Way by studying radio and infrared emissions from distant parts of the Galaxy. Measurements at these wavelengths (as well as observations of other galaxies like ours) have given us a good idea of what the Milky Way would look like if we *could* observe it from a distance.

Figure 25.5 sketches what we would see if we could view the Galaxy face-on and edge-on. The brightest part of the Galaxy consists of a thin, circular, rotating disk of stars distributed across a region about 100,000 light-years in diameter and about 2000 light-years thick. (Given how thin the disk is, perhaps a CD is a more appropriate analogy than a wheel.) The very youngest stars, and the dust and gas from which stars form, are found typically within 100 light-years of the plane of the Milky Way Galaxy. The mass of the interstellar matter is about 15% of the mass of the stars in this disk.

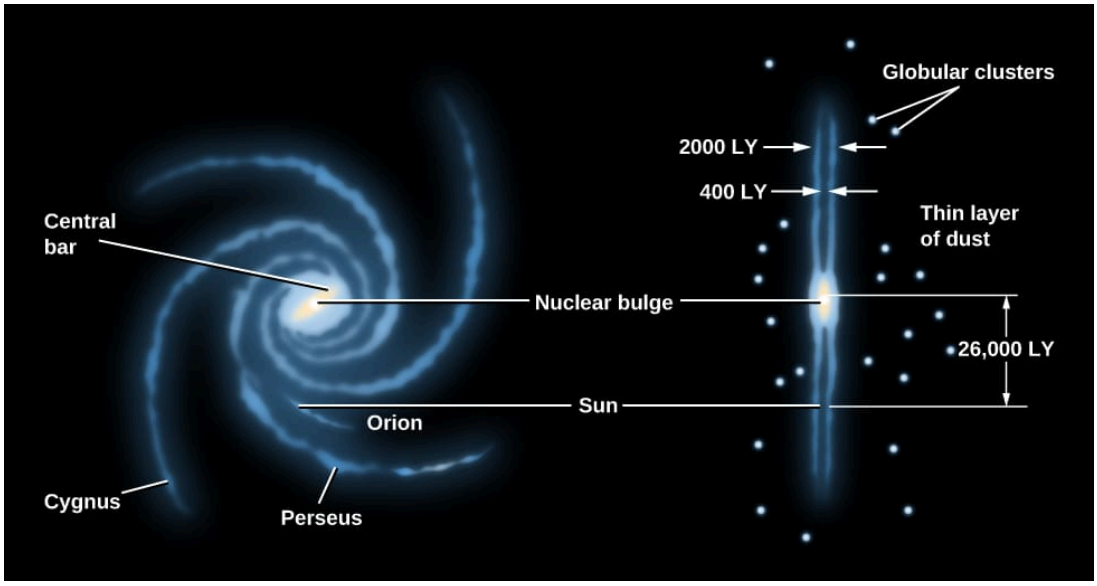


Figure 25.5 : Schematic Representation of the Galaxy. The left image shows the face-on view of the spiral disk; the right image shows the view looking edge-on along the disk. The major spiral arms are labeled. The Sun is located on the inside edge of the short Orion spur.

As the diagram in Figure 25.5 shows, the stars, gas, and dust are not spread evenly throughout the disk but are concentrated into a central bar and a series of spiral arms. Recent infrared observations have confirmed that the central bar is composed mostly of old yellow-red stars. The two main spiral arms appear to connect with the ends of the bar. They are highlighted by the blue light from young hot stars. We know many other spiral galaxies that also have bar-shaped concentrations of stars in their central regions; for that reason they are called *barred spirals*. Figure 25.6 shows two other galaxies—one without a bar and one with a strong bar—to give you a basis for comparison to our own. We will describe our spiral structure in more detail shortly. The Sun is located about halfway between the center of the Galaxy and the edge of the disk and only about 70 light-years above its central plane.



Figure 25.6 : Unbarred and Barred Spiral Galaxies. (a) This image shows the unbarred spiral galaxy M74. It contains a small central bulge of mostly old yellow-red stars, along with spiral arms that are highlighted with the blue light from young hot stars. (b) This image shows the strongly barred spiral galaxy NGC 1365. The bulge and the fainter bar both appear yellowish because the brightest stars in them are mostly old yellow and red giants. Two main spiral arms project from the ends of the bar. As in M74, these spiral arms are populated with blue stars and red patches of glowing gas—hallmarks of recent star formation. The Milky Way Galaxy is thought to have a barred spiral structure that is intermediate between these two examples. (credit a: modification of work by ESO/PESSTO/S. Smartt; credit b: modification of work by ESO)

Our thin disk of young stars, gas, and dust is embedded in a thicker but more diffuse disk of older stars; this thicker disk extends about 1000 light-years above and 1000 light-years below the midplane of the thin disk and contains only about 5% as much mass as the thin disk. The stars thin out with distance from the galactic plane and don't have a sharp edge. Approximately 2/3 of the stars in the thick disk are within 1000 light-years of midplane.

Close in to the galactic center (within about 10,000 light-years), the stars are no longer confined to the disk but form a central bulge (or nuclear bulge). When we observe with visible light, we can glimpse the stars in the bulge only in those rare directions where there happens to be relatively little interstellar dust. The first picture that actually succeeded in showing the bulge as a whole was taken at infrared wavelengths (Figure 25.7).



Figure 25.7 : Inner Part of the Milky Way Galaxy. This beautiful infrared map, showing half a billion stars, was obtained as part of the Two Micron All Sky Survey (2MASS). Because interstellar dust does not absorb infrared as strongly as visible light, this view reveals the previously hidden bulge of old stars that surrounds the center of our Galaxy, along with the Galaxy's thin disk component. (credit: modification of work by 2MASS/J. Carpenter, T. H. Jarrett, and R. Hurt)

The fact that much of the bulge is obscured by dust makes its shape difficult to determine. For a long time, astronomers assumed it was spherical. However, infrared images and other data indicate that the bulge is about two times longer than it is wide, and shaped rather like a peanut. The relationship between this elongated inner bulge and the larger bar of stars remains uncertain. At the very center of the nuclear bulge is a tremendous concentration of matter, which we will discuss later in this chapter.

In our Galaxy, the thin and thick disks and the nuclear bulge are embedded in a spherical halo of very old, faint stars that extends to a distance of at least 150,000 light-years from the galactic center. Most of the globular clusters are also found in this halo.

The mass in the Milky Way extends even farther out, well beyond the boundary of the luminous stars to a distance of at least 200,000 light-years from the center of the Galaxy. This invisible mass has been give the name *dark matter* because it emits no light and cannot be seen with any telescope. Its composition is unknown, and it can be detected only because of its gravitational effects on the motions of luminous matter that we can see. We know that this extensive dark matter halo exists because of its effects on the orbits of distant star clusters and other dwarf galaxies that are associated with the Galaxy. This mysterious halo will be a subject of the section on The Mass of the Galaxy, and the properties of dark matter will be discussed more in the chapter on The Big Bang.

Some vital statistics of the thin and thick disks and the stellar halo are given in Table 25.1, with an illustration in Figure 25.8. Note particularly how the ages of stars correlate with where they are found. As we shall see, this information holds important clues to how the Milky Way Galaxy formed.

Table 25.1: Characteristics of the Milky Way Galaxy

Property	Thin Disk	Thick Disk	Stellar Halo (Excludes Dark Matter)
Stellar mass	$4 \times 10^{10}\ M_{\text{Sun}}$	A few percent of the thin disk mass	$10^{10}\ M_{\text{Sun}}$
Luminosity	$3 \times 10^{10}\ L_{\text{Sun}}$	A few percent of the thin disk luminosity	$8 \times 10^{8}\ L_{\text{Sun}}$
Typical age of stars	1 million to 10 billion years	11 billion years	13 billion years
Heavier-element abundance	High	Intermediate	Very low
Rotation	High	Intermediate	Very low

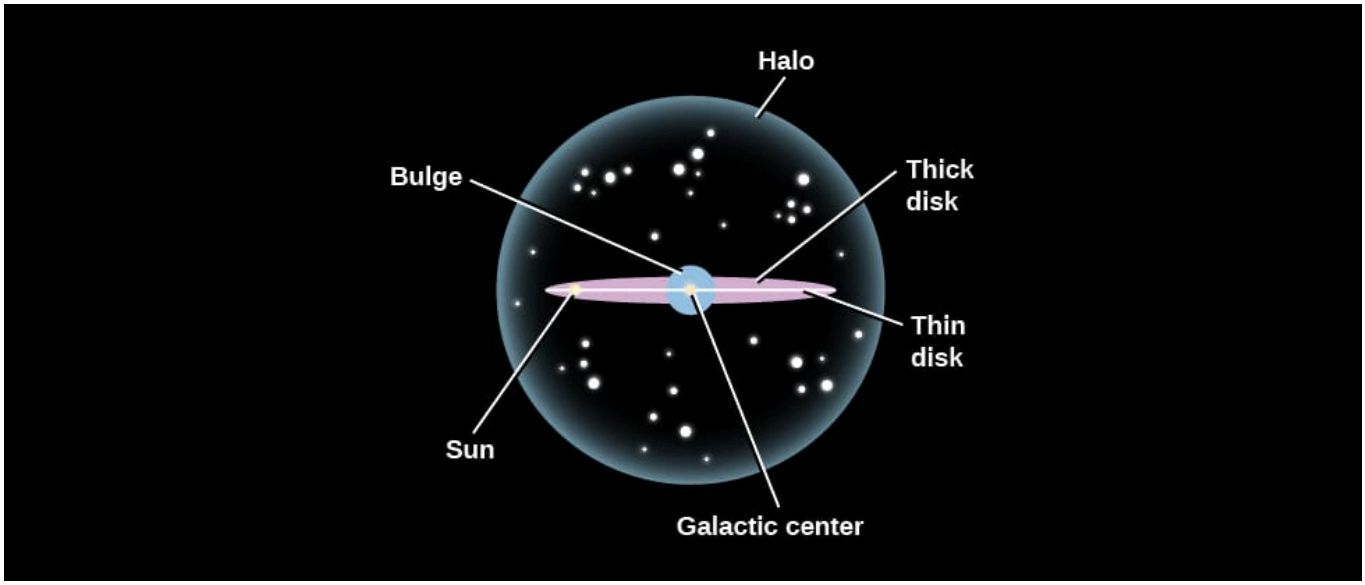


Figure 25.8 : Major Parts of the Milky Way Galaxy. This schematic shows the major components of our Galaxy.

Establishing this overall picture of the Galaxy from our dust-shrouded viewpoint inside the thin disk has been one of the great achievements of modern astronomy (and one that took decades of effort by astronomers working with a wide range of telescopes). One thing that helped enormously was the discovery that our Galaxy is not unique in its characteristics. There are many other flat, spiral-shaped islands of stars, gas, and dust in the universe. For example, the Milky Way somewhat resembles the Andromeda galaxy, which, at a distance of about 2.3 million light-years, is our nearest neighboring giant spiral galaxy. Just as you can get a much better picture of yourself if someone else takes the photo from a distance away, pictures and other diagnostic observations of nearby galaxies that resemble ours have been vital to our understanding of the properties of the Milky Way.

Link to Learning

Start this video, then click-and-drag to look in all directions from Earth in the Gaia space telescope's visualization of the stars of our galaxy.

The Milky Way viewer allows you to move to different locations in the Milky Way (using the Move and Zoom buttons) and view different stellar populations.

Making Connections: The Milky Way Galaxy in Myth and Legend

To most of us living in the twenty-first century, the Milky Way Galaxy is an elusive sight. We must make an effort to leave our well-lit homes and streets and venture beyond our cities and suburbs into less populated environments. Once the light pollution subsides to negligible levels, the Milky Way can be readily spotted arching over the sky on clear, moonless nights. The Milky Way is especially bright in late summer and early fall in the Northern Hemisphere. Some of the best places to view the Milky Way are in our national and state parks, where residential and industrial developments have been kept to a minimum. Some of these parks host special sky-gazing events that are definitely worth checking out—especially during the two weeks surrounding the new moon, when the faint stars and Milky Way don't have to compete with the Moon's brilliance.

Go back a few centuries, and these starlit sights would have been the norm rather than the exception. Before the advent of electric or even gas lighting, people relied on short-lived fires to illuminate their homes and byways. Consequently, their night skies were typically much darker. Confronted by myriad stellar patterns and the Milky Way's gauzy band of diffuse light, people of all cultures developed myths to make sense of it all.

Some of the oldest myths relating to the Milky Way are maintained by the aboriginal Australians through their rock painting and storytelling. These legacies are thought to go back tens of thousands of years, to when the aboriginal people were being "dreamed" along with the rest of the cosmos. The Milky Way played a central role as an arbiter of the Creation. Taking the form of a great serpent, it joined with the Earth serpent to dream and thus create all the creatures on Earth.

The ancient Greeks viewed the Milky Way as a spray of milk that spilled from the breast of the goddess Hera. In this legend, Zeus had secretly placed his infant son Heracles at Hera's breast while she was asleep in order to give his half-human son immortal powers. When Hera awoke and found Heracles suckling, she pushed him away, causing her milk to spray forth into the cosmos (Figure 25.9).

The dynastic Chinese regarded the Milky Way as a "silver river" that was made to separate two star-crossed lovers. To the east of the Milky Way, Zhi Nu, the weaving maiden, was identified with the bright star Vega in the constellation of Lyra the Harp. To the west of the Milky Way, her lover Niu Lang, the cowherd, was associated with the star Altair in the constellation of Aquila the Eagle. They had been exiled on opposite sides of the Milky Way by Zhi Nu's mother, the Queen of Heaven, after she heard of their secret marriage and the birth of their two children. However, once a year, they are permitted to reunite. On the seventh day of the seventh lunar month (which typically occurs in our month of August), they would meet on a bridge over the Milky Way that thousands of magpies had made (Figure 25.9). This romantic time continues to be celebrated today as Qi Xi, meaning "Double Seventh," with couples reenacting the cosmic reunion of Zhi Nu and Niu Lang.



(a)



(b)

Figure 25.9 : The Milky Way in Myth. (a) *Origin of the Milky Way* by Jacopo Tintoretto (circa 1575) illustrates the Greek myth that explains the formation of the Milky Way. (b) *The Moon of the Milky Way* by Japanese painter Tsukioka Yoshitoshi depicts the Chinese legend of Zhi Nu and Niu Lang.

To the Quechua Indians of Andean Peru, the Milky Way was seen as the celestial abode for all sorts of cosmic creatures. Arrayed along the Milky Way are myriad dark patches that they identified with partridges, llamas, a toad, a snake, a fox, and other animals. The Quechua's orientation toward the dark regions rather than the glowing band of starlight appears to be unique among all the myth makers. Likely, their access to the richly structured southern Milky Way had something to do with it.

Among Finns, Estonians, and related northern European cultures, the Milky Way is regarded as the "pathway of birds" across the night sky. Having noted that birds seasonally migrate along a north-south route, they identified this byway with the Milky Way. Recent scientific studies have shown that this myth is rooted in fact: the birds of this region use the Milky Way as a guide for their annual migrations.

Today, we regard the Milky Way as our galactic abode, where the foment of star birth and star death plays out on a grand stage, and where sundry planets have been found to be orbiting all sorts of stars. Although our perspective on the Milky Way is based on scientific investigations, we share with our forebears an affinity for telling stories of origin and transformation. In these regards, the Milky Way continues to fascinate and inspire us.

This page titled 15.1: The Architecture of the Galaxy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 25.2: The Architecture of the Galaxy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.2: The Interstellar Medium

Learning Objectives

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

- Explain how much interstellar matter there is in the Milky Way, and what its typical density is
- Describe how the interstellar medium is divided into gaseous and solid components

Astronomers refer to all the material between stars as *interstellar* matter; the entire collection of interstellar matter is called the interstellar medium (ISM). Some interstellar material is concentrated into giant clouds, each of which is known as a nebula (plural "nebulae," Latin for "clouds"). The best-known nebulae are the ones that we can see glowing or reflecting visible light; there are many pictures of these in this chapter.

Interstellar clouds do not last for the lifetime of the universe. Instead, they are like clouds on Earth, constantly shifting, merging with each other, growing, or dispersing. Some become dense and massive enough to collapse under their own gravity, forming new stars. When stars die, they, in turn, eject some of their material into interstellar space. This material can then form new clouds and begin the cycle over again.

About 99% of the material between the stars is in the form of a *gas*—that is, it consists of individual atoms or molecules. The most abundant elements in this gas are hydrogen and helium (which we saw are also the most abundant elements in the stars), but the gas also includes other elements. Some of the gas is in the form of molecules—combinations of atoms. The remaining 1% of the interstellar material is solid—frozen particles consisting of many atoms and molecules that are called *interstellar grains* or interstellar dust (Figure 20.2). A typical dust grain consists of a core of rocklike material (silicates) or graphite surrounded by a mantle of ices; water, methane, and ammonia are probably the most abundant ices.



Figure 20.2 : Various Types of Interstellar Matter. The reddish nebulae in this spectacular photograph glow with light emitted by hydrogen atoms. The darkest areas are clouds of dust that block the light from stars behind them. The upper part of the picture is filled with the bluish glow of light reflected from hot stars embedded in the outskirts of a huge, cool cloud of dust and gas. The cool supergiant star Antares can be seen as a big, reddish patch in the lower-left part of the picture. The star is shedding some of its outer atmosphere and is surrounded by a cloud of its own making that reflects the red light of the star. The red nebula in the middle right partially surrounds the star Sigma Scorpii. (To the right of Antares, you can see M4, a much more distant cluster of extremely old stars.) (credit: modification of work by ESO/Digitized Sky Survey 2)

If all the interstellar gas within the Galaxy were spread out smoothly, there would be only about one atom of gas per cm^3 in interstellar space. (In contrast, the air in the room where you are reading this book has roughly 10^{19} atoms per cm^3.) The dust

grains are even scarcer. A km^3 of space would contain only a few hundred to a few thousand tiny grains, each typically less than one ten-thousandth of a millimeter in diameter. These numbers are just averages, however, because the gas and dust are distributed in a patchy and irregular way, much as water vapor in Earth's atmosphere is often concentrated into clouds.

In some interstellar clouds, the density of gas and dust may exceed the average by as much as a thousand times or more, but even this density is more nearly a vacuum than any we can make on Earth. To show what we mean, let's imagine a vertical tube of air reaching from the ground to the top of Earth's atmosphere with a cross-section of 1 square meter. Now let us extend the same-size tube from the top of the atmosphere all the way to the edge of the observable universe—over 10 billion light-years away. Long though it is, the second tube would still contain fewer atoms than the one in our planet's atmosphere.

While the *density* of interstellar matter is very low, the volume of space in which such matter is found is huge, and so its *total mass* is substantial. To see why, we must bear in mind that stars occupy only a tiny fraction of the volume of the Milky Way Galaxy. For example, it takes light only about four seconds to travel a distance equal to the diameter of the Sun, but more than four *years* to travel from the Sun to the nearest star. Even though the spaces among the stars are sparsely populated, there's a lot of space out there!

Astronomers estimate that the total mass of gas and dust in the Milky Way Galaxy is equal to about 15% of the mass contained in stars. This means that the mass of the interstellar matter in our Galaxy amounts to about 10 billion times the mass of the Sun. There is plenty of raw material in the Galaxy to make generations of new stars and planets (and perhaps even astronomy students).

Example 20.1: Estimating Interstellar Mass

You can make a rough estimate of how much interstellar mass our Galaxy contains and also how many new stars could be made from this interstellar matter. All you need to know is how big the Galaxy is and the average density using this formula:

$$\text{total mass} = \text{volume} \times \text{density of atoms} \times \text{mass per}$$

You have to remember to use consistent units—such as meters and kilograms. We will assume that our Galaxy is shaped like a cylinder; the volume of a cylinder equals the area of its base times its height

$$V =$$

where R is the radius of the cylinder and h is its height.

Suppose that the average density of hydrogen gas in our Galaxy is one atom per cm^3. Each hydrogen atom has a mass of 1.7×10^{-27} kg. If the Galaxy is a cylinder with a diameter of 100,000 light-years and a height of 300 light-years, what is the mass of this gas? How many solar-mass stars (2.0×10^{30} kg) could be produced from this mass of gas if it were all turned into stars?

Solution

If the diameter of the Galaxy is 100,000 light-years, then the radius is 50,000 light-years. Recall that 1 light-year = 9.5×10^{12} km = 9.5×10^{17} cm, so the volume of the Galaxy is

$$V = \pi R^2 h = \pi(50{,}000 \times 9.5 \times 10^{17}\,\text{cm})^2\,(300 \times 9.5 \times 10^{17}\,\text{cm}) = 2.0 \times 10^{66}\,\text{cm}^3$$

The total mass is therefore

$$M = V \times \text{density of atoms} \times \text{mass per atom}$$

$$2.0 \times 10^{66}\,\text{cm}^3 \times (1\,\text{atom/cm}^3) \times 1.7 \times 10^{-27}\,\text{kg} = 3.5 \times 10^{39}\,\text{kg}$$

This is sufficient to make

$$N = \frac{M}{(2.0 \times 10^{30}\,\text{kg})} = 1.75 \times 10^9$$

stars equal in mass to the Sun. That's roughly 2 billion stars.

Exercise 15.2.1

You can use the same method to estimate the mass of interstellar gas around the Sun. The distance from the Sun to the nearest other star, Proxima Centauri, is 4.2 light-years. We will see in Interstellar Matter around the Sun that the gas in the immediate vicinity of the Sun is less dense than average, about 0.1 atoms per cm^3. What is the total mass of interstellar hydrogen in a sphere centered on the Sun and extending out to Proxima Centauri? How does this compare to the mass of the Sun? It is helpful to remember that the volume of a sphere is related to its radius:

$$V = (4/3)\pi R^3$$

Answer available at the end of this section.

Astronomy Basics: Naming the Nebulae

As you look at the captions for some of the spectacular photographs in this chapter and The Birth of Stars and the Discovery of Planets outside the Solar System, you will notice the variety of names given to the nebulae. A few, which in small telescopes look like something recognizable, are sometimes named after the creatures or objects they resemble. Examples include the Crab, Tarantula, and Keyhole Nebulae. But most have only numbers that are entries in a catalog of astronomical objects.

Perhaps the best-known catalog of nebulae (as well as star clusters and galaxies) was compiled by the French astronomer Charles Messier (1730–1817). Messier's passion was discovering comets, and his devotion to this cause earned him the nickname "The Comet Ferret" from King Louis XV. When comets are first seen coming toward the Sun, they look like little fuzzy patches of light; in small telescopes, they are easy to confuse with nebulae or with groupings of many stars so far away that their light is all blended together. Time and again, Messier's heart leapt as he thought he had discovered one of his treasured comets, only to find that he had "merely" observed a nebula or cluster.

In frustration, Messier set out to catalog the position and appearance of over 100 objects that could be mistaken for comets. For him, this list was merely a tool in the far more important work of comet hunting. He would be very surprised if he returned today to discover that no one recalls his comets anymore, but his catalog of "fuzzy things that are not comets" is still widely used. When Figure 20.2 refers to M4, it denotes the fourth entry in Messier's list. Visit https://www.nasa.gov/content/goddard/hubble-s-messier-catalog for a gallery of M objects as photographed with the Hubble Space Telescope.

A far more extensive listing was compiled under the title of the *New General Catalog* (*NGC*) *of Nebulae and Star Clusters* in 1888 by John Dreyer, working at the observatory in Armagh, Ireland. He based his compilation on the work of William Herschel and his son John, plus many other observers who followed them. With the addition of two further listings (called the *Index Catalogs*), Dreyer's compilation eventually included 13,000 objects. Astronomers today still use his NGC numbers when referring to most nebulae and star groups.

Answers

Exercise 15.2.1

The volume of a sphere stretching from the Sun to Proxima Centauri is:
$V = (4/3)\,\pi R^3 = (4/3)\,\pi\left(4.2\,\times\,9.5\,\times\,10^{17}\,\mathrm{cm}\right)^3 = 2.7\,\times\,10^{56}\,\mathrm{cm}^3$
Therefore, the mass of hydrogen in this sphere is:
$M = V\,\times\,\left(0.1\,\mathrm{atom/cm}^3\right)\,\times\,1.7\,\times\,10^{-27}\,\mathrm{kg} = 4.5\,\times\,10^{28}\,\mathrm{kg}$
This is only (4.5×10^{28} kg)/(2.0×10^{30} kg) = 2.2% the mass of the Sun.

This page titled 15.2: The Interstellar Medium is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 20.2: The Interstellar Medium by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.3: Interstellar Gas

Learning Objectives

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

- Name the major types of interstellar gas
- Discuss how we can observe each type
- Describe the temperature and other major properties of each type

Interstellar gas, depending on where it is located, can be as cold as a few degrees above absolute zero or as hot as a million degrees or more. We will begin our voyage through the interstellar medium by exploring the different conditions under which we find gas.

Ionized Hydrogen (H II) Regions—Gas Near Hot Stars

Some of the most spectacular astronomical photographs show interstellar gas located near hot stars (Figure 20.3). The strongest line in the visible region of the hydrogen spectrum is the red line in the Balmer series[1] (as explained in the chapter on Radiation and Spectra); this emission line accounts for the characteristic red glow in images like Figure 20.3.



Figure 20.3 : Orion Nebula. The red glow that pervades the great Orion Nebula is produced by the first line in the Balmer series of hydrogen. Hydrogen emission indicates that there are hot young stars nearby that ionize these clouds of gas. When electrons then recombine with protons and move back down into lower energy orbits, emission lines are produced. The blue color seen at the edges of some of the clouds is produced by small particles of dust that scatter the light from the hot stars. Dust can also be seen silhouetted against the glowing gas. (credit: NASA,ESA, M. Robberto (Space Telescope Science Institute/ESA) and the Hubble Space Telescope Orion Treasury Project Team)

Hot stars are able to heat nearby gas to temperatures close to 10,000 K. The ultraviolet radiation from the stars also ionizes the hydrogen (remember that during ionization, the electron is stripped completely away from the proton). Such a detached proton won't remain alone forever when attractive electrons are around; it will capture a free electron, becoming a neutral hydrogen once more. However, such a neutral atom can then absorb ultraviolet radiation again and start the cycle over. At a typical moment, most of the atoms near a hot star are in the ionized state.

Since hydrogen is the main constituent of interstellar gas, we often characterize a region of space according to whether its hydrogen is neutral or ionized. A cloud of ionized hydrogen is called an H II region. (Scientists who work with spectra use the Roman numeral I to indicate that an atom is neutral; successively higher Roman numerals are used for each higher stage of

ionization. H II thus refers to hydrogen that has lost its one electron; Fe III is iron with two electrons missing.)

The electrons that are captured by the hydrogen nuclei cascade down through the various energy levels of the hydrogen atoms on their way to the lowest level, or ground state. During each transition downward, they give up energy in the form of light. The process of converting ultraviolet radiation into visible light is called *fluorescence*. Interstellar gas contains other elements besides hydrogen. Many of them are also ionized in the vicinity of hot stars; they then capture electrons and emit light, just as hydrogen does, allowing them to be observed by astronomers. But generally, the red hydrogen line is the strongest, and that is why H II regions look red.

A fluorescent light on Earth works using the same principles as a fluorescent H II region. When you turn on the current, electrons collide with atoms of mercury vapor in the tube. The mercury is excited to a high-energy state because of these collisions. When the electrons in the mercury atoms return to lower-energy levels, some of the energy they emit is in the form of ultraviolet photons. These, in turn, strike a phosphor-coated screen on the inner wall of the light tube. The atoms in the screen absorb the ultraviolet photons and emit visible light as they cascade downward among the energy levels. (The difference is that these atoms give off a wider range of light colors, which mix to give the characteristic white glow of fluorescent lights, whereas the hydrogen atoms in an H II region give off a more limited set of colors.)

Neutral Hydrogen Clouds

The very hot stars required to produce H II regions are rare, and only a small fraction of interstellar matter is close enough to such hot stars to be ionized by them. Most of the volume of the interstellar medium is filled with neutral (nonionized) hydrogen. How do we go about looking for it?

Unfortunately, neutral hydrogen atoms at temperatures typical of the gas in interstellar space neither emit nor absorb light in the visible part of the spectrum. Nor, for the most part, do the other trace elements that are mixed with the interstellar hydrogen. However, some of these other elements can *absorb* visible light even at typical interstellar temperatures. This means that when we observe a bright source such as a hot star or a galaxy, we can sometimes see additional lines in its spectrum produced when interstellar gas absorbs light at particular frequencies (see Figure 20.4). Some of the strongest interstellar absorption lines are produced by calcium and sodium, but many other elements can be detected as well in sufficiently sensitive observations (as discussed in Radiation and Spectra).

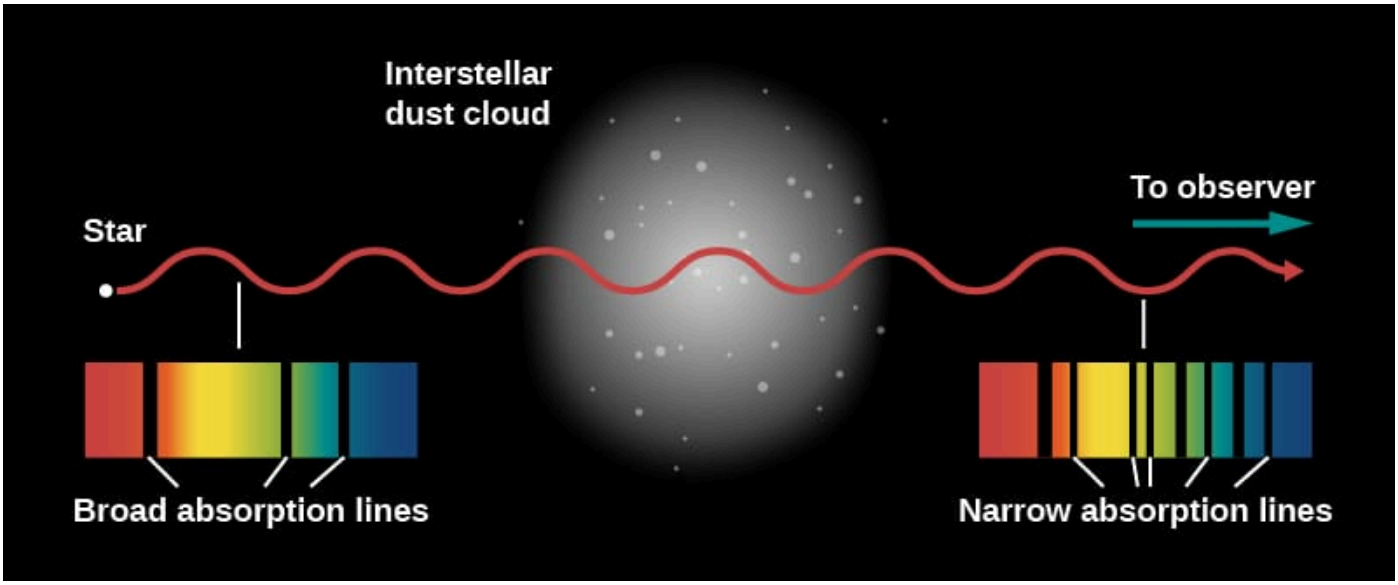


Figure 20.4 : Absorption Lines though an Interstellar Dust Cloud. When there is a significant amount of cool interstellar matter (gas with some dust) between us and a star, we can see the absorption lines of the gas in the star's spectrum. We can distinguish the two kinds of lines because, whereas the star's lines are broad, the lines from the gas are narrower.

The first evidence for absorption by interstellar clouds came from the analysis of a spectroscopic binary star (see The Stars: A Celestial Census), published in 1904. While most of the lines in the spectrum of this binary shifted alternately from longer to shorter wavelengths and back again, as we would expect from the Doppler effect for stars in orbit around each other, a few lines in the spectrum remained fixed in wavelength. Since both stars are moving in a binary system, lines that showed no motion puzzled

astronomers. The lines were also peculiar in that they were much, much narrower than the rest of the lines, indicating that the gas producing them was at a very low pressure. Subsequent work demonstrated that these lines were not formed in the star's atmosphere at all, but rather in a cold cloud of gas located between Earth and the binary star.

While these and similar observations proved there was interstellar gas, they could not yet detect hydrogen, the most common element, due to its lack of spectral features in the visible part of the spectrum. (The Balmer line of hydrogen is in the visible range, but only excited hydrogen atoms produce it. In the cold interstellar medium, the hydrogen atoms are all in the ground state and no electrons are in the higher-energy levels required to produce either emission or absorption lines in the Balmer series.) Direct detection of hydrogen had to await the development of telescopes capable of seeing very-low-energy changes in hydrogen atoms in other parts of the spectrum. The first such observations were made using radio telescopes, and radio emission and absorption by interstellar hydrogen remains one of our main tools for studying the vast amounts of cold hydrogen in the universe to this day.

In 1944, while he was still a student, the Dutch astronomer Hendrik van de Hulst predicted that hydrogen would produce a strong line at a wavelength of 21 centimeters. That's quite a long wavelength, implying that the wave has such a low frequency and low energy that it cannot come from electrons jumping between energy levels (as we discussed in Radiation and Spectra). Instead, energy is emitted when the electron does a flip, something like an acrobat in a circus flipping upright after standing on his head.

The flip works like this: a hydrogen atom consists of a proton and an electron bound together. Both the proton and the electron act is if they were spinning like tops, and spin axes of the two tops can either be pointed in the same direction (aligned) or in opposite directions (anti-aligned). If the proton and electron were spinning in opposite directions, the atom as a whole would have a very slightly lower energy than if the two spins were aligned (Figure 20.5). If an atom in the lower-energy state (spins opposed) acquired a small amount of energy, then the spins of the proton and electron could be aligned, leaving the atom in a slightly *excited state*. If the atom then lost that same amount of energy again, it would return to its ground state. The amount of energy involved corresponds to a wave with a wavelength of 21 centimeters; hence, it is known as the *21-centimeter line*.

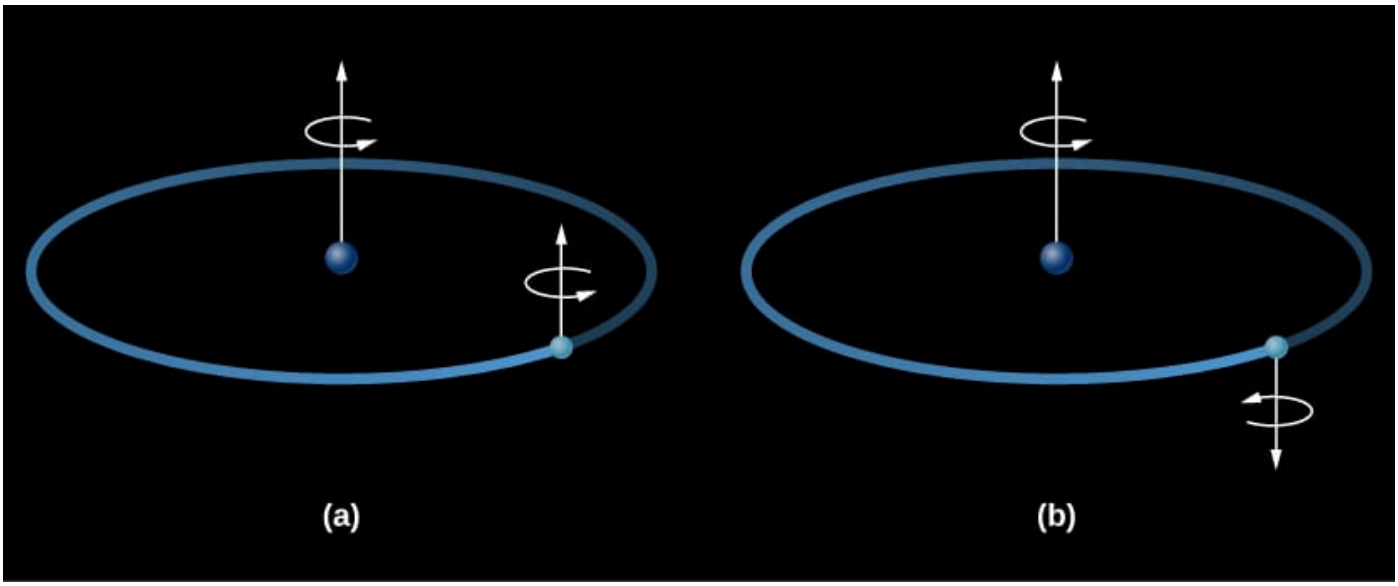


Figure 20.5 : Formation of the 21-Centimeter Line. When the electron in a hydrogen atom is in the orbit closest to the nucleus, the proton and the electron may be spinning either (a) in the same direction or (b) in opposite directions. When the electron flips over, the atom gains or loses a tiny bit of energy by either absorbing or emitting electromagnetic energy with a wavelength of 21 centimeters.

Neutral hydrogen atoms can acquire small amounts of energy through collisions with other hydrogen atoms or with free electrons. Such collisions are extremely rare in the sparse gases of interstellar space. An individual atom may wait centuries before such an encounter aligns the spins of its proton and electron. Nevertheless, over many millions of years, a significant fraction of the hydrogen atoms are excited by a collision. (Out there in cold space, that's about as much excitement as an atom typically experiences.)

An excited atom can later lose its excess energy either by colliding with another particle or by giving off a radio wave with a wavelength of 21 centimeters. If there are no collisions, an excited hydrogen atom will wait an average of about 10 million years before emitting a photon and returning to its state of lowest energy. Even though the probability that any single atom will emit a photon is low, there are so many hydrogen atoms in a typical gas cloud that collectively they will produce an observable line at 21 centimeters.

Equipment sensitive enough to detect the 21-cm line of neutral hydrogen became available in 1951. Dutch astronomers had built an instrument to detect the 21-cm waves that they had predicted, but a fire destroyed it. As a result, two Harvard physicists, Harold Ewen and Edward Purcell, made the first detection (Figure 20.6), soon followed by confirmations from the Dutch and a group in Australia. Since the detection of the 21-cm line, many other radio lines produced by both atoms and molecules have been discovered (as we will discuss in a moment), and these have allowed astronomers to map out the neutral gas throughout our home Galaxy. Astronomers have also detected neutral interstellar gas, including hydrogen, at many other wavelengths from the infrared to the ultraviolet.

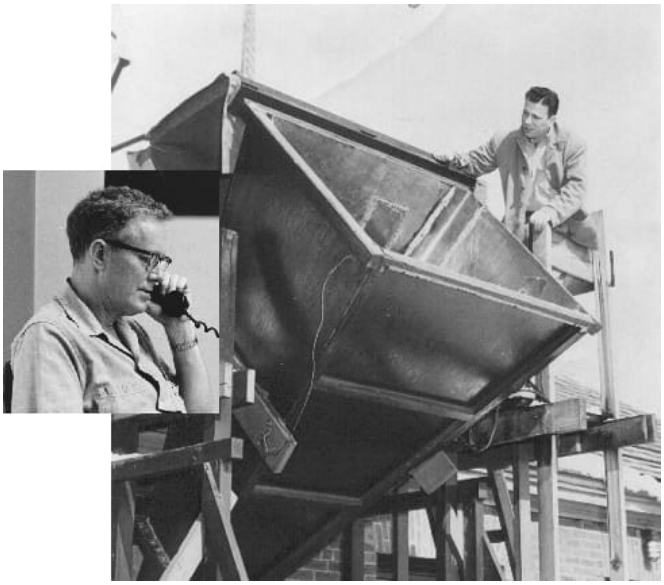


Figure 20.6 : Harold Ewen (1922–2015) and Edward Purcell (1912–1997). We see Harold Ewen in 1952 working with the horn antenna (atop the physics laboratory at Harvard) that made the first detection of interstellar 21-cm radiation. The inset shows Edward Purcell, the winner of the 1952 Nobel Prize in physics, a few years later. (credit: modification of work by NRAO)

Modern radio observations show that most of the neutral hydrogen in our Galaxy is confined to an extremely flat layer, less than 300 light-years thick, that extends throughout the disk of the Milky Way Galaxy. This gas has densities ranging from about 0.1 to about 100 atoms per cm^3, and it exists at a wide range of temperatures, from as low as about 100 K (–173 °C) to as high as about 8000 K. These regions of warm and cold gas are interspersed with each other, and the density and temperature at any particular point in space are constantly changing.

Ultra-Hot Interstellar Gas

While the temperatures of 10,000 K found in H II regions might seem warm, they are not the hottest phase of the interstellar medium. Some of the interstellar gas is at a temperature of a *million* degrees, even though there is no visible source of heat nearby. The discovery of this ultra-hot interstellar gas was a big surprise. Before the launch of astronomical observatories into space, which could see radiation in the ultraviolet and X-ray parts of the spectrum, astronomers assumed that most of the region between stars was filled with hydrogen at temperatures no warmer than those found in H II regions. But telescopes launched above Earth's atmosphere obtained ultraviolet spectra that contained interstellar lines produced by oxygen atoms that have been ionized five times. To strip five electrons from their orbits around an oxygen nucleus requires a lot of energy. Subsequent observations with orbiting X-ray telescopes revealed that the Galaxy is filled with numerous bubbles of X-ray-emitting gas. To emit X-rays, and to contain oxygen atoms that have been ionized five times, gas must be heated to temperatures of a million degrees or more.

Theorists have now shown that the source of energy producing these remarkable temperatures is the explosion of massive stars at the ends of their lives (Figure 20.7). Such explosions, called *supernovae*, will be discussed in detail in the chapter on The Death of Stars. For now, we'll just say that some stars, nearing the ends of their lives, become unstable and literally explode. These explosions launch gas into interstellar space at velocities of tens of thousands of kilometers per second (up to about 30% the speed of light). When this ejected gas collides with interstellar gas, it produces shocks that heat the gas to millions or tens of millions of degrees.



Figure 20.7 : Vela Supernova Remnant. About 11,000 years ago, a dying star in the constellation of Vela exploded, becoming as bright as the full moon in Earth's skies. You can see the faint rounded filaments from that explosion in the center of this colorful image. The edges of the remnant are colliding with the interstellar medium, heating the gas they plow through to temperatures of millions of K. Telescopes in space also reveal a glowing sphere of X-ray radiation from the remnant. (credit: Digitized Sky Survey, ESA/ESO/NASA FITS Liberator, Davide De Martin)

Astronomers estimate that one supernova explodes roughly every 100 years somewhere in the Galaxy. On average, shocks launched by supernovae sweep through any given point in the Galaxy about once every few million years. These shocks keep some interstellar space filled with gas at temperatures of millions of degrees, and they continually disturb the colder gas, keeping it in constant, turbulent motion.

Molecular Clouds

A few simple molecules out in space, such as CN and CH, were discovered decades ago because they produce absorption lines in the visible-light spectra of stars behind them. When more sophisticated equipment for obtaining spectra in radio and infrared wavelengths became available, astronomers—to their surprise—found much more complex molecules in interstellar clouds as well.

Just as atoms leave their "fingerprints" in the spectrum of visible light, so the vibration and rotation of atoms within molecules can leave spectral fingerprints in radio and infrared waves. If we spread out the radiation at such longer wavelengths, we can detect emission or absorption lines in the spectra that are characteristic of specific molecules. Over the years, experiments in our laboratories have shown us the exact wavelengths associated with changes in the rotation and vibration of many common molecules, giving us a template of possible lines against which we can now compare our observations of interstellar matter.

The discovery of complex molecules in space came as a surprise because most of interstellar space is filled with ultraviolet light from stars, and this light is capable of *dissociating* molecules (breaking them apart into individual atoms). In retrospect, however, the presence of molecules is not surprising. As we will discuss further in the next section, and have already seen above, interstellar space also contains significant amounts of dust capable of blocking out starlight. When this dust accumulates in a single location, the result is a dark cloud where ultraviolet starlight is blocked and molecules can survive. The largest of these structures are created where gravity pulls interstellar gas together to form giant molecular clouds, structures as massive as a million times the mass of the Sun. Within these, most of the interstellar hydrogen has formed the molecule H_2 (molecular hydrogen). Other, more complex molecules are also present in much smaller quantities.

Giant molecular clouds have densities of hundreds to thousands of atoms per cm^3, much denser than interstellar space is on average. As a result, though they account for a very small fraction of the volume of interstellar space, they contain a significant fraction—20–30%—of the total mass of the Milky Way's gas. Because of their high density, molecular clouds block ultraviolet starlight, the main agent for heating most interstellar gas. As a result, they tend to be extremely cold, with typical temperatures near 10 K (–263 °C). Giant molecular clouds are also the sites where new stars form, as we will discuss below.

It is in these dark regions of space, protected from starlight, that molecules can form. Chemical reactions occurring both in

the gas and on the surface of dust grains lead to much more complex compounds, hundreds of which have been identified in interstellar space. Among the simplest of these are water (H_2O), carbon monoxide (CO), which is produced by fires on Earth, and ammonia (NH_3), whose smell you recognize in strong home cleaning products. Carbon monoxide is particularly abundant in interstellar space and is the primary tool that astronomers use to study giant molecular clouds. Unfortunately, the most abundant molecule, H_2, is particularly difficult to observe directly because in most giant molecular clouds, it is too cold to emit even at radio wavelengths. CO, which tends to be present wherever H_2 is found, is a much better emitter and is often used by astronomers to trace molecular hydrogen.

The more complex molecules astronomers have found are mostly combinations of hydrogen, oxygen, carbon, nitrogen, and sulfur atoms. Many of these molecules are *organic* (those that contain carbon and are associated with the carbon chemistry of life on Earth.) They include formaldehyde (used to preserve living tissues), alcohol (see the feature box on Cocktails in Space), and antifreeze.

In 1996, astronomers discovered acetic acid (the prime ingredient of vinegar) in a cloud lying in the direction of the constellation of Sagittarius. To balance the sour with the sweet, a simple sugar (glycolaldehyde) has also been found. The largest compounds yet discovered in interstellar space are *fullerene*s, molecules in which 60 or 70 carbon atoms are arranged in a cage-like configuration (see Figure 20.8). See Table 20.1 for a list of a few of the more interesting interstellar molecules that have been found so far.

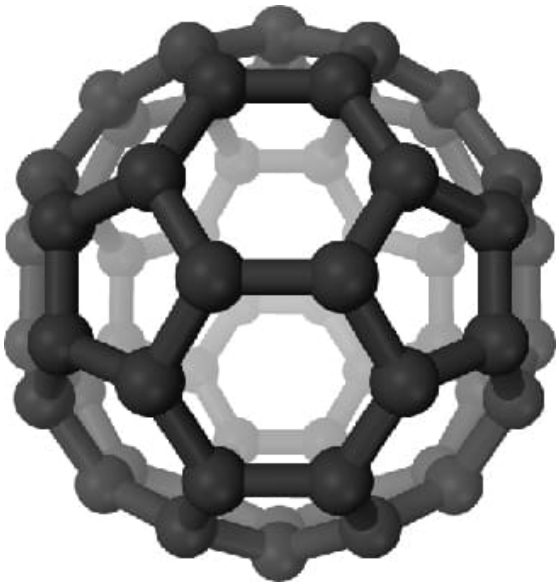


Figure 20.8 : Fullerene C60. This three-dimensional perspective shows the characteristic cage-like arrangement of the 60 carbon atoms in a molecule of fullerene C60. Fullerene C60 is also known as a “buckyball,” or as its full name, buckminsterfullerene, because of its similarity to the multisided architectural domes designed by American inventor R. Buckminster Fuller.

Some Interesting Interstellar Molecules		
Name	**Chemical Formula**	**Use on Earth**
Ammonia	NH_3	Household cleansers
Formaldehyde	H_2CO	Embalming fluid
Acetylene	HC_2H	Fuel for a welding torch
Acetic acid	$C_2H_2O_4$	The essence of vinegar
Ethyl alcohol	CH_3CH_2OH	End-of-semester parties
Ethylene glycol	$HOCH_2CH_2OH$	Antifreeze ingredient
Benzene	C_6H_6	Carbon ring, ingredient in varnishes and dyes

Table 20.1

The cold interstellar clouds also contain cyanoacetylene (HC_3N) and acetaldehyde (CH_3CHO), generally regarded as starting points for *amino acid* formation. These are building blocks of proteins, which are among the fundamental chemicals from which living organisms on Earth are constructed. The presence of these organic molecules does not imply that life exists in space, but it does show that the chemical building blocks of life can form under a wide range of conditions in the universe. As we learn more about how complex molecules are produced in interstellar clouds, we gain an increased understanding of the kinds of processes that preceded the beginnings of life on Earth billions of years ago.

Link to Learning

Interested in learning more about fullerenes, buckyballs, or buckminsterfullerenes (as they're called)? Watch a brief video from NASA's Jet Propulsion Laboratory that explains what they are and illustrates how they were discovered in space.

Making Connections: Cocktails in Space

Among the molecules astronomers have identified in interstellar clouds is alcohol, which comes in two varieties: methyl (or wood) alcohol and ethyl alcohol (the kind you find in cocktails). Ethyl alcohol is a pretty complex molecule, written by chemists as C_2H_5OH. It is quite plentiful in space (relatively speaking). In clouds where it has been identified, we detect up to one molecule for every m^3. The largest of the clouds (which can be several hundred light-years across) have enough ethyl alcohol to make 10^{28} fifths of liquor.

We need not fear, however, that future interstellar astronauts will become interstellar alcoholics. Even if a spaceship were equipped with a giant funnel 1 kilometer across and could scoop it through such a cloud at the speed of light, it would take about a thousand years to gather up enough alcohol for one standard martini.

Furthermore, the very same clouds also contain water (H_2O) molecules. Your scoop would gather them up as well, and there are a lot more of them because they are simpler and thus easier to form. For the fun of it, one astronomical paper actually calculated the proof of a typical cloud. *Proof* is the ratio of alcohol to water in a drink, where 0 proof means all water, 100 proof means half alcohol and half water, and 200 proof means all alcohol. The proof of the interstellar cloud was only 0.2, not enough to qualify as a stiff drink

Footnotes

- 1Scientists also call this red Balmer line the H-alpha line, with alpha meaning it is the first spectral line in the Balmer series.

This page titled 15.3: Interstellar Gas is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 20.3: Interstellar Gas by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.4: Cosmic Dust

Learning Objectives

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

- Describe how we can detect interstellar dust
- Understand the role and importance of infrared observations in studying dust
- Explain the terms extinction and interstellar reddening

Figure 20.9 shows a striking example of what is actually a common sight through large telescopes: a dark region on the sky that appears to be nearly empty of stars. For a long time, astronomers debated whether these dark regions were empty "tunnels" through which we looked beyond the stars of the Milky Way Galaxy into intergalactic space, or clouds of some dark material that blocked the light of the stars beyond. The astronomer William Herschel (discoverer of the planet Uranus) thought it was the former, once remarking after seeing one, "Here truly is a hole in heaven!" However, American astronomer E. E. Barnard is generally credited with showing from his extensive series of nebula photographs that the latter interpretation is the correct one (see the feature box on Edward Emerson Barnard).

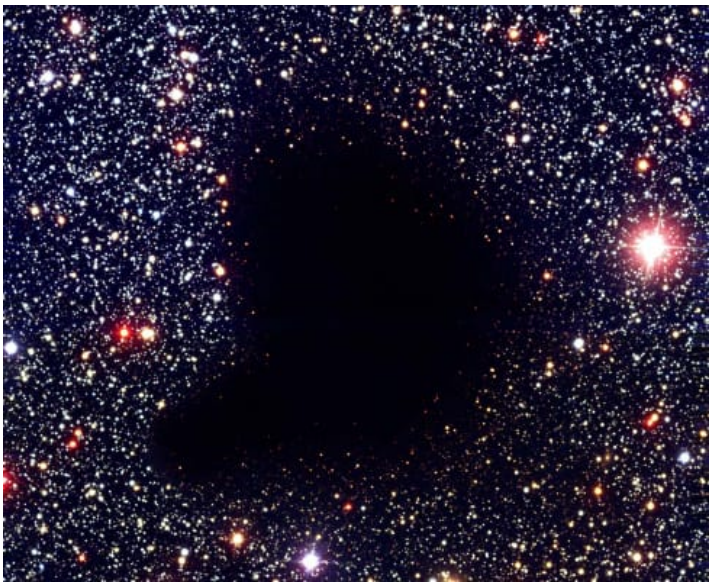


Figure 20.9 : Barnard 68. This object, first catalogued by E. E. Barnard, is a dark interstellar cloud. Its striking appearance is due to the fact that, since it is relatively close to Earth, there are no bright stars between us and it, and its dust obscures the light from the stars behind it. (It looks a little bit like a sideways heart; one astronomers sent a photo of this object to his sweetheart as a valentine.) (credit: modification of work by ESO)

Dusty clouds in space betray their presence in several ways: by blocking the light from distant stars, by emitting energy in the infrared part of the spectrum, by reflecting the light from nearby stars, and by making distant stars look redder than they really are.

Voyagers in Astronomy: Edward Emerson Barnard

Born in 1857 in Nashville, Tennessee, two months after his father died, Edward Barnard (Figure 20.10) grew up in such poor circumstances that he had to drop out of school at age nine to help support his ailing mother. He soon became an assistant to a local photographer, where he learned to love both photography and astronomy, destined to become the dual passions of his life. He worked as a photographer's aide for 17 years, studying astronomy on his own. In 1883, he obtained a job as an assistant at the Vanderbilt University Observatory, which enabled him at last to take some astronomy courses.

Married in 1881, Barnard built a house for his family that he could ill afford. But as it happened, a patent medicine manufacturer offered a $200 prize (a lot of money in those days) for the discovery of any new comet. With the determination that became characteristic of him, Barnard spent every clear night searching for comets. He discovered seven of them between 1881 and 1887, earning enough money to make the payments on his home; this "Comet House" later became a local attraction. (By the end of his life, Barnard had found 17 comets through diligent observation.)

In 1887, Barnard got a position at the newly founded Lick Observatory, where he soon locked horns with the director, Edward Holden, a blustering administrator who made Barnard's life miserable. (To be fair, Barnard soon tried to do the same for him.) Despite being denied the telescope time that he needed for his photographic work, in 1892, Barnard managed to discover the first new moon found around Jupiter since Galileo's day, a stunning observational feat that earned him world renown. Now in a position to demand more telescope time, he perfected his photographic techniques and soon began to publish the best images of the Milky Way taken up to that time. It was during the course of this work that he began to examine the dark regions among the crowded star lanes of the Galaxy and to realize that they must be vast clouds of obscuring material (rather than "holes" in the distribution of stars).

Astronomer-historian Donald Osterbrock has called Barnard an "observaholic:" his daily mood seemed to depend entirely on how clear the sky promised to be for his night of observing. He was a driven, neurotic man, concerned about his lack of formal training, fearful of being scorned, and afraid that he might somehow slip back into the poverty of his younger days. He had difficulty taking vacations and lived for his work: only serious illness could deter him from making astronomical observations.

In 1895, Barnard, having had enough of the political battles at Lick, accepted a job at the Yerkes Observatory near Chicago, where he remained until his death in 1923. He continued his photographic work, publishing compilations of his images that became classic photographic atlases, and investigating the varieties of nebulae revealed in his photographs. He also made measurements of the sizes and features of planets, participated in observations of solar eclipses, and carefully cataloged dark nebulae (see Figure 20.9). In 1916, he discovered the star with the largest proper motion, the second-closest star system to our own (see Analyzing Starlight). It is now called Barnard's Star in his honor.



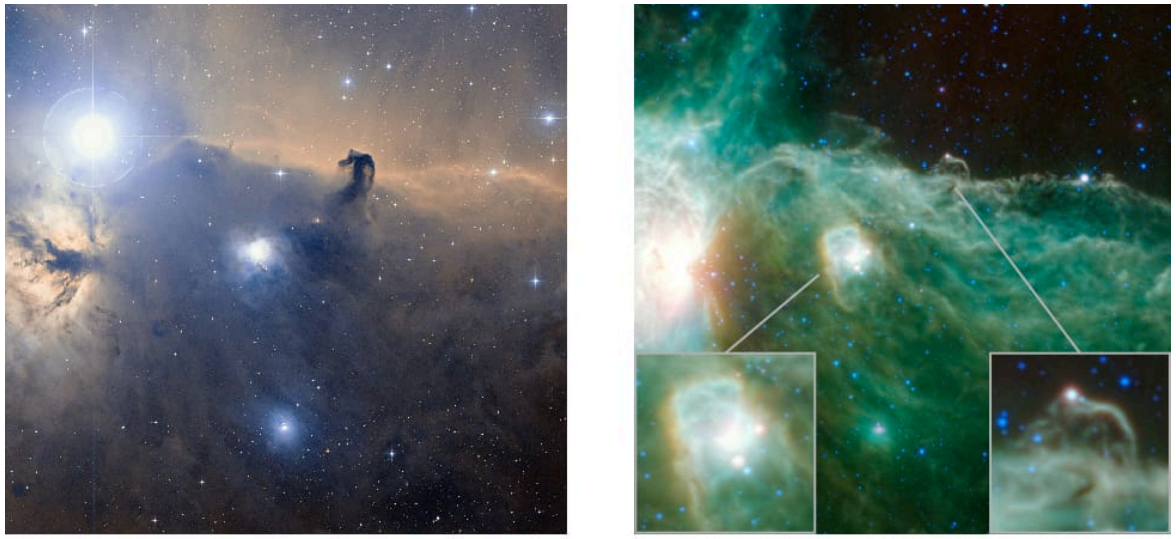
Figure 20.10 : Edward Emerson Barnard (1857–1923). Barnard's observations provided information that furthered many astronomical explorations. (credit: The Lick Observatory)

Detecting Dust

The dark cloud seen in Figure 20.9 blocks the light of the many stars that lie behind it; note how the regions in other parts of the photograph are crowded with stars. Barnard 68 is an example of a relatively dense cloud or *dark nebula* containing tiny, solid dust grains. Such opaque clouds are conspicuous on any photograph of the Milky Way, the galaxy in which the Sun is located (see the figures in The Milky Way Galaxy). The "dark rift," which runs lengthwise down a long part of the Milky Way in our sky and appears to split it in two, is produced by a collection of such obscuring clouds.

While dust clouds are too cold to radiate a measurable amount of energy in the visible part of the spectrum, they glow brightly in

the infrared (Figure 20.11). The reason is that small dust grains absorb visible light and ultraviolet radiation very efficiently. The grains are heated by the absorbed radiation, typically to temperatures from 10 to about 500 K, and re-radiate this heat at infrared wavelengths.



(a) (b)

Figure 20.11 : Visible and Infrared Images of the Horsehead Nebula in Orion. This dark cloud is one of the best-known images in astronomy, probably because it really does resemble a horse's head. The horse-head shape is an extension of a large cloud of dust that fills the lower part of the picture. (a) Seen in visible light, the dust clouds are especially easy to see against the bright background. (b) This infrared radiation image from the region of the horse head was recorded by NASA's Wide-Field Infrared Survey Explorer. Note how the regions that appear dark in visible light appear bright in the infrared. The dust is heated by nearby stars and re-radiates this heat in the infrared. Only the top of the horse's head is visible in the infrared image. Bright dots seen in the nebula below and to the left and at the top of the horse head are young, newly formed stars. The insets show the horse head and the bright nebula in more detail. (credit a: modification of work by ESO and Digitized Sky Survey; credit b: modification of work by NASA/JPL-Caltech)

Thanks to their small sizes and low temperatures, interstellar grains radiate most of their energy at infrared to microwave frequencies, with wavelengths of tens to hundreds of microns. Earth's atmosphere is opaque to radiation at these wavelengths, so emission by interstellar dust is best measured from space. Observations from above Earth's atmosphere show that dust clouds are present throughout the plane of the Milky Way (Figure 20.12).

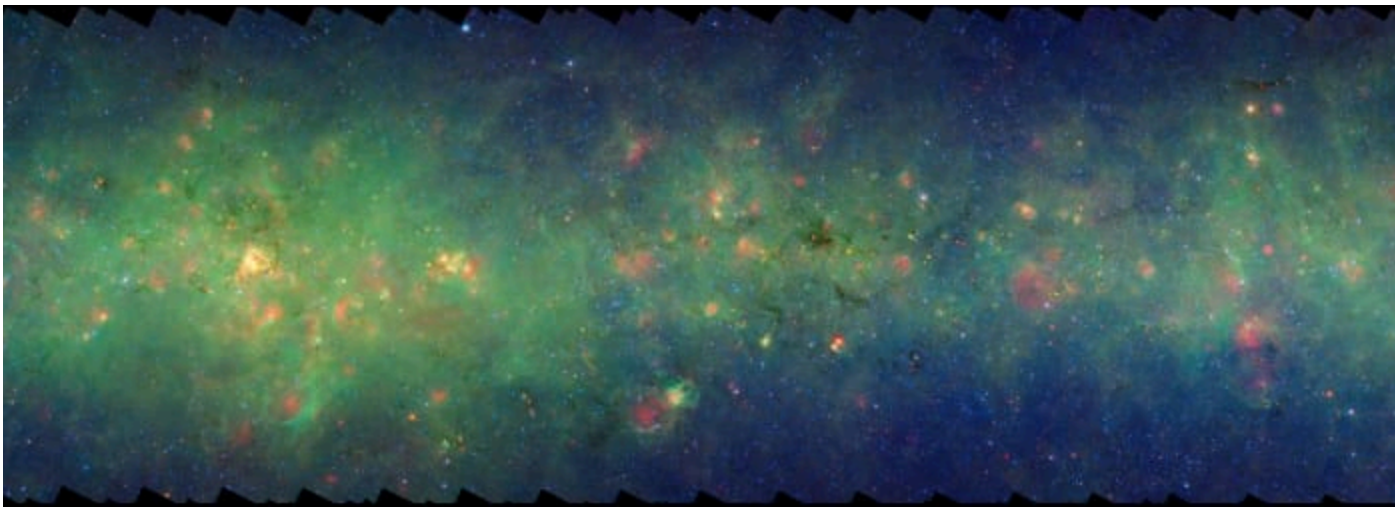


Figure 20.12 : Infrared Emission from the Plane of the Milky Way. This infrared image taken by the Spitzer Space Telescope shows a field in the plane of the Milky Way Galaxy. (Our Galaxy is in the shape of a frisbee; the plane of the Milky Way is the flat disk of that frisbee. Since the Sun, Earth, and solar system are located in the plane of the Milky Way and at a large distance from its center, we view the Galaxy edge on, much as we might look at a glass plate from its edge.) This emission is produced by tiny dust grains, which emit at 3.6 microns (blue in this image), 8.0 microns (green), and 24 microns (red). The densest regions of dust are so cold and opaque that they appear as dark clouds even at these infrared wavelengths. The red bubbles visible throughout indicate regions where the dust has been warmed up by young stars. This heating increases the emission at 24 microns, leading to the redder color in this image. (credit: modification of work by NASA/JPL-Caltech/ University of Wisconsin)

Some dense clouds of dust are close to luminous stars and scatter enough starlight to become visible. Such a cloud of dust, illuminated by starlight, is called a *reflection nebula*, since the light we see is starlight reflected off the grains of dust. One of the best-known examples is the nebulosity around each of the brightest stars in the Pleiades cluster (see Figure 20.1). The dust grains are small, and such small particles turn out to scatter light with blue wavelengths more efficiently than light at red wavelengths. A reflection nebula, therefore, usually appears bluer than its illuminating star (Figure 20.13).



Figure 20.13 : Pleiades Star Cluster. The bluish light surrounding the stars in this image is an example of a reflection nebula. Like fog around a street lamp, a reflection nebula shines only because the dust within it scatters light from a nearby bright source. The Pleiades cluster is currently passing through an interstellar cloud that contains dust grains, which scatter the light from the hot blue stars in the cluster. The Pleiades cluster is about 400 light-years from the Sun. (credit: NASA, ESA and AURA/Caltech)

Gas and dust are generally intermixed in space, although the proportions are not exactly the same everywhere. The presence of dust is apparent in many photographs of *emission nebulae* in the constellation of Sagittarius, where we see an H II region surrounded by a blue reflection nebula. Which type of nebula appears brighter depends on the kinds of stars that cause the gas and dust to glow. Stars cooler than about 25,000 K have so little ultraviolet radiation of wavelengths shorter than 91.2

nanometers—which is the wavelength required to ionize hydrogen—that the reflection nebulae around such stars outshine the emission nebulae. Stars hotter than 25,000 K emit enough ultraviolet energy that the emission nebulae produced around them generally outshine the reflection nebulae.

Interstellar Reddening

The tiny interstellar dust grains absorb some of the starlight they intercept. But at least half of the starlight that interacts with a grain is merely scattered, that is, it is redirected rather than absorbed. Since neither the absorbed nor the scattered starlight reaches us directly, both absorption and scattering make stars look dimmer. The effects of both processes are called interstellar extinction (Figure 20.14).

Astronomers first came to understand interstellar extinction around the early 1930s, as the explanation of a puzzling observation. In the early part of the twentieth century, astronomers discovered that some stars look red even though their spectral lines indicate that they must be extremely hot (and thus should look blue). The solution to this seeming contradiction turned out to be that the light from these hot stars is not only dimmed but also reddened by interstellar dust, a phenomenon known as interstellar reddening.

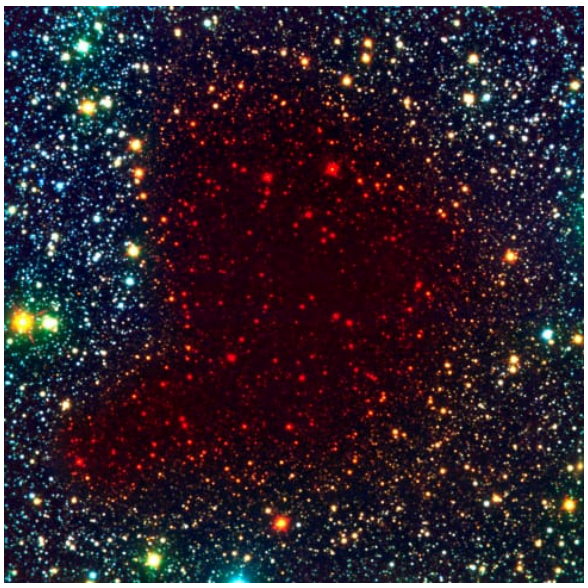


Figure 20.14 : Barnard 68 in Infrared. In this image, we see Barnard 68, the same object shown in Figure 20.9. The difference is that, in the previous image, the blue, green, and red channels showed light in the visible (or very nearly visible) part of the spectrum. In this image, the red color shows radiation emitted in the infrared at a wavelength of 2.2 microns. Interstellar extinction is much smaller at infrared than at visible wavelengths, so the stars behind the cloud become visible in the infrared channel. (credit: ESO)

Dust does not interact with all the colors of light the same way. Much of the violet, blue, and green light from these stars has been scattered or absorbed by dust, so it does not reach Earth. Some of their orange and red light, with longer wavelengths, on the other hand, more easily penetrates the intervening dust and completes its long journey through space to enter Earth-based telescopes (Figure 20.15). Thus, the star looks redder from Earth than it would if you could see it from nearby. (Strictly speaking, *reddening* is not the most accurate term for this process, since no red color is added; instead, blues and related colors are subtracted, so it should more properly be called “deblueing.”) In the most extreme cases, stars can be so reddened that they are entirely undetectable at visible wavelengths and can be seen only at infrared or longer wavelengths (Figure 20.14).

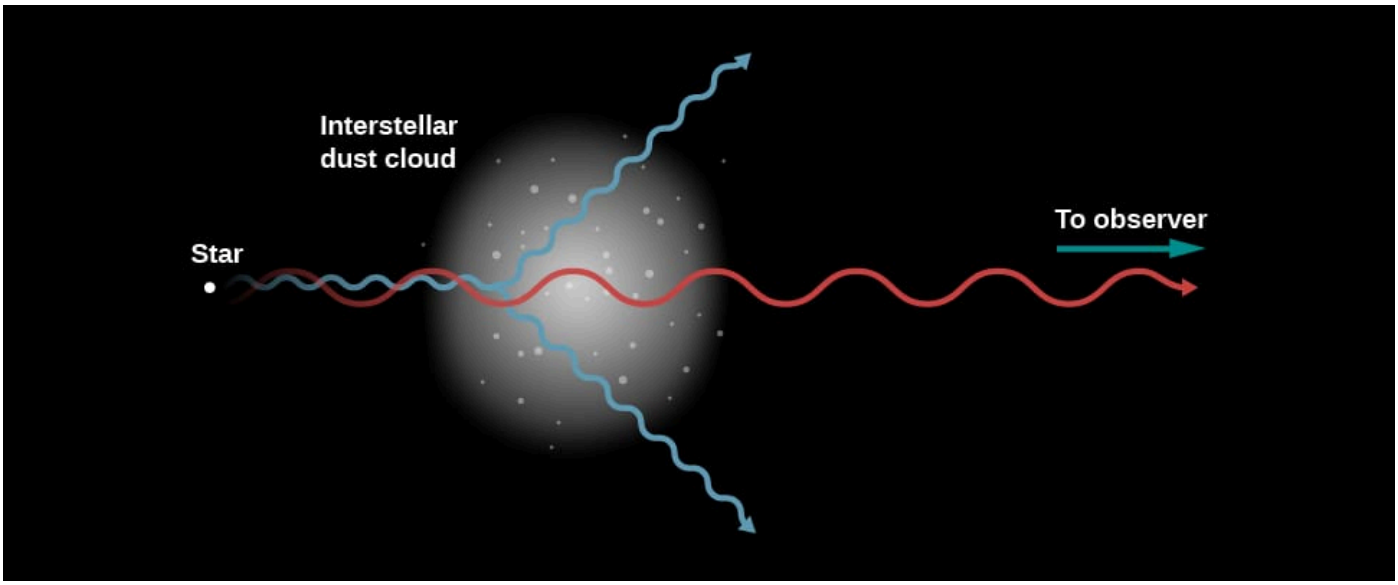


Figure 20.15 : Scattering of Light by Dust. Interstellar dust scatters blue light more efficiently than red light, thereby making distant stars appear redder and giving clouds of dust near stars a bluish hue. Here, a red ray of light from a star comes straight through to the observer, whereas a blue ray is shown scattering. A similar scattering process makes Earth's sky look blue.

We have all seen an example of reddening on Earth. The Sun appears much redder at sunset than it does at noon. The lower the Sun is in the sky, the longer the path its light must travel through the atmosphere. Over this greater distance, there is a greater chance that sunlight will be scattered. Since red light is less likely to be scattered than blue light, the Sun appears more and more red as it approaches the horizon.

By the way, scattering of sunlight is also what causes our sky to look blue, even though the gases that make up Earth's atmosphere are transparent. As sunlight comes in, it scatters from the molecules of air. The small size of the molecules means that the blue colors scatter much more efficiently than the greens, yellows, and reds. Thus, the blue in sunlight is scattered out of the beam and all over the sky. The light from the Sun that comes to your eye, on the other hand, is missing some of its blue, so the Sun looks a bit yellower, even when it is high in the sky, than it would from space.

The fact that starlight is reddened by interstellar dust means that long-wavelength radiation is transmitted through the Galaxy more efficiently than short-wavelength radiation. Consequently, if we wish to see farther in a direction with considerable interstellar material, we should look at long wavelengths. This simple fact provides one of the motivations for the development of infrared astronomy. In the infrared region at 2 microns (2000 nanometers), for example, the obscuration is only one-sixth as great as in the visible region (500 nanometers), and we can therefore study stars that are more than twice as distant before their light is blocked by interstellar dust. This ability to see farther by observing in the infrared portion of the spectrum represents a major gain for astronomers trying to understand the structure of our Galaxy or probing its puzzling, but distant, center (see The Milky Way Galaxy).

Interstellar Grains

Before we get to the details about interstellar dust, we should perhaps get one concern out of the way. Why couldn't it be the interstellar *gas* that reddens distant stars and not the dust? We already know from everyday experience that atomic or molecular gas is almost transparent. Consider Earth's atmosphere. Despite its very high density compared with that of interstellar gas, it is so transparent as to be practically invisible. (Gas does have a few specific spectral lines, but they absorb only a tiny fraction of the light as it passes through.)The quantity of *gas* required to produce the observed absorption of light in interstellar space would have to be enormous. The gravitational attraction of so great a mass of gas would affect the motions of stars in ways that could easily be detected. Such motions are not observed, and thus, the interstellar absorption cannot be the result of gases.

Although gas does not absorb much light, we know from everyday experience that tiny solid or liquid particles can be very efficient absorbers. Water vapor in the air is quite invisible. When some of that vapor condenses into tiny water droplets, however, the resulting cloud is opaque. Dust storms, smoke, and smog offer familiar examples of the efficiency with which solid particles absorb light. On the basis of arguments like these, astronomers have concluded that widely scattered s*olid* particles in interstellar space must be responsible for the observed dimming of starlight. What are these particles made of? And how did they form?

Observations like the pictures in this chapter show that a great deal of this dust exists; hence, it must be primarily composed of elements that are abundant in the universe (and in interstellar matter). After hydrogen and helium, the most abundant elements are oxygen, carbon, and nitrogen. These three elements, along with magnesium, silicon, iron—and perhaps hydrogen itself—turn out to be the most important components of interstellar dust.

Many of the dust particles can be characterized as sootlike (rich in carbon) or sandlike (containing silicon and oxygen). Grains of interstellar dust are found in meteorites and can be identified because the abundances of certain isotopes are different from what we see in other solar system material. Several different interstellar dust substances have been identified in this way in the laboratory, including graphite and diamonds. (Don't get excited; these diamonds are only a billionth of a meter in size and would hardly make an impressive engagement ring!)

The most widely accepted model pictures the grains with rocky cores that are either like soot (rich in carbon) or like sand (rich in silicates). In the dark clouds where molecules can form, these cores are covered by icy mantles (Figure 20.16). The most common ices in the grains are water (H_2O), methane (CH_4), and ammonia (NH_3)—all built out of atoms that are especially abundant in the realm of the stars. The ice mantles, in turn, are sites for some of the chemical reactions that produce complex organic molecules.

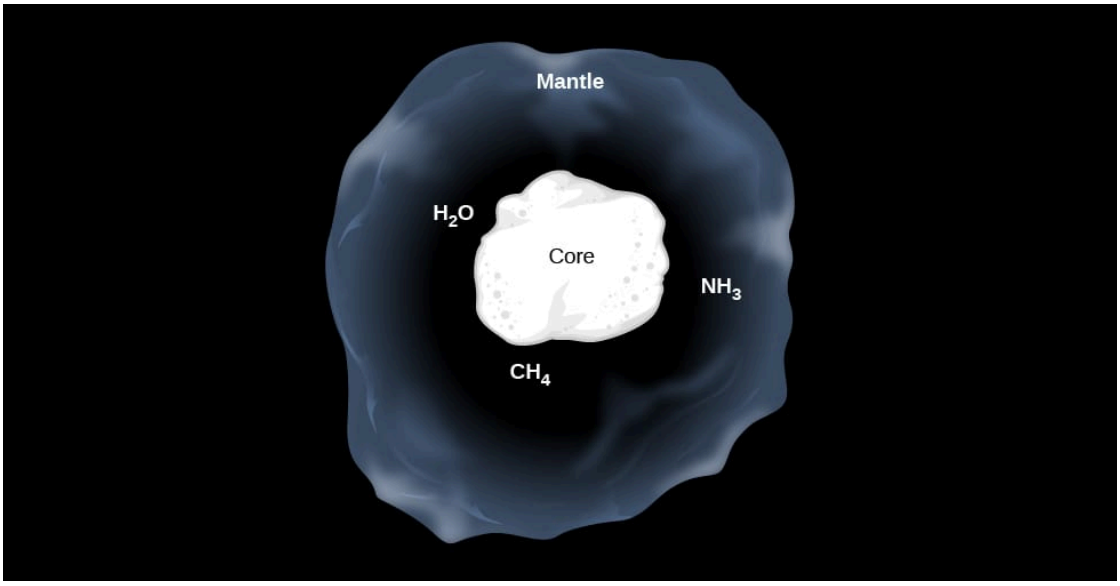


Figure 20.16 : Model of an Interstellar Dust Grain. A typical interstellar grain is thought to consist of a core of rocky material (silicates) or graphite, surrounded by a mantle of ices. Typical grain sizes are 10^{-8} to 10^{-7} meters. (This is from 1/100 to 1/10 of a micron; by contrast, human hair is about 10–200 microns wide.)

Typical individual grains must be just slightly smaller than the wavelength of visible light. If the grains were a lot smaller, they would not block the light efficiently, as Figure 20.13 and other images in this chapter show that it does.

On the other hand, if the dust grains were much larger than the wavelength of light, then starlight would not be reddened. Things that are much larger than the wavelength of light would block both blue and red light with equal efficiency. In this way we can deduce that a characteristic interstellar dust grain contains 10^6 to 10^9 atoms and has a diameter of 10^{-8} to 10^{-7} meters (10 to 100 nanometers). This is actually more like the specks of solid matter in cigarette smoke than the larger grains of dust you might find under your desk when you are too busy studying astronomy to clean properly.

This page titled 15.4: Cosmic Dust is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 20.4: Cosmic Dust by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.5: The Life Cycle of Cosmic Material

Learning Objectives

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

- Explain how interstellar matter flows into and out of our Galaxy and transforms from one phase to another, and understand how star formation and evolution affects the properties of the interstellar medium
- Explain how the heavy elements and dust grains found in interstellar space got there and describe how dust grains help produce molecules that eventually find their way into planetary systems

Flows of Interstellar Gas

The most important thing to understand about the interstellar medium is that it is not static. Interstellar gas orbits through the Galaxy, and as it does so, it can become more or less dense, hotter and colder, and change its state of ionization. A particular parcel of gas may be neutral hydrogen at some point, then find itself near a young, hot star and become part of an H II region. The star may then explode as a supernova, heating the nearby gas up to temperatures of millions of degrees. Over millions of years, the gas may cool back down and become neutral again, before it collects into a dense region that gravity gathers into a giant molecular cloud (Figure 20.18)

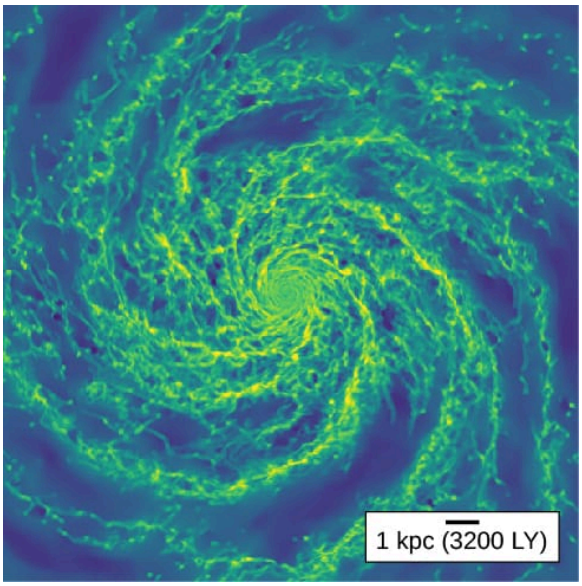


Figure 20.18 : Large-Scale Distribution of Interstellar Matter. This image is from a computer simulation of the Milky Way Galaxy's interstellar medium as a whole. The majority of gas, visible in greenish colors, is neutral hydrogen. In the densest regions in the spiral arms, shown in yellow, the gas is collected into giant molecular clouds. Low-density holes in the spiral arms, shown in blue, are the result of supernova explosions. (credit: modification of work by Mark Krumholz)

At any given time in the Milky Way, the majority of the interstellar gas by mass and volume is in the form of atomic hydrogen. The much-denser molecular clouds occupy a tiny fraction of the volume of interstellar space but add roughly 30% to the total mass of gas between the stars. Conversely, the hot gas produced by supernova explosions contributes a negligible mass but occupies a significant fraction of the volume of interstellar space. H II regions, though they are visually spectacular, constitute only a very small fraction of either the mass or volume of interstellar material.

However, the interstellar medium is not a closed system. Gas from intergalactic space constantly falls onto the Milky Way due to its gravity, adding new gas to the interstellar medium. Conversely, in giant molecular clouds where gas collects together due to gravity, the gas can collapse to form new stars, as discussed in The Birth of Stars and the Discovery of Planets outside the Solar System. This process locks interstellar matter into stars. As the stars age, evolve, and eventually die, massive stars lose

a large fraction of their mass, and low-mass stars lose very little. On average, roughly one-third of the matter incorporated into stars goes back into interstellar space. Supernova explosions have so much energy that they can drive interstellar mass out of the Galaxy and back into intergalactic space. Thus, the total amount of mass of the interstellar medium is set by a competition between the gain of mass from intergalactic space, the conversion of interstellar mass into stars, and the loss of interstellar mass back into intergalactic space due to supernovae. This entire process is known as the baryon cycle—baryon is from the Greek word for "heavy," and the cycle has this name because it is the repeating process that the heavier components of the universe—the atoms—undergo.

The Cycle of Dust and Heavy Elements

While much of the mass of the interstellar medium is material accreted during the last few billion years from intergalactic space, this is not true of the elements heavier than hydrogen and helium, or of the dust. Instead, these components of the interstellar medium were made inside stars in the Milky Way, which returned them to the interstellar medium at the end of their lives. We will talk more about this process in later chapters, but for now just bear in mind what we learned in The Sun: A Nuclear Powerhouse. What stars "do for a living" is fuse heavier elements from lighter ones, producing energy in the process. As stars mature, they begin to lose some of the newly made elements to the reservoir of interstellar matter.

The same is true of dust grains. Dust forms when grains can condense in regions where gas is dense and cool. One place where the right conditions are found is the winds from luminous cool stars (the red giants and supergiants we discussed in The Stars: A Celestial Census). Grains can also condense in the matter thrown off by a supernova explosion as the ejected gases begin to cool.

The dust grains produced by stars may grow even further when they spend time in the dense parts of the interstellar medium, inside molecular clouds. In these environments, grains can stick together or gather additional atoms from the gas around them, growing larger. They also facilitate the production of other compounds, including some of the more complex molecules we discussed earlier.

The surfaces of the dust grains (see Cosmic Dust)—which would seem very large if you were an atom—provide "nooks and crannies" where these atoms can stick long enough to find partners and form molecules. (Think of the dust grains as "interstellar social clubs" where lonely atoms can meet and form meaningful relationships.) Eventually, the dust grains become coated with ices. The presence of the dust shields the molecules inside the clouds from ultraviolet radiation and cosmic rays that would break them up.

When stars finally begin to form within the cloud, they heat the grains and evaporate the ices. The gravitational attraction of the newly forming stars also increases the density of the surrounding cloud material. Many more chemical reactions take place on the surfaces of grains in the gas surrounding the newly forming stars, and these areas are where organic molecules are formed. These molecules can be incorporated into newly formed planetary systems, and the early Earth may have been seeded in just such a way.

Indeed, scientists speculate that some of the water on Earth may have come from interstellar grains. Recent observations from space have shown that water is abundant in dense interstellar clouds. Since stars are formed from this material, water must be present when solar systems, including our own, come into existence. The water in our oceans and lakes may have come initially from water locked into the rocky material that accreted to form Earth. Alternatively, the water may have been brought to Earth when asteroids and comets (formed from the same cloud that made the planets) later impacted it. Scientists estimate that one comet impact every thousand years during Earth's first billion years would have been enough to account for the water we see today. Of course, both sources may have contributed to the water we now enjoy drinking and swimming in.

Any interstellar grains that are incorporated into newly forming stars (instead of the colder planets and smaller bodies around them) will be destroyed by their high temperatures. But eventually, each new generation of stars will evolve to become red giants, with stellar winds of their own. Some of these stars will also become supernovae and explode. Thus, the process of recycling cosmic material can start all over again.

This page titled 15.5: The Life Cycle of Cosmic Material is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 20.6: The Life Cycle of Cosmic Material by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.6: Spiral Structure

Learning Objectives

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

- Describe the structure of the Milky Way Galaxy and how astronomers discovered it
- Compare theoretical models for the formation of spiral arms in disk galaxies

Astronomers were able to make tremendous progress in mapping the spiral structure of the Milky Way after the discovery of the 21-cm line that comes from cool hydrogen (see Between the Stars: Gas and Dust in Space). Remember that the obscuring effect of interstellar dust prevents us from seeing stars at large distances in the disk at visible wavelengths. However, radio waves of 21-cm wavelength pass right through the dust, enabling astronomers to detect hydrogen atoms throughout the Galaxy. More recent surveys of the infrared emission from stars in the disk have provided a similar dust-free perspective of our Galaxy's stellar distribution. Despite all this progress over the past fifty years, we are still just beginning to pin down the precise structure of our Galaxy.

The Arms of the Milky Way

Our radio observations of the disk's gaseous component indicate that the Galaxy has two major spiral arms that emerge from the bar and several fainter arms and shorter spurs. You can see a recently assembled map of our Galaxy's arm structure—derived from studies in the infrared—in Figure 25.10.

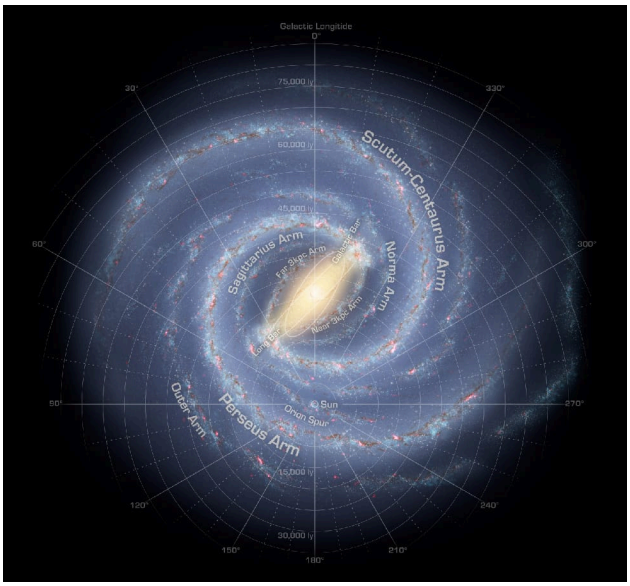


Figure 25.10 : Milky Way Bar and Arms. Here, we see the Milky Way Galaxy as it would look from above. This image, assembled from data from NASA's Spitzer mission, shows that the Milky Way Galaxy has a modest bar in its central regions. Two spiral arms, Scutum-Centaurus and Perseus, emerge from the ends of the bar and wrap around the bulge. The Sagittarius and Outer arms have fewer stars than the other two arms. (credit: modification of work by NASA/JPL-Caltech/R. Hurt (SSC/Caltech))

The Sun is near the inner edge of a short arm called the Orion Spur, which is about 10,000 light-years long and contains such conspicuous features as the Cygnus Rift (the great dark nebula in the summer Milky Way) and the bright Orion Nebula. Figure 25.11 shows a few other objects that share this small section of the Galaxy with us and are easy to see. Remember, the farther away we try to look from our own arm, the more the dust in the Galaxy builds up and makes it hard to see with visible light.

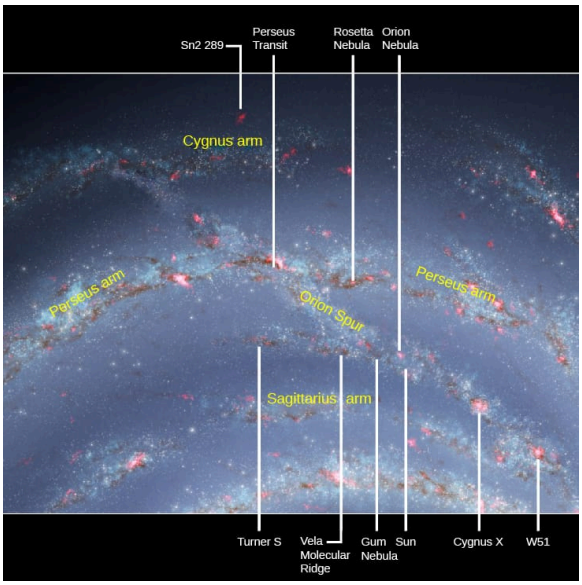


Figure 25.11 : Orion Spur. The Sun is located in the Orion Spur, which is a minor spiral arm located between two other arms. In this diagram, the white lines point to some other noteworthy objects that share this feature of the Milky Way Galaxy with the Sun. (credit: modification of work by NASA/JPL-Caltech)

Formation of Spiral Structure

At the Sun's distance from its center, the Galaxy does not rotate like a solid wheel or a CD inside your player. Instead, the way individual objects turn around the center of the Galaxy is more like the solar system. Stars, as well as the clouds of gas and dust, obey Kepler's third law. Objects farther from the center take longer to complete an orbit around the Galaxy than do those closer to the center. In other words, stars (and interstellar matter) in larger orbits in the Galaxy trail behind those in smaller ones. This effect is called differential galactic rotation.

Differential rotation would appear to explain why so much of the material in the disk of the Milky Way is concentrated into elongated features that resemble spiral arms. No matter what the original distribution of the material might be, the differential rotation of the Galaxy can stretch it out into spiral features. Figure 25.12 shows the development of spiral arms from two irregular blobs of interstellar matter. Notice that as the portions of the blobs closest to the galactic center move faster, those farther out trail behind.

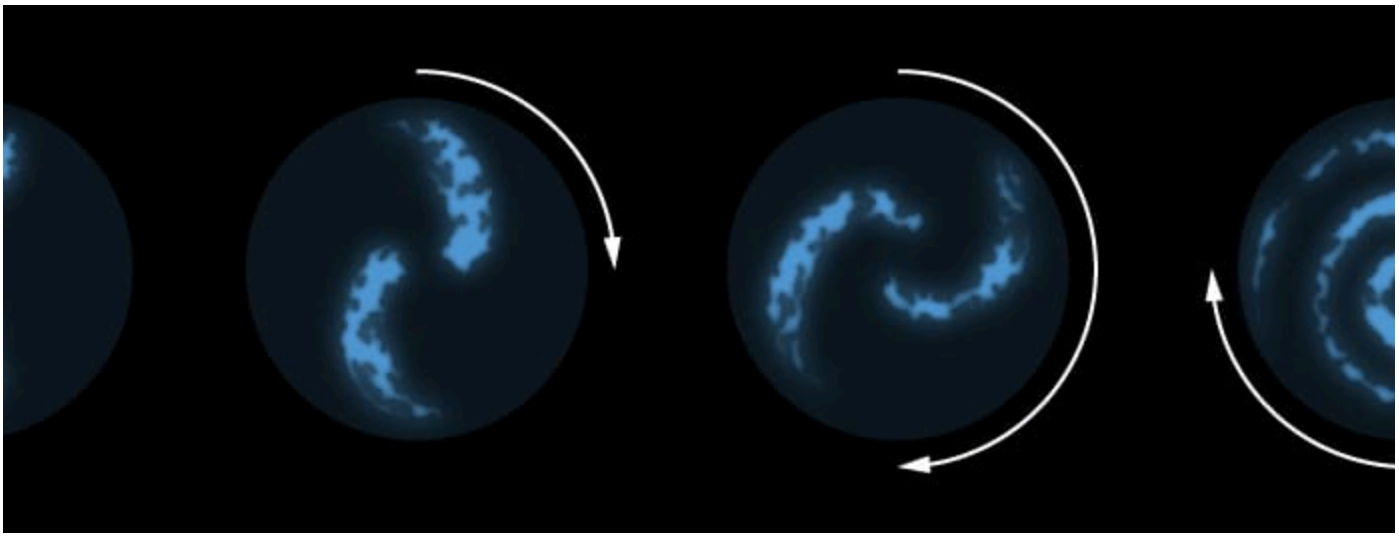


Figure 25.12 : Simplified Model for the Formation of Spiral Arms. This sketch shows how spiral arms might form from irregular clouds of interstellar material stretched out by the different rotation rates throughout the Galaxy. The regions farthest from the galactic center take longer to complete their orbits and thus lag behind the inner regions. If this were the only mechanism for creating spiral arms, then over time the spiral arms would completely wind up and disappear. Since many galaxies have spiral arms, they must be long-lived, and there must be other processes at work to maintain them.

But this picture of spiral arms presents astronomers with an immediate problem. If that's all there were to the story, differential rotation—over the roughly 13-billion-year history of the Galaxy—would have wound the Galaxy's arms tighter and tighter until all semblance of spiral structure had disappeared. But did the Milky Way actually have spiral arms when it formed 13 billion years ago? And do spiral arms, once formed, last for that long a time?

With the advent of the Hubble Space Telescope, it has become possible to observe the structure of very distant galaxies and to see what they were like shortly after they began to form more than 13 billion years ago. What the observations show is that galaxies in their infancy had bright, clumpy star-forming regions, but no regular spiral structure.

Over the next few billion years, the galaxies began to "settle down." The galaxies that were to become spirals lost their massive clumps and developed a central bulge. The turbulence in these galaxies decreased, rotation began to dominate the motions of the stars and gas, and stars began to form in a much quieter disk. Smaller star-forming clumps began to form fuzzy, not-very-distinct spiral arms. Bright, well-defined spiral arms began to appear only when the galaxies were about 3.6 billion years old. Initially, there were two well-defined arms. Multi-armed structures in galaxies like we see in the Milky Way appeared only when the universe was about 8 billion years old.

We will discuss the history of galaxies in more detail in The Evolution and Distribution of Galaxies. But, even from our brief discussion, you can get the sense that the spiral structures we now observe in mature galaxies have come along later in the full story of how things develop in the universe.

Scientists have used supercomputer calculations to model the formation and evolution of the arms. These calculations follow the motions of up to 100 million "star particles" to see whether gravitational forces can cause them to form spiral structure. What these calculations show is that giant molecular clouds (which we discussed in Between the Stars: Gas and Dust in Space) have enough gravitational influence over their surroundings to initiate the formation of structures that look like spiral arms. These arms then become self-perpetuating and can survive for at least several billion years. The arms may change their brightness over time as star formation comes and goes, but they are not temporary features. The concentration of matter in the arms exerts sufficient gravitational force to keep the arms together over long periods of time.

This page titled 15.6: Spiral Structure is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 25.3: Spiral Structure by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.7: The Mass of the Galaxy

Learning Objectives

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

- Describe historical attempts to determine the mass of the Galaxy
- Interpret the observed rotation curve of our Galaxy to suggest the presence of dark matter whose distribution extends well beyond the Sun's orbit

When we described the sections of the Milky Way, we said that the stars are now known to be surrounded by a much larger halo of invisible matter. Let's see how this surprising discovery was made.

Kepler Helps Weigh the Galaxy

The Sun, like all the other stars in the Galaxy, orbits the center of the Milky Way. Our star's orbit is nearly circular and lies in the Galaxy's disk. The speed of the Sun in its orbit is about 200 kilometers per second, which means it takes us approximately 225 million years to go once around the center of the Galaxy. We call the period of the Sun's revolution the *galactic year*. It is a long time compared to human time scales; during the entire lifetime of Earth, only about 20 galactic years have passed. This means that we have gone only a tiny fraction of the way around the Galaxy in all the time that humans have gazed into the sky.

We can use the information about the Sun's orbit to estimate the mass of the Galaxy (just as we could "weigh" the Sun by monitoring the orbit of a planet around it—see Orbits and Gravity). Let's assume that the Sun's orbit is circular and that the Galaxy is roughly spherical, (we know the Galaxy is shaped more like a disk, but to simplify the calculation we will make this assumption, which illustrates the basic approach). Long ago, Newton showed that if you have matter distributed in the shape of a sphere, then it is simple to calculate the pull of gravity on some object just outside that sphere: you can assume that gravity acts as if all the matter were concentrated at a point in the center of the sphere. For our calculation, then, we can assume that all the mass that lies inward of the Sun's position is concentrated at the center of the Galaxy, and that the Sun orbits that point from a distance of about 26,000 light-years.

This is the sort of situation to which Kepler's third law (as modified by Newton) can be directly applied. Plugging numbers into Kepler's formula, we can calculate the sum of the masses of the Galaxy and the Sun. However, the mass of the Sun is completely trivial compared to the mass of the Galaxy. Thus, for all practical purposes, the result (about 100 billion times the mass of the Sun) is the mass of the Milky Way. More sophisticated calculations based on more sophisticated models give a similar result.

Our estimate tells us how much mass is contained in the volume inside the Sun's orbit. This is a good estimate for the total mass of the Galaxy only if hardly any mass lies outside the Sun's orbit. For many years astronomers thought this assumption was reasonable. The number of bright stars and the amount of *luminous matter* (meaning any material from which we can detect electromagnetic radiation) both drop off dramatically at distances of more than about 30,000 light-years from the galactic center. Little did we suspect how wrong our assumption was.

A Galaxy of Mostly Invisible Matter

In science, what seems to be a reasonable assumption can later turn out to be wrong (which is why we continue to do observations and experiments every chance we get). There is a lot more to the Milky Way than meets the eye (or our instruments). While there is relatively little luminous matter beyond 30,000 light-years, we now know that a lot of *invisible matter* exists at great distances from the galactic center.

We can understand how astronomers detected this invisible matter by remembering that according to Kepler's third law, objects orbiting at large distances from a massive object will move more slowly than objects that are closer to that central mass. In the case of the solar system, for example, the outer planets move more slowly in their orbits than the planets close to the Sun.

There are a few objects, including globular clusters and some nearby small satellite galaxies, that lie well outside the luminous

boundary of the Milky Way. If most of the mass of our Galaxy were concentrated within the luminous region, then these very distant objects should travel around their galactic orbits at lower speeds than, for example, the Sun does.

It turns out, however, that the few objects seen at large distances from the luminous boundary of the Milky Way Galaxy are *not* moving more slowly than the Sun. There are some globular clusters and RR Lyrae stars between 30,000 and 150,000 light-years from the center of the Galaxy, and their orbital velocities are even greater than the Sun's (Figure 25.13).

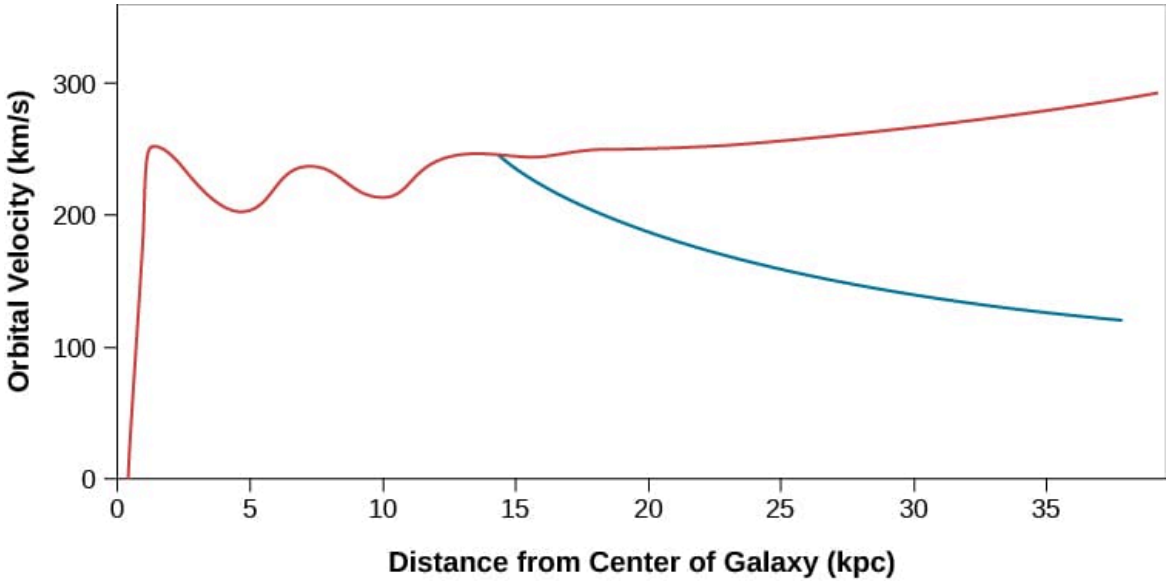


Figure 25.13 : Rotation Curve of the Galaxy. The orbital speed of carbon monoxide (CO) and hydrogen (H) gas at different distances from the center of the Milky Way Galaxy is shown in red. The blue curve shows what the rotation curve would look like if all the matter in the Galaxy were located inside a radius of 50,000 light-years. Instead of going down, the speed of gas clouds farther out remains high, indicating a great deal of mass beyond the Sun's orbit. The horizontal axis shows the distance from the galactic center in kiloparsecs (where a kiloparsec equals 3,260 light-years).

What do these higher speeds mean? Kepler's third law tells us how fast objects must orbit a source of gravity if they are neither to fall in (because they move too slowly) nor to escape (because they move too fast). If the Galaxy had only the mass calculated by Kepler, then the high-speed outer objects should long ago have escaped the grip of the Milky Way. The fact that they have not done so means that our Galaxy must have more gravity than can be supplied by the luminous matter—in fact, a *lot* more gravity. The high speed of these outer objects tells us that the source of this extra gravity must extend outward from the center far beyond the Sun's orbit.

If the gravity were supplied by stars or by something else that gives off radiation, we should have spotted this additional outer material long ago. We are therefore forced to the reluctant conclusion that this matter is invisible and has, except for its gravitational pull, gone entirely undetected.

Studies of the motions of the most remote globular clusters and the small galaxies that orbit our own show that the total mass of the Galaxy is at least 2×10^{12} M_{Sun}, which is about twenty times greater than the amount of luminous matter. Moreover, the dark matter (as astronomers have come to call the invisible material) extends to a distance of at least 200,000 light-years from the center of the Galaxy. Observations indicate that this dark matter halo is almost but not quite spherical.

The obvious question is: what is the dark matter made of? Let's look at a list of "suspects" taken from our study of astronomy so far. Since this matter is invisible, it clearly cannot be in the form of ordinary stars. And it cannot be gas in any form (remember that there has to be a lot of it). If it were neutral hydrogen gas, its 21-cm wavelength spectral-line emission would have been detected as radio waves. If it were ionized hydrogen, it should be hot enough to emit visible radiation. If a lot of hydrogen atoms out there had combined into hydrogen molecules, these should produce dark features in the ultraviolet spectra of objects lying beyond the Galaxy, but such features have not been seen. Nor can the dark matter consist of interstellar dust, since in the required quantities, the dust would significantly obscure the light from distant galaxies.

What are our other possibilities? The dark matter cannot be a huge number of black holes (of stellar mass) or old neutron stars, since interstellar matter falling onto such objects would produce more X-rays than are observed. Also, recall that the formation of black holes and neutron stars is preceded by a substantial amount of mass loss, which scatters heavy elements into space to be

incorporated into subsequent generations of stars. If the dark matter consisted of an enormous number of any of those objects, they would have blown off and recycled a lot of heavier elements over the history of the Galaxy. In that case, the young stars we observe in our Galaxy today would contain much greater abundances of heavy elements than they actually do.

Brown dwarfs and lone Jupiter-like planets have also been ruled out. First of all, there would have to be an awful lot of them to make up so much dark matter. But we have a more direct test of whether so many low-mass objects could actually be lurking out there. As we learned in Black Holes and Curved Spacetime, the general theory of relativity predicts that the path traveled by light is changed when it passes near a concentration of mass. It turns out that when the two objects appear close enough together in the sky, the mass closer to us can bend the light from farther away. With just the right alignment, the image of the more distant object also becomes significantly brighter. By looking for the temporary brightening that occurs when a dark matter object in our own Galaxy moves across the path traveled by light from stars in the Magellanic Clouds, astronomers have now shown that the dark matter cannot be made up of a lot of small objects with masses between one-millionth and one-tenth the mass of the Sun.

What's left? One possibility is that the dark matter is composed of exotic subatomic particles of a type not yet detected on Earth. Very sophisticated (and difficult) experiments are now under way to look for such particles. Stay tuned to see whether anything like that turns up.

We should add that the problem of dark matter is by no means confined to the Milky Way. Observations show that dark matter must also be present in other galaxies (whose outer regions also orbit too fast "for their own good"—they also have flat rotation curves). As we will see, dark matter even exists in great clusters of galaxies whose members are now known to move around under the influence of far more gravity than can be accounted for by luminous matter alone.

Stop a moment and consider how astounding the conclusion we have reached really is. Perhaps as much as 95% of the mass in our Galaxy (and many other galaxies) is not only invisible, but we do not even know what it is made of. The stars and raw material we can observe may be merely the tip of the cosmic iceberg; underlying it all may be other matter, perhaps familiar, perhaps startlingly new. Understanding the nature of this dark matter is one of the great challenges of astronomy today; you will learn more about this in A Universe of (Mostly) Dark Matter and Dark Energy.

Link to Learning

You can use the Dark Matter Simulator to explore the distribution of invisible dark matter in our galaxy. Use the controls in the upper right to change the amount of dark matter at different distances from the center of the galaxy. You can then compare the orbital speeds in your simulated galaxy with orbital speeds measured in the Milky Way.

This page titled 15.7: The Mass of the Galaxy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 25.4: The Mass of the Galaxy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.8: Stellar Populations in the Galaxy

Learning Objectives

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

- Distinguish between population I and population II stars according to their locations, motions, heavy-element abundances, and ages
- Explain why the oldest stars in the Galaxy are poor in elements heavier than hydrogen and helium, while stars like the Sun and even younger stars are typically richer in these heavy elements

In the first section of his chapter, we described the thin disk, thick disk, and stellar halo. Look back at Table 25.1 and note some of the patterns. Young stars lie in the thin disk, are rich in metals, and orbit the Galaxy's center at high speed. The stars in the halo are old, have low abundances of elements heavier than hydrogen and helium, and have highly elliptical orbits randomly oriented in direction (see Figure 25.19). Halo stars can plunge through the disk and central bulge, but they spend most of their time far above or below the plane of the Galaxy. The stars in the thick disk are intermediate between these two extremes. Let's first see why age and heavier-element abundance are correlated and then see what these correlations tell us about the origin of our Galaxy.

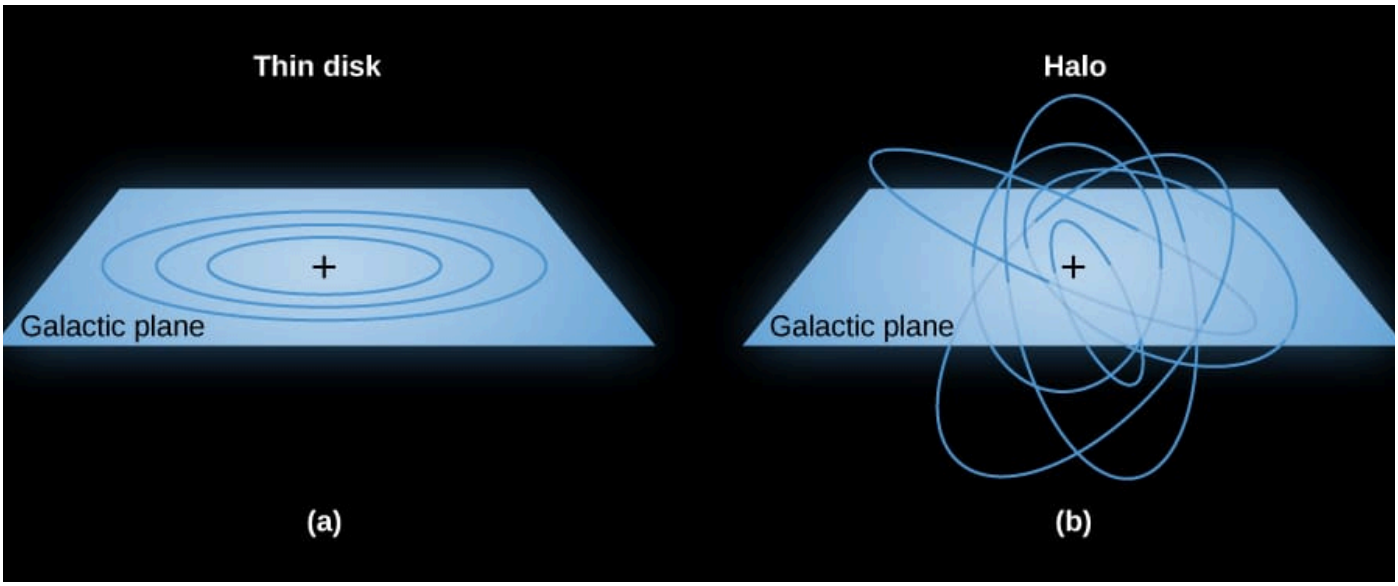


Figure 25.19 : How Objects Orbit the Galaxy. (a) In this image, you see stars in the thin disk of our Galaxy in nearly circular orbits. (b) In this image, you see the motion of stars in the Galaxy's halo in randomly oriented and elliptical orbits.

Two Kinds of Stars

The discovery that there are two different kinds of stars was first made by Walter Baade during World War II. As a German national, Baade was not allowed to do war research as many other U.S.-based scientists were doing, so he was able to make regular use of the Mount Wilson telescopes in southern California. His observations were aided by the darker skies that resulted from the wartime blackout of Los Angeles.

Among the things a large telescope and dark skies enabled Baade to examine carefully were *other* galaxies—neighbors of our Milky Way Galaxy. We will discuss other galaxies in the next chapter (Galaxies), but for now we will just mention that the nearest Galaxy that resembles our own (with a similar disk and spiral structure) is often called the Andromeda galaxy, after the constellation in which we find it.

Baade was impressed by the similarity of the mainly reddish stars in the Andromeda galaxy's nuclear bulge to those in our Galaxy's globular clusters and the halo. He also noted the difference in color between all these and the bluer stars found in the spiral arms

near the Sun (Figure 25.20). On this basis, he called the bright blue stars in the spiral arms population I and all the stars in the halo and globular clusters population II.



Figure 25.20 : Andromeda Galaxy (M31). This neighboring spiral looks similar to our own Galaxy in that it is a disk galaxy with a central bulge. Note the bulge of older, yellowish stars in the center, the bluer and younger stars in the outer regions, and the dust in the disk that blocks some of the light from the bulge. (credit: Adam Evans)

We now know that the populations differ not only in their locations in the Galaxy, but also in their chemical composition, age, and orbital motions around the center of the Galaxy. Population I stars are found only in the disk and follow nearly circular orbits around the galactic center. Examples are bright supergiant stars, main-sequence stars of high luminosity (spectral classes O and B), which are concentrated in the spiral arms, and members of young open star clusters. Interstellar matter and molecular clouds are found in the same places as population I stars.

Population II stars show no correlation with the location of the spiral arms. These objects are found throughout the Galaxy. Some are in the disk, but many others follow eccentric elliptical orbits that carry them high above the galactic disk into the halo. Examples include stars surrounded by planetary nebulae and RR Lyrae variable stars. The stars in globular clusters, found almost entirely in the Galaxy's halo, are also classified as population II.

Today, we know much more about stellar evolution than astronomers did in the 1940s, and we can determine the ages of stars. Population I includes stars with a wide range of ages. While some are as old as 10 billion years, others are still forming today. For example, the Sun, which is about 5 billion years old, is a population I star. But so are the massive young stars in the Orion Nebula that have formed in the last few million years. Population II, on the other hand, consists entirely of old stars that formed very early in the history of the Galaxy; typical ages are 11 to 13 billion years.

We also now have good determinations of the compositions of stars. These are based on analyses of the stars' detailed spectra. Nearly all stars appear to be composed mostly of hydrogen and helium, but their abundances of the heavier elements differ. In the Sun and other population I stars, the heavy elements (those heavier than hydrogen and helium) account for 1–4% of the total stellar mass. Population II stars in the outer galactic halo and in globular clusters have much lower abundances of the heavy elements—often less than one-hundredth the concentrations found in the Sun and in rare cases even lower. The oldest population II star discovered to date has less than one ten-millionth as much iron as the Sun, for example.

As we discussed in earlier chapters, heavy elements are created deep within the interiors of stars. They are added to the Galaxy's reserves of raw material when stars die, and their material is recycled into new generations of stars. Thus, as time goes on, stars are born with larger and larger supplies of heavy elements. Population II stars formed when the abundance of elements heavier than hydrogen and helium was low. Population I stars formed later, after mass lost by dying members of the first generations of stars had seeded the interstellar medium with elements heavier than hydrogen and helium. Some are still forming now, when further generations have added to the supply of heavier elements available to new stars.

The Real World

With rare exceptions, we should never trust any theory that divides the world into just two categories. While they can provide a starting point for hypotheses and experiments, they are often oversimplifications that need refinement a research continue. The idea of two populations helped organize our initial thoughts about the Galaxy, but we now know it cannot explain everything we observe. Even the different structures of the Galaxy—disk, halo, central bulge—are not so cleanly separated in terms of their locations, ages, and the heavy element content of the stars within them.

The exact definition of the Galaxy's disk depends on what objects we use to define it, and, as we saw earlier, it has no sharp boundary. The hottest young stars and their associated gas and dust clouds are mostly in a region about 200 light-years thick. Older stars define a thicker disk that is about 2000 light-years thick. Halo stars spend most of their time high above or below the disk but pass through it on their highly elliptical orbits and so are sometimes found relatively near the Sun.

The highest density of stars is found in the central bulge, that bar-shaped inner region of the Galaxy. There are a few hot, young stars in the bulge, but most of the bulge stars are more than 10 billion years old. Yet unlike the halo stars of similar age, the abundance of heavy elements in the bulge stars is about the same as in the Sun. Why would that be?

Astronomers think that star formation in the crowded nuclear bulge occurred very rapidly just after the Milky Way Galaxy formed. After a few million years, the first generation of massive and short-lived stars then expelled heavy elements in supernova explosions and thereby enriched subsequent generations of stars. Thus, even stars that formed in the bulge more than 10 billion years ago started with a good supply of heavy elements.

Exactly the opposite occurred in the Small Magellanic Cloud, a small galaxy near the Milky Way, visible from Earth's Southern Hemisphere. Even the youngest stars in this galaxy are deficient in heavy elements. We think this is because the little galaxy is not especially crowded, and star formation has occurred quite slowly. As a result there have been, so far, relatively few supernova explosions. Smaller galaxies also have more trouble holding onto the gas expelled by supernova explosions in order to recycle it. Low-mass galaxies exert only a modest gravitational force, and the high-speed gas ejected by supernovae can easily escape from them.

Which elements a star is endowed with thus depends not only on when the star formed in the history of its galaxy, but also on how many stars in its part of the galaxy had already completed their lives by the time the star is ready to form.

This page titled 15.8: Stellar Populations in the Galaxy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 25.6: Stellar Populations in the Galaxy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.9: The Formation of the Galaxy

Learning Objectives

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

- Describe the roles played by the collapse of a single cloud and mergers with other galaxies in building the Milky Way Galaxy we see today
- Provide examples of globular clusters and satellite galaxies affected by the Milky Way's strong gravity.

Information about stellar populations holds vital clues to how our Galaxy was built up over time. The flattened disk shape of the Galaxy suggests that it formed through a process similar to the one that leads to the formation of a protostar (see The Birth of Stars and the Discovery of Planets outside the Solar System). Building on this idea, astronomers first developed models that assumed the Galaxy formed from a single rotating cloud. But, as we shall see, this turns out to be only part of the story.

The Protogalactic Cloud and the Monolithic Collapse Model

Because the oldest stars—those in the halo and in globular clusters—are distributed in a sphere centered on the nucleus of the Galaxy, it makes sense to assume that the *protogalactic* cloud that gave birth to our Galaxy was roughly spherical. The oldest stars in the halo have ages of 12 to 13 billion years, so we estimate that the formation of the Galaxy began about that long ago. (See the chapter on The Big Bang for other evidence that galaxies in general began forming a little more than 13 billion years ago.) Then, just as in the case of star formation, the protogalactic cloud collapsed and formed a thin rotating disk. Stars born before the cloud collapsed did not participate in the collapse, but have continued to orbit in the halo to the present day (Figure 25.21).

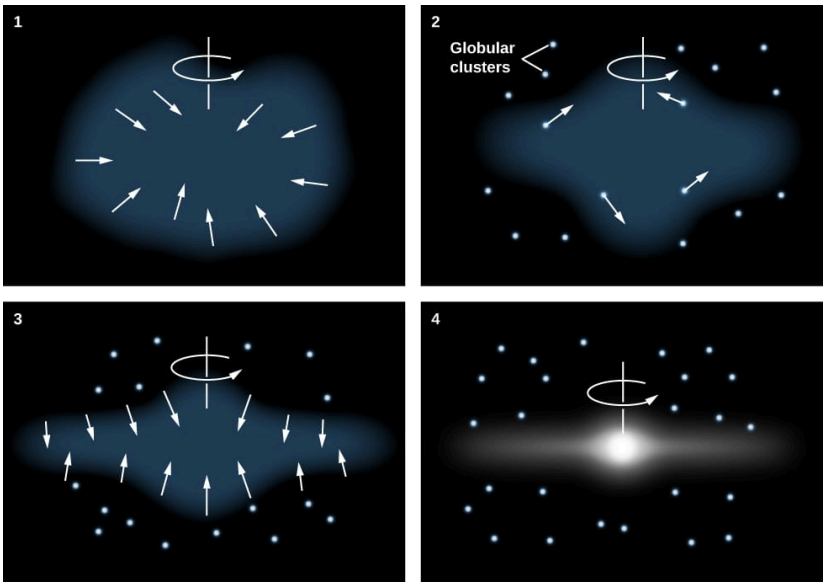


Figure 25.21 : Monolithic Collapse Model for the Formation of the Galaxy. According to this model, the Milky Way Galaxy initially formed from a rotating cloud of gas that collapsed due to gravity. Halo stars and globular clusters either formed prior to the collapse or were formed elsewhere. Stars in the disk formed later, when the gas from which they were made was already "contaminated" with heavy elements produced in earlier generations of stars.

Gravitational forces caused the gas in the thin disk to fragment into clouds or clumps with masses like those of star clusters. These individual clouds then fragmented further to form stars. Since the oldest stars in the disk are nearly as old as the youngest stars in the halo, the collapse must have been rapid (astronomically speaking), requiring perhaps no more than a few hundred million years.

Collision Victims and the Multiple Merger Model

In past decades, astronomers have learned that the evolution of the Galaxy has not been quite as peaceful as this monolithic collapse model suggests. In 1994, astronomers discovered a small new galaxy in the direction of the constellation of Sagittarius. The Sagittarius dwarf galaxy is currently about 70,000 light-years away from Earth and 50,000 light-years from the center of the Galaxy. It is the closest galaxy known (Figure 25.22). It is very elongated, and its shape indicates that it is being torn apart by our Galaxy's gravitational tides—just as Comet Shoemaker-Levy 9 was torn apart when it passed too close to Jupiter in 1992.

The Sagittarius galaxy is much smaller than the Milky Way and is about 10,000 times less massive than our Galaxy. All of the stars in the Sagittarius dwarf galaxy seem destined to end up in the bulge and halo of the Milky Way. But don't sound the funeral bells for the little galaxy quite yet; the ingestion of the Sagittarius dwarf will take another 100 million years or so, and the stars themselves will survive.

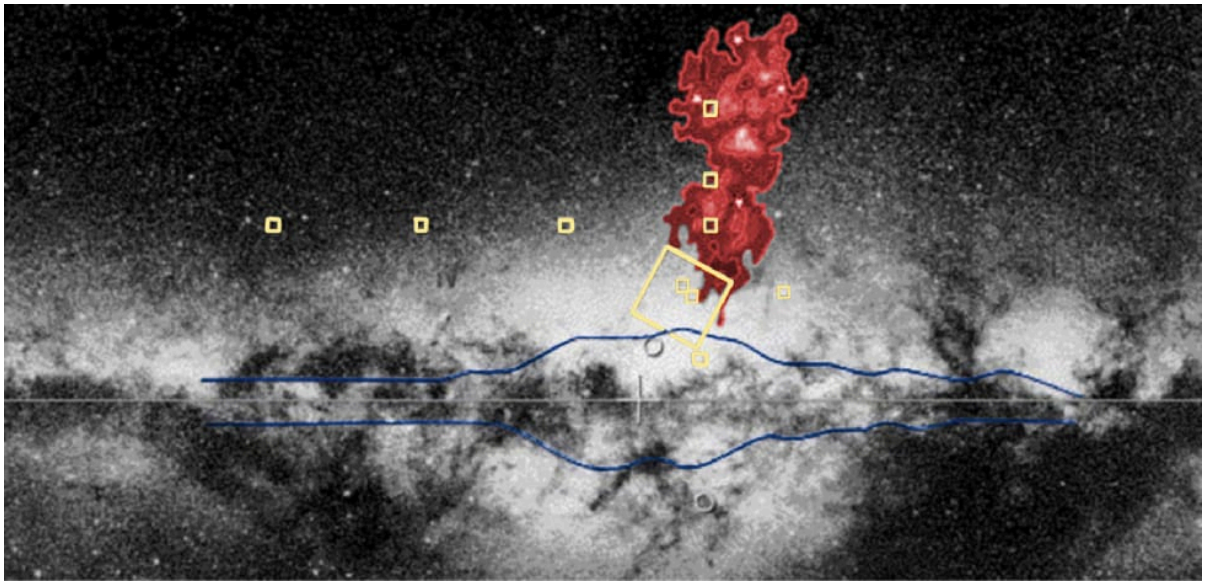


Figure 25.22 : Sagittarius Dwarf. In 1994, British astronomers discovered a galaxy in the constellation of Sagittarius, located only about 50,000 light-years from the center of the Milky Way and falling into our Galaxy. This image covers a region approximately 70° × 50° and combines a black-and-white view of the disk of our Galaxy with a red contour map showing the brightness of the dwarf galaxy. The dwarf galaxy lies on the other side of the galactic center from us. The white stars in the red region mark the locations of several globular clusters contained within the Sagittarius dwarf galaxy. The cross marks the galactic center. The horizontal line corresponds to the galactic plane. The blue outline on either side of the galactic plane corresponds to the infrared image in Figure 25.7. The boxes mark regions where detailed studies of individual stars led to the discovery of this galaxy. (credit: modification of work by R. Ibata (UBC), R. Wyse (JHU), R. Sword (IoA))

Since that discovery, evidence has been found for many more close encounters between our Galaxy and other neighbor galaxies. When a small galaxy ventures too close, the force of gravity exerted by our Galaxy tugs harder on the near side than on the far side. The net effect is that the stars that originally belonged to the small galaxy are spread out into a long stream that orbits through the halo of the Milky Way (Figure 25.23).

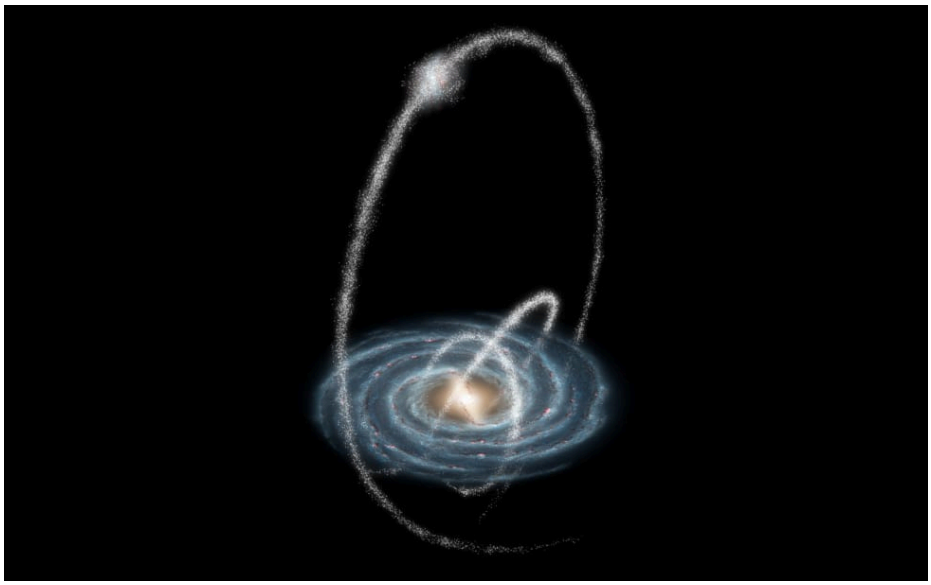


Figure 25.23 : Streams in the Galactic Halo. When a small galaxy is swallowed by the Milky Way, its member stars are stripped away and form streams of stars in the galactic halo. This image is based on calculations of what some of these tidal streams might look like if the Milky Way swallowed 50 dwarf galaxies over the past 10 billion years. (credit: modification of work by NASA/JPL-Caltech/R. Hurt (SSC/Caltech))

Such a tidal stream can maintain its identity for billions of years. To date, astronomers have now identified streams originating from 12 small galaxies that ventured too close to the much larger Milky Way. Six more streams are associated with globular clusters. It has been suggested that large globular clusters, like Omega Centauri, are actually dense nuclei of cannibalized dwarf galaxies. The globular cluster M54 is now thought to be the nucleus of the Sagittarius dwarf we discussed earlier, which is currently merging with the Milky Way (Figure 25.24). The stars in the outer regions of such galaxies are stripped off by the gravitational pull of the Milky Way, but the central dense regions may survive.

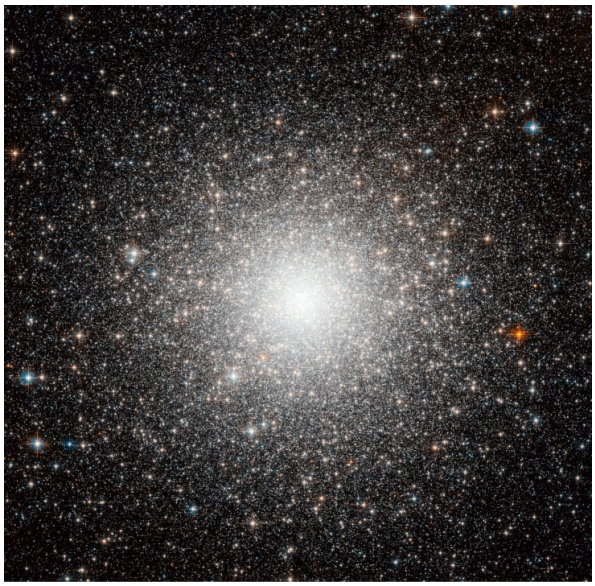


Figure 25.24 : Globular Cluster M54. This beautiful Hubble Space Telescope image shows the globular cluster that is now believed to be the nucleus of the Sagittarius Dwarf Galaxy. (credit: ESA/Hubble & NASA)

Calculations indicate that the Galaxy's thick disk may be a product of one or more such collisions with other galaxies. Accretion of a satellite galaxy would stir up the orbits of the stars and gas clouds originally in the thin disk and cause them to move higher above and below the mid-plane of the Galaxy. Meanwhile, the Galaxy's stars would add to the fluffed-up mix. If such a collision happened about 10 billion years ago, then any gas in the two galaxies that had not yet formed into stars would have had plenty of time to settle back down into the thin disk. The gas could then have begun forming subsequent generations of population I stars. This timing is also consistent with the typical ages of stars in the thick disk.

The Milky Way has more collisions in store. An example is the Canis Major dwarf galaxy, which has a mass of about 1% of the mass of the Milky Way. Already long tidal tails have been stripped from this galaxy, which have wrapped themselves around the Milky Way three times. Several of the globular clusters found in the Milky Way may also have come from the Canis Major dwarf, which is expected to merge gradually with the Milky Way over about the next billion years.

In about 4 billion years, the Milky Way itself will be swallowed up, since it and the Andromeda galaxy are on a collision course. Our computer models show that after a complex interaction, the two will merge to form a larger, more rounded galaxy (Figure 25.25).

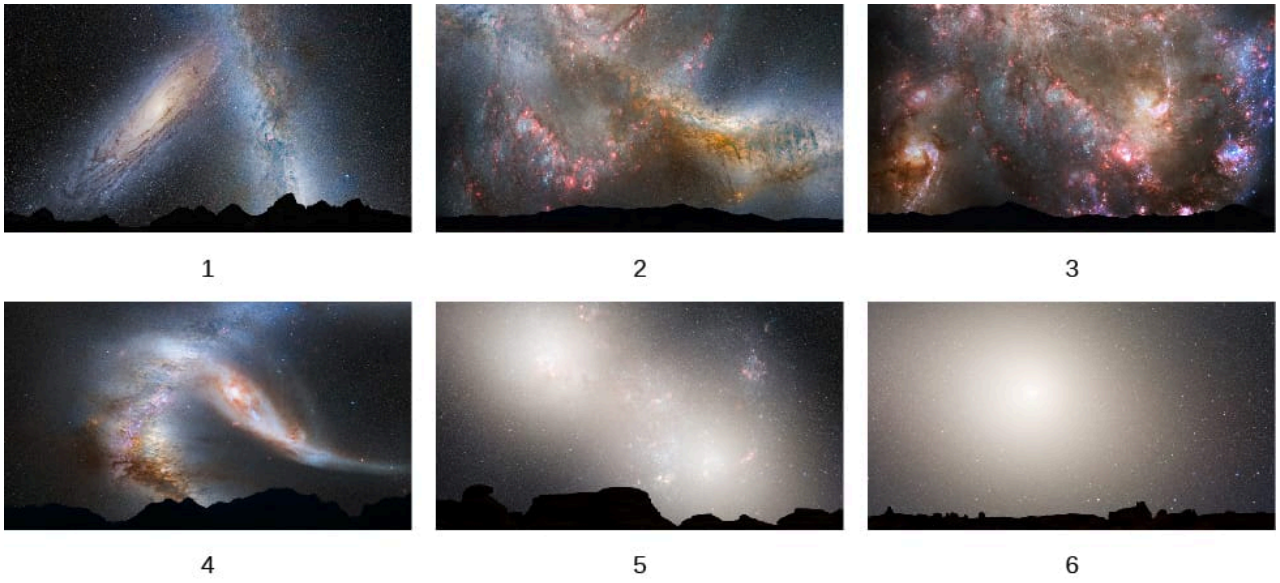


Figure 25.25 : Collision of the Milky Way with Andromeda. In about 3 billion years, the Milky Way Galaxy and Andromeda Galaxy will begin a long process of colliding, separating, and then coming back together to form an elliptical galaxy. The whole interaction will take 3 to 4 billion years. These computer-simulated images show the following sequence: (1) In 3.75 billion years, Andromeda has approached the Milky Way. (2) New star formation fills the sky 3.85 billion years from now. (3) Star formation continues at 3.9 billion years. (4) The galaxy shapes change as they interact, with Andromeda being stretched and our Galaxy becoming warped, about 4 billion years from now. (5) In 5.1 billion years, the cores of the two galaxies are bright lobes. (6) In 7 billion years, the merged galaxies form a huge elliptical galaxy whose brightness fills the night sky. This artist's illustrations show events from a vantage point 25,000 light-years from the center of the Milky Way. However, we should mention that the Sun may not be at that distance throughout the sequence of events, as the collision readjusts the orbits of many stars within each galaxy. (credit: NASA; ESA; Z. Levay, R. van der Marel, STScI; T. Hallas, and A. Mellinger)

We are thus coming to realize that "environmental influences" (and not just a galaxy's original characteristics) play an important role in determining the properties and development of our Galaxy. In future chapters we will see that collisions and mergers are a major factor in the evolution of many other galaxies as well.

This page titled 15.9: The Formation of the Galaxy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 25.7: The Formation of the Galaxy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

15.10: The Center of the Galaxy

Learning Objectives

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

- Describe the radio and X-ray observations that indicate energetic phenomena are occurring at the galactic center
- Explain what has been revealed by high-resolution near-infrared imaging of the galactic center
- Discuss how these near-infrared images, when combined with Kepler's third law of motion, can be used to derive the mass of the central gravitating object

At the beginning of this chapter, we hinted that the core of our Galaxy contains a large concentration of mass. In fact, we now have evidence that the very center contains a black hole with a mass equivalent to 4.6 million Suns and that all this mass fits within a sphere that has less than the diameter of Mercury's orbit. Such monster black holes are called supermassive black holes by astronomers, to indicate that the mass they contain is far greater than that of the typical black hole created by the death of a single star. It is amazing that we have very convincing evidence that this black hole really does exist. After all, recall from the chapter on Black Holes and Curved Spacetime that we cannot see a black hole directly because by definition it radiates no energy. And we cannot even see into the center of the Galaxy in visible light because of absorption by the interstellar dust that lies between us and the galactic center. Light from the central region of the Galaxy is dimmed by a factor of a trillion (10^{12}) by all this dust.

Fortunately, we are not so blind at other wavelengths. Infrared and radio radiation, which have long wavelengths compared to the sizes of the interstellar dust grains, flow unimpeded past the dust particles and so reach our telescopes with hardly any dimming. In fact, the very bright radio source in the nucleus of the Galaxy, now known as Sagittarius A* (pronounced "Sagittarius A-star" and abbreviated Sgr A*), was the first cosmic radio source astronomers discovered.

A Journey toward the Center

Let's take a voyage to the mysterious heart of our Galaxy and see what's there. Figure 25.14 is a radio image of a region about 1500 light-years across, centered on Sagittarius A, a bright radio source that contains the smaller Sagittarius A^*. Much of the radio emission comes from hot gas heated either by clusters of hot stars (the stars themselves do not produce radio emission and can't be seen in the image) or by supernova blast waves. Most of the hollow circles visible on the radio image are supernova remnants. The other main source of radio emission is from electrons moving at high speed in regions with strong magnetic fields. The bright thin arcs and "threads" on the figure show us where this type of emission is produced.

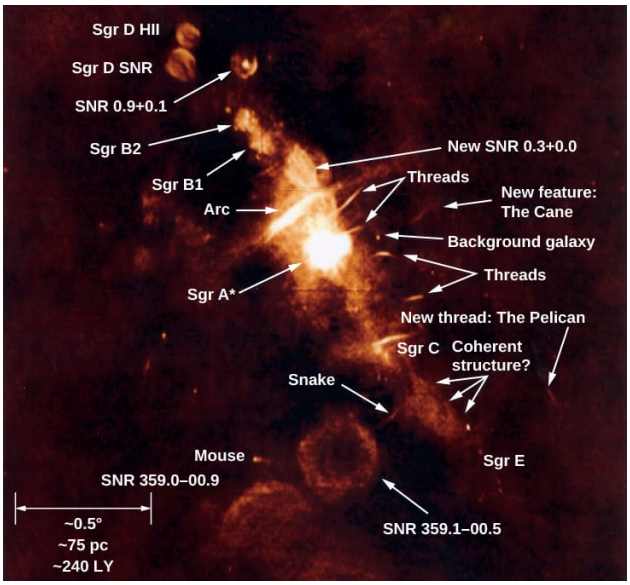


Figure 25.14 : Radio Image of Galactic Center Region. This radio map of the center of the Galaxy (at a wavelength of 90 centimeters) was constructed from data obtained with the Very Large Array (VLA) of radio telescopes in Socorro, New Mexico. Brighter regions are more intense in radio waves. The galactic center is inside the region labeled Sagittarius A. Sagittarius B1 and B2 are regions of active star formation. Many filaments or threadlike features are seen, as well as a number of shells (labeled SNR), which are supernova remnants. The scale bar at the bottom left is about 240 light-years long. Notice that radio astronomers also give fanciful animal names to some of the structures, much as visible-light nebulae are sometimes given the names of animals they resemble. (credit: modification of work by N. E. Kassim, D. S. Briggs, T. J. W. Lazio, T. N. LaRosa, and J. Imamura (NRL/RSD))

Now let's focus in on the central region using a more energetic form of electromagnetic radiation. Figure 25.15 shows the X-ray emission from a smaller region 400 light-years wide and 900 light-years across centered in Sagittarius A*. Seen in this picture are hundreds of hot white dwarfs, neutron stars, and stellar black holes with accretion disks glowing with X-rays. The diffuse haze in the picture is emission from gas that lies among the stars and is at a temperature of 10 million K.



Figure 25.15 : Galactic Center in X-Rays. This artificial-color mosaic of 30 images taken with the Chandra X-ray satellite shows a region 400 × 900 light-years in extent and centered on Sagittarius A*, the bright white source in the center of the picture. The X-ray-emitting point sources are white dwarfs, neutron stars, and stellar black holes. The diffuse "haze" is emission from gas at a temperature of 10 million K. This hot gas is flowing away from the center out into the rest of the Galaxy. The colors indicate X-ray energy bands: red (low energy), green (medium energy), and blue (high energy). (credit: modification of work by NASA/CXC/ UMass/D. Wang et al.)

As we approach the center of the Galaxy, we find the supermassive black hole Sagittarius A*. There are also thousands of stars within a few lightyears of Sagittarius A*. Most of these are old, reddish main-sequence stars. But there are also about a hundred

hot OB stars that must have formed within the last few million years. There is as yet no good explanation for how stars could have formed recently so close to a supermassive black hole. Perhaps they formed in a dense cluster of stars that was originally at a larger distance from the black hole and subsequently migrated closer.

There is currently no star formation at the galactic center, but there is lots of dust and molecular gas that is revolving around the black hole, along with some ionized gas streamers that are heated by the hot stars. Figure 25.16 is a radio map that shows these gas streamers.

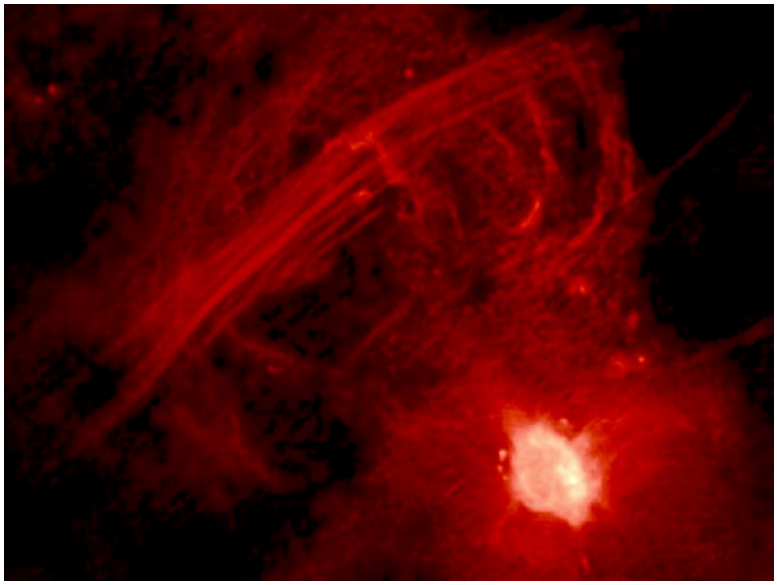


Figure 25.16 : Sagittarius A. This image, taken with the Very Large Array of radio telescopes, shows the radio emission from hot, ionized gas in the center of the Milky Way. The lines slanting across the top of the image are gas streamers. Sagittarius A* is the bright spot in the lower right. (credit: modification of work by Farhad Zadeh et al. (Northwestern), VLA, NRAO)

Finding the Heart of the Galaxy

Just what is Sagittarius A*, which lies right at the center our Galaxy? To establish that there really is a black hole there, we must show that there is a very large amount of mass crammed into a very tiny volume. As we saw in Black Holes and Curved Spacetime, proving that a black hole exists is a challenge because the black hole itself emits no radiation. What astronomers must do is prove that a black hole is the only possible explanation for our observations—that a small region contains far more mass than could be accounted for by a very dense cluster of stars or something else made of ordinary matter.

To put some numbers with this discussion, the radius of the event horizon of a *galactic black hole* with a mass of about 4 million M_{Sun} would be only about 17 times the size of the Sun—the equivalent of a single red giant star. The corresponding density within this region of space would be much higher than that of any star cluster or any other ordinary astronomical object. Therefore, we must measure both the diameter of Sagittarius A* and its mass. Both radio and infrared observations are required to give us the necessary evidence.

First, let's look at how the mass can be measured. If we zero in on the inner few light-days of the Galaxy with an infrared telescope equipped with adaptive optics, we see a region crowded with individual stars (Figure 25.17). These stars have now been observed for almost two decades, and astronomers have detected their rapid orbital motions around the very center of the Galaxy.

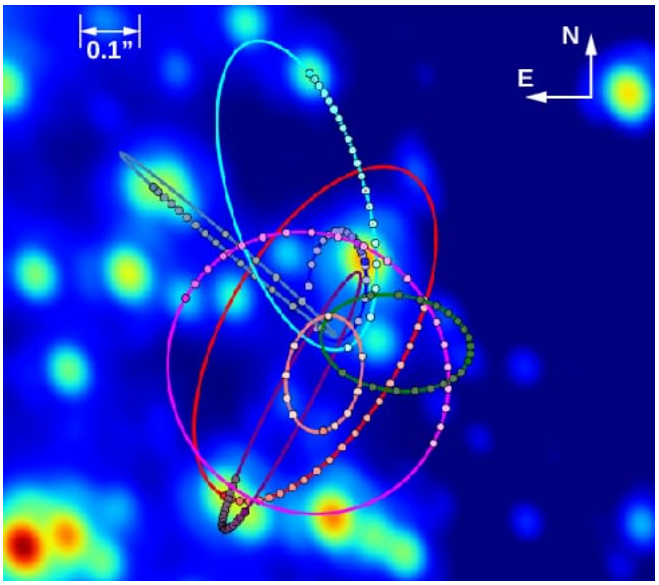


Figure 25.17 : Near-Infrared View of the Galactic Center. This image shows the inner 1 arcsecond, or 0.13 light-year, at the center of the Galaxy, as observed with the giant Keck Telescope. Tracks of the orbiting stars measured from 1995 to 2014 have been added to this "snapshot." The stars are moving around the center very fast, and their tracks are all consistent with a single massive "gravitator" that resides in the very center of this image. (credit: modification of work by Andrea Ghez, UCLA Galactic Center Group, W.M. Keck Observatory Laser Team)

Link to Learning

Check out an animated version of Figure 25.17, showing the motion of the stars over the years.

If we combine observations of their periods and the size of their orbits with Kepler's third law, we can estimate the mass of the object that keeps them in their orbits. One of the stars has been observed for its full orbit of 15.6 years. Its closest approach takes it to a distance of only 124 AU or about 17 light-hours from the black hole. This orbit, when combined with observations of other stars close to the galactic center, indicates that a mass of 4.6 million M_{Sun} must be concentrated inside the orbit—that is, within 17 light-hours of the center of the Galaxy.

Even tighter limits on the size of the concentration of mass at the center of the Galaxy come from radio astronomy, which provided the first clue that a black hole might lie at the center of the Galaxy. As matter spirals inward toward the event horizon of a black hole, it is heated in a whirling *accretion disk* and produces radio radiation. (Such accretion disks were explained in Black Holes and Curved Spacetime.) Measurements of the size of the accretion disk with the Very Long Baseline Array, which provides very high spatial resolution, show that the diameter of the radio source Sagittarius A^* is no larger than about 0.3 AU, or about the size of Mercury's orbit. (In light units, that's only 2.5 light-*minutes*!)

The observations thus show that 4.6 million solar masses are crammed into a volume that has a diameter that is no larger than the orbit of Mercury. If this were anything other than a supermassive black hole—low-mass stars that emit very little light or neutron stars or a very large number of small black holes— calculations show that these objects would be so densely packed that they would collapse to a single black hole within a hundred thousand years. That is a very short time compared with the age of the Galaxy, which probably began forming more than 13 billion years ago. Since it seems very unlikely that we would have caught such a complex cluster of objects just before it collapsed, the evidence for a supermassive black hole at the center of the Galaxy is convincing indeed.

Finding the Source

Where did our galactic black hole come from? The origin of supermassive black holes in galaxies like ours is currently an active field of research. One possibility is that a large cloud of gas near the center of the Milky Way collapsed directly to form a black hole. Since we find large black holes at the centers of most other large galaxies (see Active Galaxies, Quasars, and Supermassive Black Holes)—even ones that are very young—this collapse probably would have taken place when the Milky Way was just beginning to take shape. The initial mass of this black hole might have been only a few tens of solar masses. Another

way it could have started is that a massive star might have exploded to leave behind a seed black hole, or a dense cluster of stars might have collapsed into a black hole.

Once a black hole exists at the center of a galaxy, it can grow over the next several billion years by devouring nearby stars and gas clouds in the crowded central regions. It can also grow by merging with other black holes.

It appears that the monster black hole at the center of our Galaxy is not finished "eating." At the present time, we observe clouds of gas and dust falling into the galactic center at the rate of about 1 M_{Sun} per thousand years. Stars are also on the black hole's menu. The density of stars near the galactic center is high enough that we would expect a star to pass near the black hole and be swallowed by it every ten thousand years or so. As this happens, some of the energy of infall is released as radiation. As a result, the center of the Galaxy might flare up and even briefly outshine all the stars in the Milky Way. Other objects might also venture too close to the black hole and be pulled in. How great a flare we observe would depend on the mass of the object falling in.

In 2013, the Chandra X-ray satellite detected a flare from the center of our Galaxy that was 400 times brighter than the usual output from Sagittarius A*. A year later, a second flare, only half as bright, was also detected. This is much less energy than swallowing a whole star would produce. There are two theories to account for the flares. First, an asteroid might have ventured too close to the black hole and been heated to a very high temperature before being swallowed up. Alternatively, the flares might have involved interactions of the magnetic fields near the galactic center in a process similar to the one described for solar flares (see The Sun: A Garden-Variety Star). Astronomers continue to monitor the galactic center area for flares or other activity. Although the monster in the center of the Galaxy is not close enough to us to represent any danger, we still want to keep our eyes on it.

Voyagers in Astronomy: Andrea Ghez

A lover of puzzles, Andrea Ghez has been pursuing one of the greatest mysteries in astronomy: what strange entity lurks within the center of our Milky Way Galaxy?



Figure 25.18 : Andrea Ghez. Research by Ghez and her team has helped shape our understanding of supermassive black holes. (credit: modification of work by John D. and Catherine T. MacArthur Foundation)

As a child living in Chicago during the late 1960s, Andrea Ghez (Figure 25.18) was fascinated by the Apollo Moon landings. But she was also drawn to ballet and to solving all sorts of puzzles. By high school, she had lost the ballet bug in favor of competing in field hockey, playing the flute, and digging deeper into academics. Her undergraduate years at MIT were punctuated by a number of changes in her major—from mathematics to chemistry, mechanical engineering, aerospace engineering, and finally physics—where she felt her options were most open. As a physics major, she became involved in astronomical research under the guidance of one of her instructors. Once she got to do some actual observing at Kitt Peak National Observatory in Arizona, and later at Cerro Tololo Inter-American Observatory in Chile, Ghez had found her calling.

Pursuing her graduate studies at Caltech, she stuck with physics but oriented her efforts toward observational astrophysics, an area where Caltech had access to cutting-edge facilities. Though initially attracted to studying the black holes that were suspected of dwelling inside most massive galaxies, Ghez ended up spending most of her graduate study and later postdoctoral research at the University of Arizona studying stars in formation. By taking very high-resolution (detailed) imaging of regions where new stars are born, she discovered that most stars form as members of binary systems. As technologies advanced, she was able to track the orbits danced by these stellar pairings and thereby could ascertain their respective masses.

Now an astronomy professor at UCLA, Ghez has since used similar high-resolution imaging techniques to study the orbits of stars in the innermost core of the Milky Way. These orbits take years to delineate, so Ghez and her science team have logged more than 20 years of taking super-resolution infrared images with the giant Keck telescopes in Hawaii. Based on the resulting stellar orbits, the UCLA Galactic Center Group has settled (as we saw) on a gravitational solution that requires the presence of a supermassive black hole with a mass equivalent to 4.6 million Suns—all nestled within a space smaller than that occupied by our solar system. Ghez's achievements have been recognized with one of the "genius" awards given by the MacArthur Foundation. More recently, her team discovered glowing clouds of warm ionized gas that co-orbit with the stars but may be more vulnerable to the disruptive effects of the central black hole. By monitoring these clouds, the team hopes to better understand the evolution of supermassive black holes and their immediate environs. They also hope to test Einstein's theory of general relativity by carefully scrutinizing the orbits of stars that careen closest to the intensely gravitating black hole.

In 2020, Ghez received the Nobel Prize in Physics for her work on the black hole at the center of the Galaxy. You can see her explain her work in non-technical language in this video. Besides her pioneering work as an astronomer, Ghez competes as a master swimmer, enjoys family life as a mother of two children, and actively encourages other women to pursue scientific careers.

This page titled 15.10: The Center of the Galaxy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 25.5: The Center of the Galaxy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16: THE EXPANDING UNIVERSE OF GALAXIES

16.1: The Discovery of Galaxies
16.2: The Extragalactic Distance Scale
16.3: The Expanding Universe
16.4: Types of Galaxies
16.5: Properties of Galaxies
16.6: Quasars
16.7: Supermassive Black Holes- What Quasars really are
16.8: Quasars as Probes of Evolution in the Universe

This page titled 16: The Expanding Universe of Galaxies is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

16.1: The Discovery of Galaxies

Learning Objectives

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

- Describe the discoveries that confirmed the existence of galaxies that lie far beyond the Milky Way Galaxy
- Explain why galaxies used to be called nebulae and why we don't include them in that category any more

Growing up at a time when the Hubble Space Telescope orbits above our heads and giant telescopes are springing up on the great mountaintops of the world, you may be surprised to learn that we were not sure about the existence of other galaxies for a very long time. The very idea that other galaxies exist used to be controversial. Even into the 1920s, many astronomers thought the Milky Way encompassed *all* that exists in the universe. The evidence found in 1924 that meant our Galaxy is not alone was one of the great scientific discoveries of the twentieth century.

It was not that scientists weren't asking questions. They questioned the composition and structure of the universe as early as the eighteenth century. However, with the telescopes available in earlier centuries, galaxies looked like small fuzzy patches of light that were difficult to distinguish from the star clusters and gas-and-dust clouds that are part of our own Galaxy. All objects that were not sharp points of light were given the same name, *nebulae*, the Latin word for "clouds." Because their precise shapes were often hard to make out and no techniques had yet been devised for measuring their distances, the nature of the nebulae was the subject of much debate.

As early as the eighteenth century, the philosopher Immanuel Kant (1724–1804) suggested that some of the nebulae might be distant systems of stars (other Milky Ways), but the evidence to support this suggestion was beyond the capabilities of the telescopes of that time.

Other Galaxies

By the early twentieth century, some nebulae had been correctly identified as star clusters, and others (such as the Orion Nebula) as gaseous nebulae. Most nebulae, however, looked faint and indistinct, even with the best telescopes, and their distances remained unknown. (For more on how such nebulae are named, by the way, see the feature box on Naming the Nebulae in the chapter on interstellar matter.) If these nebulae were nearby, with distances comparable to those of observable stars, they were most likely clouds of gas or groups of stars within our Galaxy. If, on the other hand, they were remote, far beyond the edge of the Galaxy, they could be other star systems containing billions of stars.

To determine what the nebulae are, astronomers had to find a way of measuring the distances to at least some of them. When the 2.5-meter (100-inch) telescope on Mount Wilson in Southern California went into operation, astronomers finally had the large telescope they needed to settle the controversy.

Working with the 2.5-meter telescope, Edwin Hubble was able to resolve individual stars in several of the brighter spiral-shaped nebulae, including M31, the great spiral in Andromeda (Figure 26.2). Among these stars, he discovered some faint variable stars that—when he analyzed their light curves—turned out to be cepheids. Here were reliable indicators that Hubble could use to measure the distances to the nebulae using the technique pioneered by Henrietta Leavitt (see the chapter on Celestial Distances). After painstaking work, he estimated that the Andromeda galaxy was about 900,000 light-years away from us. At that enormous distance, it had to be a separate galaxy of stars located well outside the boundaries of the Milky Way. Today, we know the Andromeda galaxy is actually slightly more than twice as distant as Hubble's first estimate, but his conclusion about its true nature remains unchanged.



Figure 26.2 : Andromeda Galaxy. Also known by its catalog number M31, the Andromeda galaxy is a large spiral galaxy very similar in appearance to, and slightly larger than, our own Galaxy. At a distance of about 2.5 million light-years, Andromeda is the spiral galaxy that is nearest to our own in space. Here, it is seen with two of its satellite galaxies, M32 (top) and M110 (bottom). (credit: Adam Evans)

No one in human history had ever measured a distance so great. When Hubble's paper on the distances to nebulae was read before a meeting of the American Astronomical Society on the first day of 1925, the entire room erupted in a standing ovation. A new era had begun in the study of the universe, and a new scientific field—extragalactic astronomy—had just been born.

Voyagers in Astronomy: Edwin Hubble: Expanding the Universe

The son of a Missouri insurance agent, Edwin Hubble (Figure 26.3) graduated from high school at age 16. He excelled in sports, winning letters in track and basketball at the University of Chicago, where he studied both science and languages. Both his father and grandfather wanted him to study law, however, and he gave in to family pressure. He received a prestigious Rhodes scholarship to Oxford University in England, where he studied law with only middling enthusiasm. Returning to be the United States, he spent a year teaching high school physics and Spanish as well as coaching basketball, while trying to determine his life's direction.



Figure 26.3 : Edwin Hubble (1889–1953). Edwin Hubble established some of the most important ideas in the study of galaxies.

The pull of astronomy eventually proved too strong to resist, and so Hubble went back to the University of Chicago for graduate work. Just as he was about to finish his degree and accept an offer to work at the soon-to-be-completed 2.5-meter telescope, the United States entered World War I, and Hubble enlisted as an officer. Although the war had ended by the time he arrived in Europe, he received more officer's training abroad and enjoyed a brief time of further astronomical study at Cambridge before being sent home.

In 1919, at age 30, he joined the staff at Mount Wilson and began working with the world's largest telescope. Ripened by experience, energetic, disciplined, and a skillful observer, Hubble soon established some of the most important ideas in modern astronomy. He showed that other galaxies existed, classified them on the basis of their shapes, found a pattern to their motion (and thus put the notion of an expanding universe on a firm observational footing), and began a lifelong program to study the distribution of galaxies in the universe. Although a few others had glimpsed pieces of the puzzle, it was Hubble who put it all together and showed that an understanding of the large-scale structure of the universe was feasible.

His work brought Hubble much renown and many medals, awards, and honorary degrees. As he became better known (he was the first astronomer to appear on the cover of *Time* magazine), he and his wife enjoyed and cultivated friendships with movie stars and writers in Southern California. Hubble was instrumental (if you'll pardon the pun) in the planning and building of the 5-meter telescope on Palomar Mountain, and he had begun to use it for studying galaxies when he passed away from a stroke in 1953.

When astronomers built a space telescope that would allow them to extend Hubble's work to distances he could only dream about, it seemed natural to name it in his honor. It was fitting that observations with the Hubble Space Telescope (and his foundational work on expansion of the universe) contributed to the 2011 Nobel Prize in Physics, given for the discovery that the expansion of the universe is accelerating (a topic we will expand upon in the chapter on The Big Bang).

This page titled 16.1: The Discovery of Galaxies is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 26.2: The Discovery of Galaxies by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.2: The Extragalactic Distance Scale

Learning Objectives

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

- Describe the use of variable stars to estimate distances to galaxies
- Explain how standard bulbs and the Tully-Fisher relation can be used to estimate distances to galaxies

To determine many of the properties of a galaxy, such as its luminosity or size, we must first know how far away it is. If we know the distance to a galaxy, we can convert how bright the galaxy appears to us in the sky into its true luminosity because we know the precise way light is dimmed by distance. (The same galaxy 10 times farther away, for example, would look 100 times dimmer.) But the measurement of galaxy distances is one of the most difficult problems in modern astronomy: all galaxies are far away, and most are so distant that we cannot even make out individual stars in them.

For decades after Hubble's initial work, the techniques used to measure galaxy distances were relatively inaccurate, and different astronomers derived distances that differed by as much as a factor of two. (Imagine if the distance between your home or dorm and your astronomy class were this uncertain; it would be difficult to make sure you got to class on time.) In the past few decades, however, astronomers have devised new techniques for measuring distances to galaxies; most importantly, all of them give the same answer to within an accuracy of about 10%. As we will see, this means we may finally be able to make reliable estimates of the size of the universe.

Variable Stars

Before astronomers could measure distances to other galaxies, they first had to establish the scale of cosmic distances using objects in our own Galaxy. We described the chain of these distance methods in Celestial Distances (and we recommend that you review that chapter if it has been a while since you've read it). Astronomers were especially delighted when they discovered that they could measure distances using certain kinds of intrinsically luminous *variable stars*, such as cepheids, which can be seen at very large distances (Figure 26.11).

After the variables in nearby galaxies had been used to make distance measurements for a few decades, Walter Baade showed that there were actually two kinds of cepheids and that astronomers had been unwittingly mixing them up. As a result, in the early 1950s, the distances to all of the galaxies had to be increased by about a factor of two. We mention this because we want you to bear in mind, as you read on, that science is always a study in progress. Our first tentative steps in such difficult investigations are always subject to future revision as our techniques become more reliable.

The amount of work involved in finding cepheids and measuring their periods can be enormous. Hubble, for example, obtained 350 long-exposure photographs of the Andromeda galaxy over a period of 18 years and was able to identify only 40 cepheids. Even though cepheids are fairly luminous stars, they can be detected in only about 30 of the nearest galaxies with the world's largest ground-based telescopes.

As mentioned in Celestial Distances, one of the main projects carried out during the first years of operation of the Hubble Space Telescope was the measurement of cepheids in more distant galaxies to improve the accuracy of the extragalactic distance scale. Recently, astronomers working with the Hubble Space Telescope have extended such measurements out to 108 million light-years—a triumph of technology and determination.

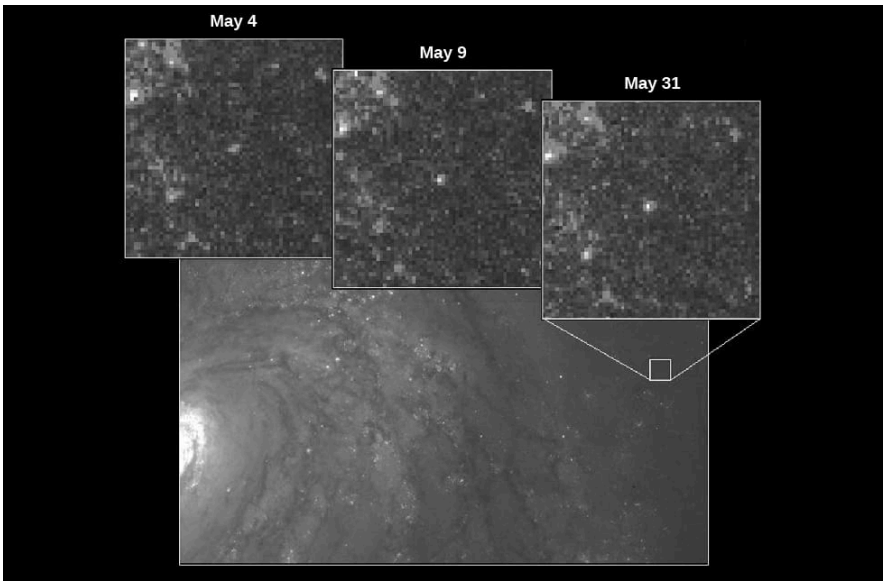


Figure 26.11 : Cepheid Variable Star. In 1994, using the Hubble Space Telescope, astronomers were able to make out an individual cepheid variable star in the galaxy M100 and measure its distance to be 56 million light-years. The insets show the star on three different nights; you can see that its brightness is indeed variable. (credit: modification of work by Wendy L. Freedman, Observatories of the Carnegie Institution of Washington, and NASA/ESA)

Nevertheless, we can only use cepheids to measure distances within a small fraction of the universe of galaxies. After all, to use this method, we must be able to resolve single stars and follow their subtle variations. Beyond a certain distance, even our finest space telescopes cannot help us do this. Fortunately, there are other ways to measure the distances to galaxies.

Standard Bulbs

We discussed in Celestial Distances the great frustration that astronomers felt when they realized that the stars in general were not standard *bulbs*. If every light bulb in a huge auditorium is a standard 100-watt bulb, then bulbs that look brighter to us must be closer, whereas those that look dimmer must be farther away. If every star were a standard luminosity (or wattage), then we could similarly "read off" their distances based on how bright they appear to us. Alas, as we have learned, neither stars nor galaxies come in one standard-issue luminosity. Nonetheless, astronomers have been searching for objects out there that do act in some way like a standard bulb—that have the same intrinsic (built-in) brightness wherever they are.

A number of suggestions have been made for what sorts of objects might be effective standard bulbs, including the brightest supergiant stars, planetary nebulae (which give off a lot of ultraviolet radiation), and the average globular cluster in a galaxy. One object turns out to be particularly useful: the type Ia supernova. These supernovae involve the explosion of a white dwarf in a binary system (see The Evolution of Binary Star Systems) Observations show that supernovae of this type all reach nearly the same luminosity (about $4.5 \times 10^9 L_{Sun}$) at maximum light. With such tremendous luminosities, these supernovae have been detected out to a distance of more than 8 billion light-years and are therefore especially attractive to astronomers as a way of determining distances on a large scale (Figure 26.12).



Figure 26.12 : Type Ia Supernova. The bright object at the bottom left of center is a type Ia supernova near its peak intensity. The supernova easily outshines its host galaxy. This extreme increase and luminosity help astronomers use Ia supernova as standard bulbs. (credit: NASA, ESA, A. Riess (STScI))

Several other kinds of standard bulbs visible over great distances have also been suggested, including the overall brightness of, for example, giant ellipticals and the brightest member of a galaxy cluster. Type Ia supernovae, however, have proved to be the most accurate standard bulbs, and they can be seen in more distant galaxies than the other types of calibrators. As we will see in the chapter on The Big Bang, observations of this type of supernova have profoundly changed our understanding of the evolution of the universe.

Other Measuring Techniques

Another technique for measuring galactic distances makes use of an interesting relationship noticed in the late 1970s by Brent Tully of the University of Hawaii and Richard Fisher of the National Radio Astronomy Observatory. They discovered that the luminosity of a spiral galaxy is related to its rotational velocity (how fast it spins). Why would this be true?

The more mass a galaxy has, the faster the objects in its outer regions must orbit. A more massive galaxy has more stars in it and is thus more luminous (ignoring dark matter for a moment). Thinking back to our discussion from the previous section, we can say that if the mass-to-light ratios for various spiral galaxies are pretty similar, then we can estimate the luminosity of a spiral galaxy by measuring its mass, and we can estimate its mass by measuring its rotational velocity.

Tully and Fisher used the 21-cm line of cold hydrogen gas to determine how rapidly material in spiral galaxies is orbiting their centers (you can review our discussion of the 21-cm line in Between the Stars: Gas and Dust in Space). Since 21-cm radiation from stationary atoms comes in a nice narrow line, the width of the 21-cm line produced by a whole rotating galaxy tells us the range of orbital velocities of the galaxy's hydrogen gas. The broader the line, the faster the gas is orbiting in the galaxy, and the more massive and luminous the galaxy turns out to be.

It is somewhat surprising that this technique works, since much of the mass associated with galaxies is dark matter, which does not contribute at all to the luminosity but does affect the rotation speed. There is also no obvious reason why the mass-to-light ratio should be similar for all spiral galaxies. Nevertheless, observations of nearer galaxies (where we have other ways of measuring distance) show that measuring the rotational velocity of a galaxy provides an accurate estimate of its intrinsic luminosity. Once we know how luminous the galaxy really is, we can compare the luminosity to the apparent brightness and use the difference to calculate its distance.

While the Tully-Fisher relation works well, it is limited—we can only use it to determine the distance to a spiral galaxy. There are other methods that can be used to estimate the distance to an elliptical galaxy; however, those methods are beyond the scope of our introductory astronomy course.

Table 26.2 lists the type of galaxy for which each of the distance techniques is useful, and the range of distances over which the

technique can be applied.

Table 26.2: Some Methods for Estimating Distance to Galaxies

Method	Galaxy Type	Approximate Distance Range (millions of light-years)
Planetary nebulae	All	0–70
Cepheid variables	Spiral, irregulars	0–110
Tully-Fisher relation	Spiral	0–300
Type Ia supernovae	All	0–11,000
Redshifts (Hubble's law)	All	300–13,000

This page titled 16.2: The Extragalactic Distance Scale is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 26.5: The Extragalactic Distance Scale by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.3: The Expanding Universe

Learning Objectives

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

- Describe the discovery that galaxies getting farther apart as the universe evolves
- Explain how to use Hubble's law to determine distances to remote galaxies
- Describe models for the nature of an expanding universe
- Explain the variation in Hubble's constant

We now come to one of the most important discoveries ever made in astronomy—the fact that the universe is expanding. Before we describe how the discovery was made, we should point out that the first steps in the study of galaxies came at a time when the techniques of spectroscopy were also making great strides. Astronomers using large telescopes could record the spectrum of a faint star or galaxy on photographic plates, guiding their telescopes so they remained pointed to the same object for many hours and collected more light. The resulting spectra of galaxies contained a wealth of information about the composition of the galaxy and the velocities of these great star systems.

Slipher's Pioneering Observations

Curiously, the discovery of the expansion of the universe began with the search for Martians and other solar systems. In 1894, the controversial (and wealthy) astronomer Percival Lowell established an observatory in Flagstaff, Arizona, to study the planets and search for life in the universe. Lowell thought that the spiral nebulae might be solar systems in the process of formation. He therefore asked one of the observatory's young astronomers, Vesto M. Slipher (Figure 26.13), to photograph the spectra of some of the spiral nebulae to see if their spectral lines might show chemical compositions like those expected for newly forming planets.



Figure 26.13 : Vesto M. Slipher (1875–1969). Slipher spent his entire career at the Lowell Observatory, where he discovered the large radial velocities of galaxies. (credit: Lowell Observatory)

The Lowell Observatory's major instrument was a 24-inch refracting telescope, which was not at all well suited to observations of faint spiral nebulae. With the technology available in those days, photographic plates had to be exposed for 20 to 40 hours to produce a good spectrum (in which the positions of the lines could reveal a galaxy's motion). This often meant continuing to expose the same photograph over several nights. Beginning in 1912, and making heroic efforts over a period of about 20 years, Slipher managed to photograph the spectra of more than 40 of the spiral nebulae (which would all turn out to be galaxies).

To his surprise, the spectral lines of most galaxies showed an astounding redshift. By “redshift” we mean that the lines in the spectra are displaced toward longer wavelengths (toward the red end of the visible spectrum). Recall from the chapter on Radiation and Spectra that a redshift is seen when the source of the waves is moving away from us. Slipher’s observations showed that most spirals are racing away at huge speeds; the highest velocity he measured was 1800 kilometers per second.

Only a few spirals—such as the Andromeda and Triangulum Galaxies and M81—all of which are now known to be our close neighbors, turned out to be approaching us. All the other galaxies were moving away. Slipher first announced this discovery in 1914, years before Hubble showed that these objects were other galaxies and before anyone knew how far away they were. No one at the time quite knew what to make of this discovery.

Hubble’s Law

The profound implications of Slipher’s work became apparent only during the 1920s. Georges Lemaître was a Belgian priest and a trained astronomer. In 1927, he published a paper in French in an obscure Belgian journal in which he suggested that we live in an expanding universe. The title of the paper (translated into English) is “A Homogenous Universe of Constant Mass and Growing Radius Accounting for the Radial Velocity of Extragalactic Nebulae.” Lemaître had discovered that Einstein’s equations of relativity were consistent with an expanding universe (as had the Russian scientist Alexander Friedmann independently in 1922). Lemaître then went on to use Slipher’s data to support the hypothesis that the universe actually is expanding and to estimate the rate of expansion. Initially, scientists paid little attention to this paper, perhaps because the Belgian journal was not widely available.

In the meantime, Hubble was making observations of galaxies with the 2.5-meter telescope on Mt. Wilson, which was then the world’s largest. Hubble carried out the key observations in collaboration with a remarkable man, Milton Humason, who dropped out of school in the eighth grade and began his astronomical career by driving a mule train up the trail on Mount Wilson to the observatory (Figure 26.14). In those early days, supplies had to be brought up that way; even astronomers hiked up to the mountaintop for their turns at the telescope. Humason became interested in the work of the astronomers and, after marrying the daughter of the observatory’s electrician, took a job as janitor there. After a time, he became a night assistant, helping the astronomers run the telescope and record data. Eventually, he made such a mark that he became a full astronomer at the observatory.

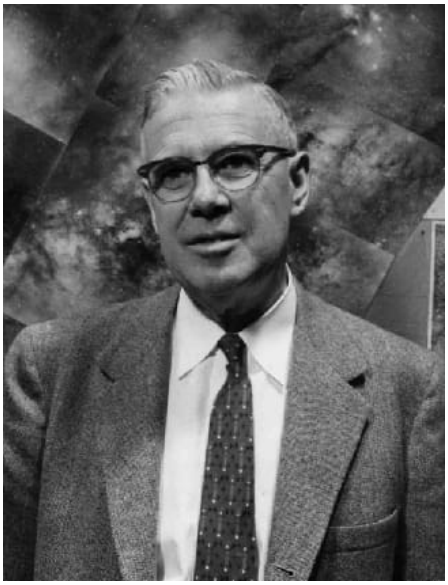


Figure 26.14 : Milton Humason (1891–1972). Humason was Hubble’s collaborator on the great task of observing, measuring, and classifying the characteristics of many galaxies. (credit: Caltech Archives)

By the late 1920s, Humason was collaborating with Hubble by photographing the spectra of faint galaxies with the 2.5-meter telescope. (By then, there was no question that the spiral nebulae were in fact galaxies.) Hubble had found ways to improve the accuracy of the estimates of distances to spiral galaxies, and he was able to measure much fainter and more distant galaxies than Slipher could observe with his much-smaller telescope. When Hubble laid his own distance estimates next to measurements of the recession velocities (the speed with which the galaxies were moving away), he found something stunning: there was a relationship between distance and velocity for galaxies. *The more distant the galaxy, the faster it was receding from us*.

In 1931, Hubble and Humason jointly published the seminal paper where they compared distances and velocities of remote galaxies moving away from us at speeds as high as 20,000 kilometers per second and were able to show that the recession velocities of galaxies are directly proportional to their distances from us (Figure 26.15), just as Lemaître had suggested.

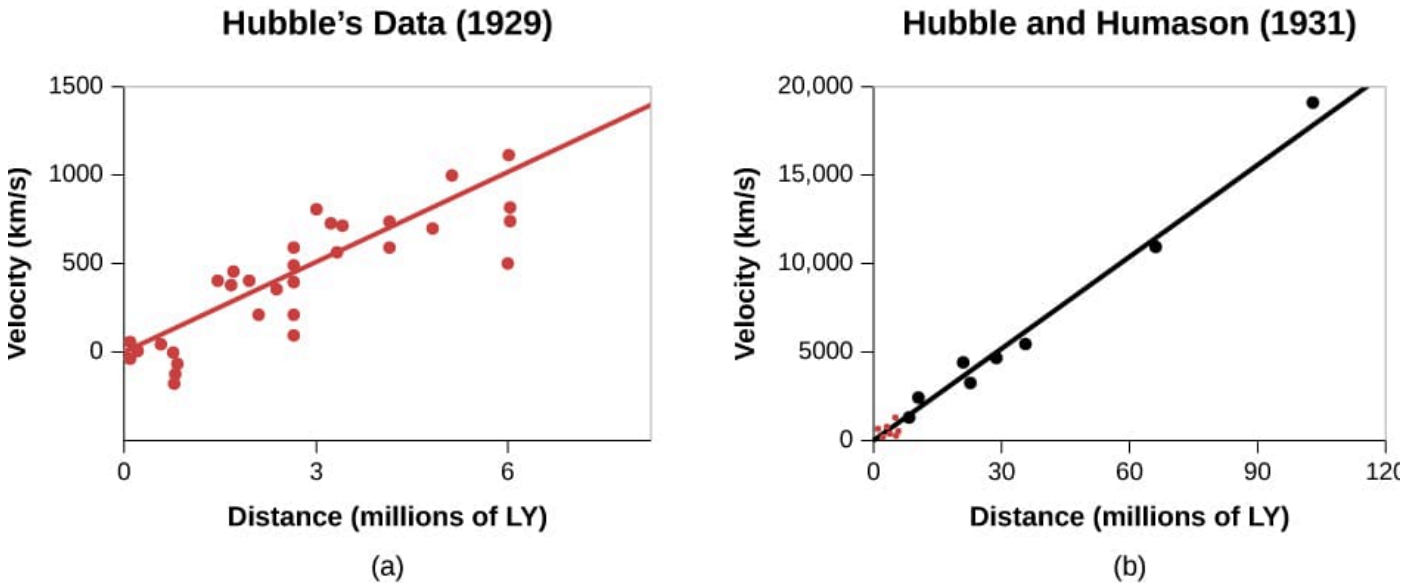


Figure 26.15 : Hubble's Law. (a) These data show Hubble's original velocity-distance relation, adapted from his 1929 paper in the *Proceedings of the National Academy of Sciences*. (b) These data show Hubble and Humason's velocity-distance relation, adapted from their 1931 paper in *The Astrophysical Journal*. The red dots at the lower left are the points in the diagram in the 1929 paper. Comparison of the two graphs shows how rapidly the determination of galactic distances and redshifts progressed in the 2 years between these publications.

We now know that this relationship holds for every galaxy except a few of the nearest ones. Nearly all of the galaxies that are approaching us turn out to be part of the Milky Way's own group of galaxies, which have their own individual motions, just as birds flying in a group may fly in slightly different directions at slightly different speeds even though the entire flock travels through space together.

Written as a formula, the relationship between velocity and distance is

$$v =$$

where *v* is the velocity, *d* is the distance, and *H* is a number called the Hubble constant. This equation is now known as Hubble's law.[1]

Astronomy Basics: Constants of Proportionality

Mathematical relationships such as Hubble's law are pretty common in life. To take a simple example, suppose your college or university hires you to call rich alumni and ask for donations. You are paid $2.50 for each call; the more calls you can squeeze in between studying astronomy and other courses, the more money you take home. We can set up a formula that connects *p*, your pay, and *n*, the number of calls

$$p = A \times n$$

where *A* is the alumni constant, with a value of $2.50. If you make 20 calls, you will earn $2.50 times 20, or $50.

Suppose your boss forgets to tell you what you will get paid for each call. You can calculate the alumni constant that governs your pay by keeping track of how many calls you make and noting your gross pay each week. If you make 100 calls the first week and are paid $250, you can deduce that the constant is $2.50 (in units of dollars per call). Hubble, of course, had no "boss" to tell him what his constant would be—he *had* to calculate its value from the measurements of distance and velocity.

Astronomers express the value of Hubble's constant in units that relate to how they measure speed and distance for galaxies. In this book, we will use kilometers per second per million light-years as that unit. For many years, estimates of the value of the

Hubble constant have been in the range of 15 to 30 kilometers per second per million light-years The most recent work appears to be converging on a value near 22 kilometers per second per million light-years If H is 22 kilometers per second per million light-years, a galaxy moves away from us at a speed of 22 kilometers per second for every million light-years of its distance. As an example, a galaxy 100 million light-years away is moving away from us at a speed of 2200 kilometers per second.

Hubble's law tells us something fundamental about the universe. Since all but the nearest galaxies appear to be in motion away from us, with the most distant ones moving the fastest, we must be living in an expanding universe. We will explore the implications of this idea shortly, as well as in the final chapters of this text. For now, we will just say that Hubble's observation underlies all our theories about the origin and evolution of the universe.

Hubble's Law and Distances

The regularity expressed in Hubble's law has a built-in bonus: it gives us a new way to determine the distances to remote galaxies. First, we must reliably establish Hubble's constant by measuring both the distance and the velocity of many galaxies in many directions to be sure Hubble's law is truly a universal property of galaxies. But once we have calculated the value of this constant and are satisfied that it applies everywhere, much more of the universe opens up for distance determination. Basically, if we can obtain a spectrum of a galaxy, we can immediately tell how far away it is.

The procedure works like this. We use the spectrum to measure the speed with which the galaxy is moving away from us. If we then put this speed and the Hubble constant into Hubble's law equation, we can solve for the distance.

Example 26.1: Hubble's Law

Hubble's law ($v = H \times d$) allows us to calculate the distance to any galaxy. Here is how we use it in practice.

We have measured Hubble's constant to be 22 km/s per million light-years. This means that if a galaxy is 1 million light-years farther away, it will move away 22 km/s faster. So, if we find a galaxy that is moving away at 18,000 km/s, what does Hubble's law tells us about the distance to the galaxy?

Solution

$$d = \frac{v}{H} = \frac{18{,}000\ \text{km/s}}{\frac{22\ \text{km/s}}{1\ \text{million light-years}}} = \frac{18{,}000}{22} \times \frac{1\ \text{million light-years}}{1} = 818\ \text{million light-years}$$

Note how we handled the units here: the km/s in the numerator and denominator cancel, and the factor of million light-years in the denominator of the constant must be divided correctly before we get our distance of 818 million light-years.

Exercise 16.3.1

Using 22 km/s/million light-years for Hubble's constant, what recessional velocity do we expect to find if we observe a galaxy at 500 million light-years?

Answer available at the end of this section.

Variation of Hubble's Constant

The use of redshift is potentially a very important technique for determining distances because as we have seen, most of our methods for determining galaxy distances are limited to approximately the nearest few hundred million light-years (and they have large uncertainties at these distances). The use of Hubble's law as a distance indicator requires only a spectrum of a galaxy and a measurement of the Doppler shift, and with large telescopes and modern spectrographs, spectra can be taken of extremely faint galaxies.

But, as is often the case in science, things are not so simple. This technique works if, and only if, the Hubble constant has been truly constant throughout the entire life of the universe. When we observe galaxies billions of light-years away, we are seeing them

as they were billions of years ago. What if the Hubble "constant" was different billions of years ago? Before 1998, astronomers thought that, although the universe is expanding, the expansion should be slowing down, or decelerating, because the overall gravitational pull of all matter in the universe would have a dominant, measureable effect. If the expansion is decelerating, then the Hubble constant should be decreasing over time.

The discovery that type Ia supernovae are standard bulbs gave astronomers the tool they needed to observe extremely distant galaxies and measure the rate of expansion billions of years ago. The results were completely unexpected. It turns out that the expansion of the universe is *accelerating* over time! What makes this result so astounding is that there is no way that existing physical theories can account for this observation. While a decelerating universe could easily be explained by gravity, there was no force or property in the universe known to astronomers that could account for the acceleration. In The Big Bang chapter, we will look in more detail at the observations that led to this totally unexpected result and explore its implications for the ultimate fate of the universe.

In any case, if the Hubble constant is not really a constant when we look over large spans of space and time, then the calculation of galaxy distances using the Hubble constant won't be accurate. As we shall see in the chapter on The Big Bang, the accurate calculation of distances requires a model for how the Hubble constant has changed over time. The farther away a galaxy is (and the longer ago we are seeing it), the more important it is to include the effects of the change in the Hubble constant. For galaxies within a few billion light-years, however, the assumption that the Hubble constant is indeed constant gives good estimates of distance.

Models for an Expanding Universe

At first, thinking about Hubble's law and being a fan of the work of Copernicus and Harlow Shapley, you might be shocked. Are all the galaxies really moving *away from us*? Is there, after all, something special about our position in the universe? Worry not; the fact that galaxies are receding from us and that more distant galaxies are moving away more rapidly than nearby ones shows only that the universe is expanding uniformly.

A uniformly expanding universe is one that is expanding at the same rate everywhere. In such a universe, we and all other observers, no matter where they are located, must observe a proportionality between the velocities and distances of equivalently remote galaxies. (Here, we are ignoring the fact that the Hubble constant is not constant over all time, but if at any given time in the evolution of the universe the Hubble constant has the same value everywhere, this argument still works.)

To see why, first imagine a ruler made of stretchable rubber, with the usual lines marked off at each centimeter. Now suppose someone with strong arms grabs each end of the ruler and slowly stretches it so that, say, it doubles in length in 1 minute (Figure 26.16). Consider an intelligent ant sitting on the mark at 2 centimeters—a point that is not at either end nor in the middle of the ruler. He measures how fast other ants, sitting at the 4-, 7-, and 12-centimeter marks, move away from him as the ruler stretches.

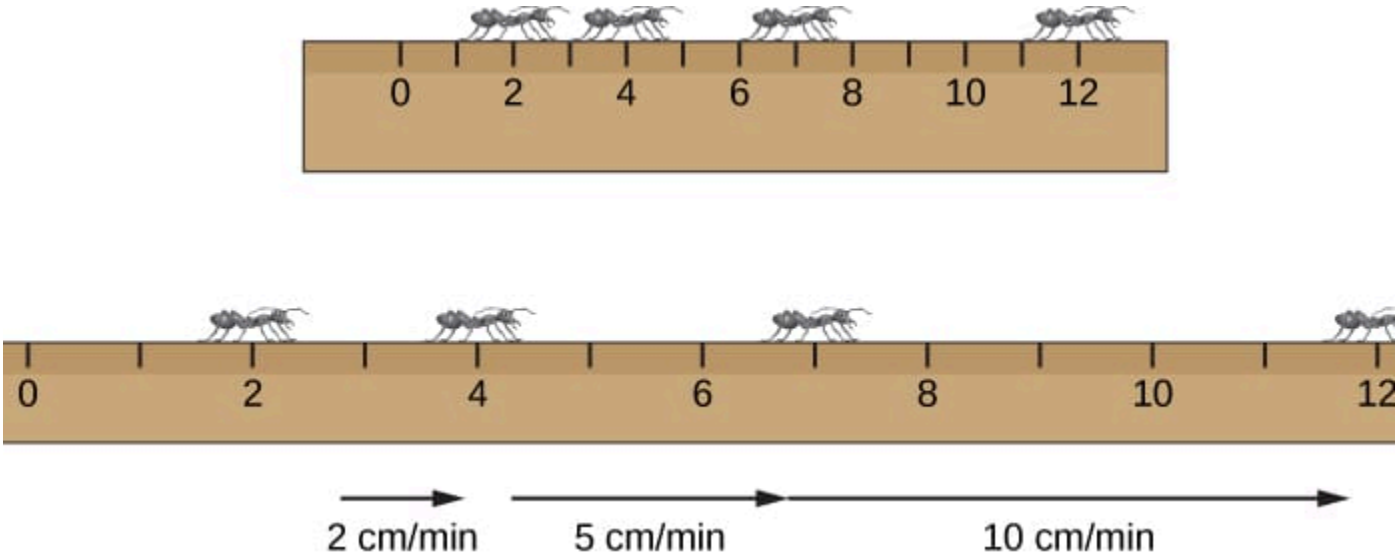


Figure 26.16 : Stretching a Ruler. Ants on a stretching ruler see other ants move away from them. The speed with which another ant moves away is proportional to its distance.

The ant at 4 centimeters, originally 2 centimeters away from our ant, has doubled its distance in 1 minute; it therefore moved away at a speed of 2 centimeters per minute. The ant at the 7-centimeters mark, which was originally 5 centimeters away from our ant, is now 10 centimeters away; it thus had to move at 5 centimeters per minute. The one that started at the 12-centimeters mark, which was 10 centimeters away from the ant doing the counting, is now 20 centimeters away, meaning it must have raced away at a speed of 10 centimeters per minute. Ants at different distances move away at different speeds, and their speeds are proportional to their distances (just as Hubble's law indicates for galaxies). Yet, notice in our example that all the ruler was doing was stretching uniformly. Also, notice that none of the ants were actually moving of their own accord, it was the stretching of the ruler that moved them apart.

Now let's repeat the analysis, but put the intelligent ant on some other mark—say, on 7 or 12 centimeters. We discover that, as long as the ruler stretches uniformly, this ant also finds every other ant moving away at a speed proportional to its distance. In other words, the kind of relationship expressed by Hubble's law can be explained by a uniform stretching of the "world" of the ants. And all the ants in our simple diagram will see the other ants moving away from them as the ruler stretches.

For a three-dimensional analogy, let's look at the loaf of raisin bread in Figure 26.17. The chef has accidentally put too much yeast in the dough, and when she sets the bread out to rise, it doubles in size during the next hour, causing all the raisins to move farther apart. On the figure, we again pick a representative raisin (that is not at the edge or the center of the loaf) and show the distances from it to several others in the figure (before and after the loaf expands).

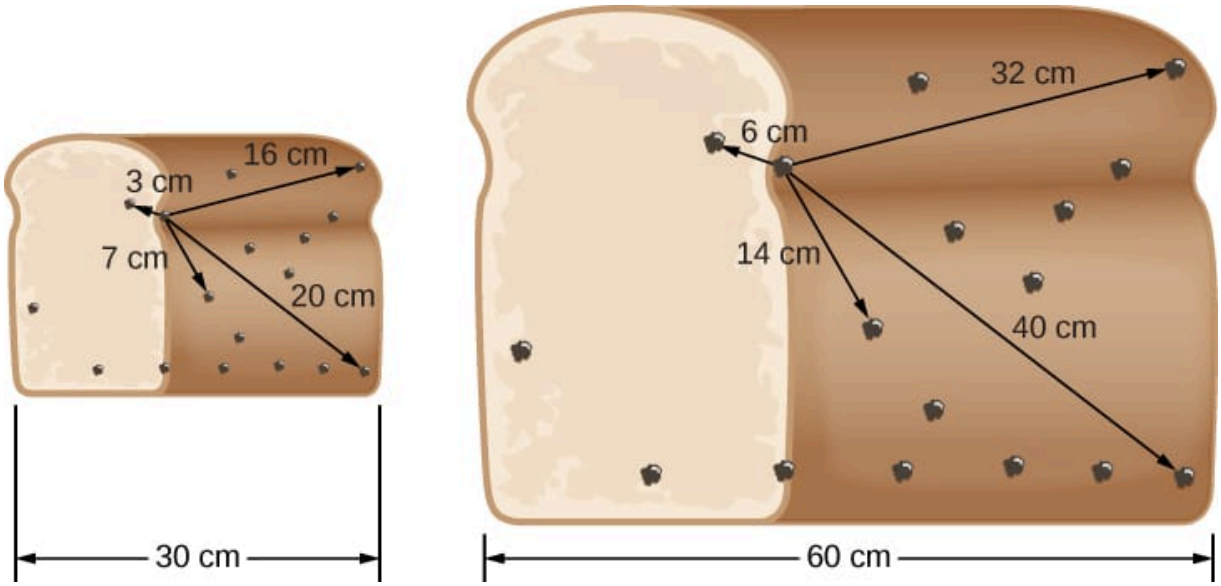


Figure 26.17 : Expanding Raisin Bread. As the raisin bread rises, the raisins "see" other raisins moving away. More distant raisins move away faster in a uniformly expanding bread.

Measure the increases in distance and calculate the speeds for yourself on the raisin bread, just like we did for the ruler. You will see that, since each distance doubles during the hour, each raisin moves away from our selected raisin at a speed proportional to its distance. The same is true no matter which raisin you start with.

Our two analogies are useful for clarifying our thinking, but you must not take them literally. On both the ruler and the raisin bread, there are points that are at the end or edge. You can use these to pinpoint the middle of the ruler and the loaf. While our models of the universe have some resemblance to the properties of the ruler and the loaf, the universe has no boundaries, no edges, and no center (all mind-boggling ideas that we will discuss in a later chapter).

What is useful to notice about both the ants and the raisins is that they themselves did not "cause" their motion. It isn't as if the raisins decided to take a trip away from each other and then hopped on a hoverboard to get away. No, in both our analogies, it was the stretching of the medium (the ruler or the bread) that moved the ants or the raisins farther apart. In the same way, we will see in The Big Bang chapter that the galaxies don't have rocket motors propelling them away from each other. Instead, they are passive participants in the *expansion of space*. As space stretches, the galaxies are carried farther and farther apart much as the ants and the raisins were. (If this notion of the "stretching" of space surprises or bothers you, now would be a good time to review the information about spacetime in Black Holes and Curved Spacetime. We will discuss these ideas further as our discussion broadens from galaxies to the whole universe.)

The expansion of the universe, by the way, does not imply that the individual galaxies and clusters of galaxies themselves are expanding. Neither raisins nor the ants in our analogy grow in size as the loaf expands. Similarly, gravity holds galaxies and clusters of galaxies together, and they get farther away from each other—without themselves changing in size—as the universe expands.

Footnotes

- 1In 2020, the International Astronomical Union suggested that it would be more fair to call it the Hubble-Lemaitre law. In telling the history in this textbook, we have acknowledged the role Lemaître played, and we urge our readers to keep his contributions in mind as they read on.

Answers

Exercise 16.3.1

$$v = d \times H = 500\,\text{million light-years} \times \frac{22\,\text{km/s}}{1\,\text{million light-years}} = 11{,}000\,\text{km/s}$$

This page titled 16.3: The Expanding Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 26.6: The Expanding Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.4: Types of Galaxies

Learning Objectives

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

- Describe the properties and features of elliptical, spiral, and irregular galaxies
- Explain what may cause a galaxy's appearance to change over time

Having established the existence of other galaxies, Hubble and others began to observe them more closely—noting their shapes, their contents, and as many other properties as they could measure. This was a daunting task in the 1920s when obtaining a single photograph or spectrum of a galaxy could take a full night of tireless observing. Today, larger telescopes and electronic detectors have made this task less difficult, although observing the most distant galaxies (those that show us the universe in its earliest phases) still requires enormous effort.

The first step in trying to understand a new type of object is often simply to describe it. Remember, the first step in understanding stellar spectra was simply to sort them according to appearance (see Analyzing Starlight). As it turns out, the biggest and most luminous galaxies come in one of two basic shapes: either they are flatter and have spiral arms, like our own Galaxy, or they appear to be elliptical (blimp- or cigar-shaped). Many smaller galaxies, in contrast, have an irregular shape.

Spiral Galaxies

Our own Galaxy and the Andromeda galaxy are typical, large spiral galaxies (see Figure 26.2). They consist of a central bulge, a halo, a disk, and spiral arms. Interstellar material is usually spread throughout the disks of spiral galaxies. Bright emission nebulae and hot, young stars are present, especially in the spiral arms, showing that new star formation is still occurring. The disks are often dusty, which is especially noticeable in those systems that we view almost edge on (Figure 26.4).

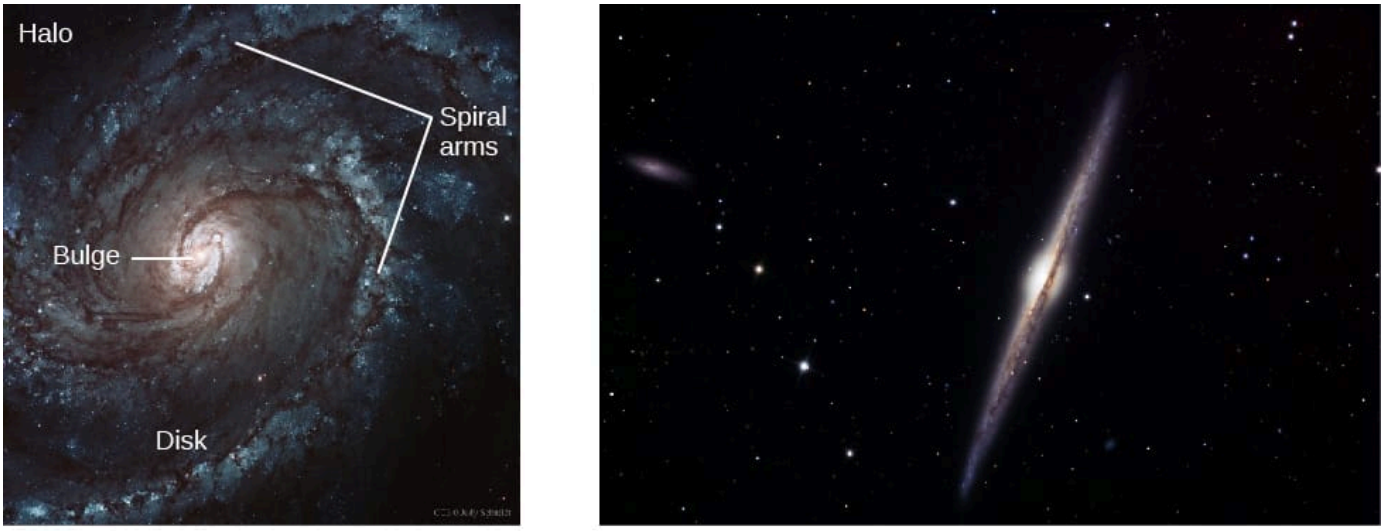


Figure 26.4 : Spiral Galaxies. (a) The spiral arms of M100, shown here, are bluer than the rest of the galaxy, indicating young, high-mass stars and star-forming regions. (b) We view this spiral galaxy, NGC 4565, almost exactly edge on, and from this angle, we can see the dust in the plane of the galaxy; it appears dark because it absorbs the light from the stars in the galaxy. (credit a: modification of work by Hubble Legacy Archive, NASA, ESA, and Judy Schmidt; credit b: modification of work by "Jschulman555"/ Wikimedia)

In galaxies that we see face on, the bright stars and emission nebulae make the arms of spirals stand out like those of a pinwheel on the fourth of July. Open star clusters can be seen in the arms of nearer spirals, and globular clusters are often visible in their halos. Spiral galaxies contain a mixture of young and old stars, just as the Milky Way does. All spirals rotate, and the direction of

their spin is such that the arms appear to trail much like the wake of a boat.

About two-thirds of the nearby spiral galaxies have boxy or peanut-shaped bars of stars running through their centers (Figure 26.5). Showing great originality, astronomers call these galaxies barred spirals.



Figure 26.5 : Barred Spiral Galaxy. NGC 1300, shown here, is a barred spiral galaxy. Note that the spiral arms begin at the ends of the bar. (credit: NASA, ESA, and the Hubble Heritage Team(STScI/AURA))

As we noted in The Milky Way Galaxy chapter, our Galaxy has a modest bar too (see Figure 25.10). The spiral arms usually begin from the ends of the bar. The fact that bars are so common suggests that they are long lived; it may be that most spiral galaxies form a bar at some point during their evolution.

In both barred and unbarred spiral galaxies, we observe a range of different shapes. At one extreme, the central bulge is large and luminous, the arms are faint and tightly coiled, and bright emission nebulae and supergiant stars are inconspicuous. Hubble, who developed a system of classifying galaxies by shape, gave these galaxies the designation Sa. Galaxies at this extreme may have no clear spiral arm structure, resulting in a lens-like appearance (they are sometimes referred to as lenticular galaxies). These galaxies seem to share as many properties with elliptical galaxies as they do with spiral galaxies

At the other extreme, the central bulge is small and the arms are loosely wound. In these Sc galaxies, luminous stars and emission nebulae are very prominent. Our Galaxy and the Andromeda galaxy are both intermediate between the two extremes. Photographs of spiral galaxies, illustrating the different types, are shown in Figure 26.6, along with elliptical galaxies for comparison.

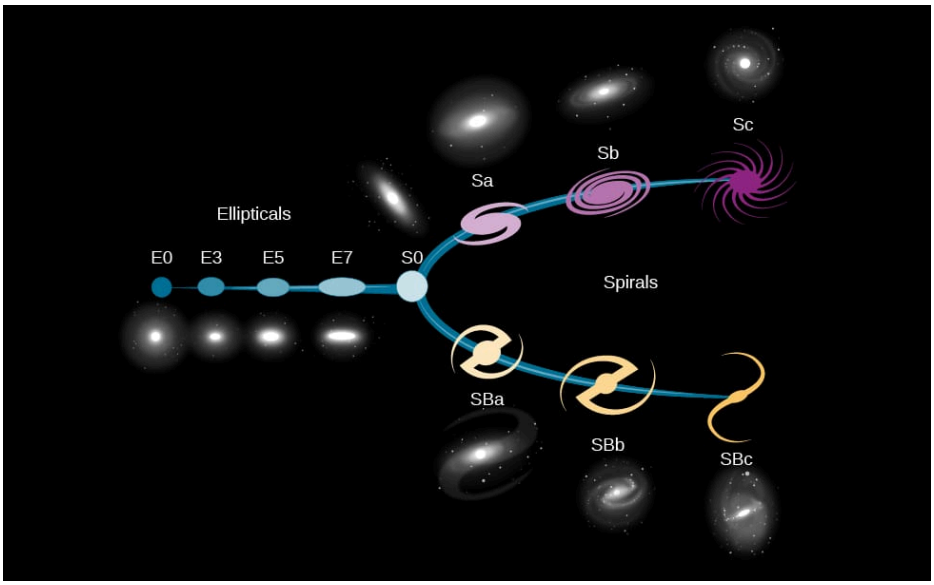


Figure 26.6 : Hubble Classification of Galaxies. This figure shows Edwin Hubble's original classification of galaxies. Elliptical galaxies are on the left. On the right, you can see the basic spiral shapes illustrated, alongside images of actual barred and unbarred spirals. (credit: modification of work by NASA, ESA)

The luminous parts of spiral galaxies appear to range in diameter from about 20,000 to more than 100,000 light-years. Recent studies have found that there is probably a large amount of galactic material that extends well beyond the apparent edge of galaxies. This material appears to be thin, cold gas that is difficult to detect in most observations.

From the observational data available, the masses of the visible portions of spiral galaxies are estimated to range from 1 billion to 1 trillion Suns (10^9 to 10^{12} M_{Sun}). The total luminosities of most spirals fall in the range of 100 million to 100 billion times the luminosity of our Sun (10^8 to 10^{11} L_{Sun}). Our Galaxy and M31 are relatively large and massive, as spirals go. There is also considerable dark matter in and around the galaxies, just as there is in the Milky Way; we deduce its presence from how fast stars in the outer parts of the Galaxy are moving in their orbits.

Elliptical Galaxies

Elliptical galaxies consist almost entirely of old stars and have shapes that are spheres or ellipsoids (somewhat squashed spheres) (Figure 26.7). They contain no trace of spiral arms. Their light is dominated by older reddish stars (the population II stars discussed in The Milky Way Galaxy). In the larger nearby ellipticals, many globular clusters can be identified. Dust and emission nebulae are not conspicuous in elliptical galaxies, but many do contain a small amount of interstellar matter.



Figure 26.7 : Elliptical Galaxies. (a) ESO 325-G004 is a giant elliptical galaxy. Other elliptical galaxies can be seen around the edges of this image. (b) This elliptical galaxy probably originated from the collision of two spiral galaxies. (credit a: modification of work by NASA, ESA, and The Hubble Heritage Team (STScI/AURA); credit b: modification of work by ESA/Hubble, NASA)

Elliptical galaxies show various degrees of flattening, ranging from systems that are approximately spherical to those that approach the flatness of spirals. The rare giant ellipticals (for example, ESO 325-G004 in Figure 26.7) reach luminosities of 10^{11} L_{Sun}. The mass in a giant elliptical can be as large as 10^{13} M_{Sun}. The diameters of these large galaxies extend over several hundred thousand light-years and are considerably larger than the largest spirals. Although individual stars orbit the center of an elliptical galaxy, the orbits are not all in the same direction, as occurs in spirals. Therefore, ellipticals don't appear to rotate in a systematic way, making it difficult to estimate how much dark matter they contain.

We find that elliptical galaxies range all the way from the giants, just described, to dwarfs, which may be the most common kind of galaxy. *Dwarf ellipticals* (sometimes called dwarf spheroidals) escaped our notice for a long time because they are very faint and difficult to see. An example of a dwarf elliptical is the Leo I Dwarf Spheroidal galaxy shown in Figure 26.8. The luminosity of this typical dwarf is about equal to that of the brightest globular clusters.

Intermediate between the giant and dwarf elliptical galaxies are systems such as M32 and M110, the two companions of the Andromeda galaxy. While they are often referred to as dwarf ellipticals, these galaxies are significantly larger than galaxies such as Leo I.



Figure 26.8 : Dwarf Elliptical Galaxy. M32, a dwarf elliptical galaxy and one of the companions to the giant Andromeda galaxy M31. M32 is a dwarf by galactic standards, as it is only 2400 light-years across. (credit: NOAO/AURA/NSF)

Irregular Galaxies

Hubble classified galaxies that do not have the regular shapes associated with the categories we just described into the catchall bin of an irregular galaxy, and we continue to use his term. Typically, irregular galaxies have lower masses and luminosities than spiral galaxies. Irregular galaxies often appear disorganized, and many are undergoing relatively intense star formation activity. They contain both young population I stars and old population II stars.

The two best-known irregular galaxies are the Large Magellanic Cloud and Small Magellanic Cloud (Figure 26.9), which are at a distance of a little more than 160,000 light-years away and are among our nearest extragalactic neighbors. Their names reflect the fact that Ferdinand Magellan and his crew, making their round-the-world journey, were the first European travelers to notice them. Although not visible from the United States and Europe, these two systems are prominent from the Southern Hemisphere, where they look like wispy clouds in the night sky. Since they are only about one-tenth as distant as the Andromeda galaxy, they present an excellent opportunity for astronomers to study nebulae, star clusters, variable stars, and other key objects in the setting of another galaxy. For example, the Large Magellanic Cloud contains the 30 Doradus complex (also known as the Tarantula Nebula), one of the largest and most luminous groups of supergiant stars known in any galaxy.



Figure 26.9 : 4-Meter Telescope at Cerro Tololo Inter-American Observatory Silhouetted against the Southern Sky. The Milky Way is seen to the right of the dome, and the Large and Small Magellanic Clouds are seen to the left. (credit: Roger Smith/NOAO/AURA/NSF)

The Small Magellanic Cloud is considerably less massive than the Large Magellanic Cloud, and it is six times longer than it is wide. This narrow wisp of material points directly toward our Galaxy like an arrow. The Small Magellanic Cloud was most likely contorted into its current shape through gravitational interactions with the Milky Way. A large trail of debris from this interaction between the Milky Way and the Small Magellanic Cloud has been strewn across the sky and is seen as a series of gas clouds moving at abnormally high velocity, known as the Magellanic Stream. We will see that this kind of interaction between galaxies will help explain the irregular shapes of this whole category of small galaxies,

Link to Learning

View this beautiful album showcasing the different types of galaxies that have been photographed by the Hubble Space Telescope.

Galaxy Evolution

Encouraged by the success of the H-R diagram for stars (see Analyzing Starlight), astronomers studying galaxies hoped to find some sort of comparable scheme, where differences in appearance could be tied to different evolutionary stages in the life of galaxies. Wouldn't it be nice if every elliptical galaxy evolved into a spiral, for example, just as every main-sequence star evolves into a red giant? Several simple ideas of this kind were tried, some by Hubble himself, but none stood the test of time (and observation).

Because no simple scheme for evolving one type of galaxy into another could be found, astronomers then tended to the opposite point of view. For a while, most astronomers thought that all galaxies formed very early in the history of the universe and that the differences between them had to do with the rate of star formation. Ellipticals were those galaxies in which all the interstellar matter was converted rapidly into stars. Spirals were galaxies in which star formation occurred slowly over the entire lifetime of the galaxy. This idea turned out to be too simple as well.

Today, we understand that at least some galaxies have changed types over the billions of years since the universe began. As we shall see in later chapters, collisions and mergers between galaxies may dramatically change spiral galaxies into elliptical galaxies. Even isolated spirals (with no neighbor galaxies in sight) can change their appearance over time. As they consume their gas, the rate of star formation will slow down, and the spiral arms will gradually become less conspicuous. Over long periods, spirals therefore begin to look more like the galaxies at the middle of Figure 26.6 (which astronomers refer to as S0 types).

Over the past several decades, the study of how galaxies evolve over the lifetime of the universe has become one of the most active fields of astronomical research. We will discuss the evolution of galaxies in more detail in The Evolution and Distribution

of Galaxies, but let's first see in a little more detail just what different galaxies are like.

This page titled 16.4: Types of Galaxies is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 26.3: Types of Galaxies by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.5: Properties of Galaxies

Learning Objectives

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

- Describe the methods through which astronomers can estimate the mass of a galaxy
- Characterize each type of galaxy by its mass-to-light ratio

The technique for deriving the masses of galaxies is basically the same as that used to estimate the mass of the Sun, the stars, and our own Galaxy. We measure how fast objects in the outer regions of the galaxy are orbiting the center, and then we use this information along with Kepler's third law to calculate how much mass is inside that orbit.

Masses of Galaxies

Astronomers can measure the rotation speed in spiral galaxies by obtaining spectra of either stars or gas, and looking for wavelength shifts produced by the Doppler effect. Remember that the faster something is moving toward or away from us, the greater the shift of the lines in its spectrum. Kepler's law, together with such observations of the part of the Andromeda galaxy that is bright in visible light, for example, show it to have a galactic mass of about 4×10^{11} M_{Sun} (enough material to make 400 billion stars like the Sun).

The total mass of the Andromeda galaxy is greater than this, however, because we have not included the mass of the material that lies beyond its visible edge. Fortunately, there is a handful of objects—such as isolated stars, star clusters, and satellite galaxies—beyond the visible edge that allows astronomers to estimate how much additional matter is hidden out there. Recent studies show that the amount of dark matter beyond the visible edge of Andromeda may be as large as the mass of the bright portion of the galaxy. Indeed, using Kepler's third law and the velocities of its satellite galaxies, the Andromeda galaxy is estimated to have a mass closer to 1.4×10^{12} M_{Sun}. The mass of the Milky Way Galaxy is estimated to be 8.5×10^{11} M_{Sun}, and so our Milky Way is turning out to be somewhat smaller than Andromeda.

Elliptical galaxies do not rotate in a systematic way, so we cannot determine a rotational velocity; therefore, we must use a slightly different technique to measure their mass. Their stars are still orbiting the galactic center, but not in the organized way that characterizes spirals. Since elliptical galaxies contain stars that are billions of years old, we can assume that the galaxies themselves are not flying apart. Therefore, if we can measure the various speeds with which the stars are moving in their orbits around the center of the galaxy, we can calculate how much mass the galaxy must contain in order to hold the stars within it.

In practice, the spectrum of a galaxy is a composite of the spectra of its many stars, whose different motions produce different Doppler shifts (some red, some blue). The result is that the lines we observe from the entire galaxy contain the combination of many Doppler shifts. When some stars provide blueshifts and others provide redshifts, they create a wider or broader absorption or emission feature than would the same lines in a hypothetical galaxy in which the stars had no orbital motion. Astronomers call this phenomenon line broadening. The amount by which each line broadens indicates the range of speeds at which the stars are moving with respect to the center of the galaxy. The range of speeds depends, in turn, on the force of gravity that holds the stars within the galaxies. With information about the speeds, it is possible to calculate the mass of an elliptical galaxy.

Table 26.1 summarizes the range of masses (and other properties) of the various types of galaxies. Interestingly enough, the most and least massive galaxies are ellipticals. On average, irregular galaxies have less mass than spirals.

Table 26.1: Characteristics of the Different Types of Galaxies

Characteristic	Spirals	Ellipticals	Irregulars
Mass (M_{Sun})	10^9 to 10^{12}	10^5 to 10^{13}	10^8 to 10^{11}
Diameter (thousands of light-years)	15 to 150	3 to >700	3 to 30
Luminosity (L_{Sun})	10^8 to 10^{11}	10^6 to 10^{11}	10^7 to 2×10^9

Characteristic	Spirals	Ellipticals	Irregulars
Populations of stars	Old and young	Old	Old and young
Interstellar matter	Gas and dust	Almost no dust; little gas	Much gas; some have little dust, some much dust
Mass-to-light ratio in the visible part	2 to 10	10 to 20	1 to 10
Mass-to-light ratio for total galaxy	100	100	?

Mass-to-Light Ratio

A useful way of characterizing a galaxy is by noting the ratio of its mass (in units of the Sun's mass) to its light output (in units of the Sun's luminosity). This single number tells us roughly what kind of stars make up most of the luminous population of the galaxy, and it also tells us whether a lot of dark matter is present. For stars like the Sun, the mass-to-light ratio is 1 by our definition.

Galaxies are not, of course, composed entirely of stars that are identical to the Sun. The overwhelming majority of stars are less massive and less luminous than the Sun, and usually these stars contribute most of the mass of a system without accounting for very much light. The mass-to-light ratio for low-mass stars is greater than 1 (you can verify this using the data in Table 18.3). Therefore, a galaxy's mass-to-light ratio is also generally greater than 1, with the exact value depending on the ratio of high-mass stars to low-mass stars.

Galaxies in which star formation is still occurring have many massive stars, and their mass-to-light ratios are usually in the range of 1 to 10. Galaxies consisting mostly of an older stellar population, such as ellipticals, in which the massive stars have already completed their evolution and have ceased to shine, have mass-to-light ratios of 10 to 20.

But these figures refer only to the inner, conspicuous parts of galaxies (Figure 26.10). In The Milky Way Galaxy and above, we discussed the evidence for dark matter in the outer regions of our own Galaxy, extending much farther from the galactic center than do the bright stars and gas. Recent measurements of the rotation speeds of the outer parts of nearby galaxies, such as the Andromeda galaxy we discussed earlier, suggest that they too have extended distributions of dark matter around the visible disk of stars and dust. This largely invisible matter adds to the mass of the galaxy while contributing nothing to its luminosity, thus increasing the mass-to-light ratio. If dark invisible matter is present in a galaxy, its mass-to-light ratio can be as high as 100. The two different mass-to-light ratios measured for various types of galaxies are given in Table 26.1.



Figure 26.10 : M101, the Pinwheel Galaxy. This galaxy is a face-on spiral at a distance of 21 million light-years. M101 is almost twice the diameter of the Milky Way, and it contains at least 1 trillion stars. (credit: NASA, ESA, K. Kuntz (Johns Hopkins University), F. Bresolin (University of Hawaii), J. Trauger (Jet Propulsion Lab), J. Mould (NOAO), Y.-H. Chu (University of Illinois, Urbana), and STScI)

These measurements of other galaxies support the conclusion already reached from studies of the rotation of our own Galaxy—namely, that most of the material in the universe cannot at present be observed directly in any part of the electromagnetic spectrum. An understanding of the properties and distribution of this invisible matter is crucial to our understanding of galaxies. It's becoming clearer and clearer that, through the gravitational force it exerts, dark matter plays a dominant role in galaxy formation and early evolution. There is an interesting parallel here between our time and the time during which Edwin Hubble was receiving his training in astronomy. By 1920, many scientists were aware that astronomy stood on the brink of important breakthroughs—if only the nature and behavior of the nebulae could be settled with better observations. In the same way, many astronomers today feel we may be closing in on a far more sophisticated understanding of the large-scale structure of the universe—if only we can learn more about the nature and properties of dark matter. If you follow astronomy articles in the news (as we hope you will), you should be hearing more about dark matter in the years to come.

This page titled 16.5: Properties of Galaxies is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 26.4: Properties of Galaxies by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.6: Quasars

Learning Objectives

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

- Describe how quasars were discovered
- Explain how astronomers determined that quasars are at the distances implied by their redshifts
- Justify the statement that the enormous amount of energy produced by quasars is generated in a very small volume of space

The name “quasars” started out as short for “quasi-stellar radio sources” (here “quasi-stellar” means “sort of like stars”). The discovery of radio sources that appeared point-like, just like stars, came with the use of surplus World War II radar equipment in the 1950s. Although few astronomers would have predicted it, the sky turned out to be full of strong sources of radio waves. As they improved the images that their new radio telescopes could make, scientists discovered that some radio sources were in the same location as faint blue “stars.” No known type of star in our Galaxy emits such powerful radio radiation. What then were these “quasi-stellar radio sources”?

Redshifts: The Key to Quasars

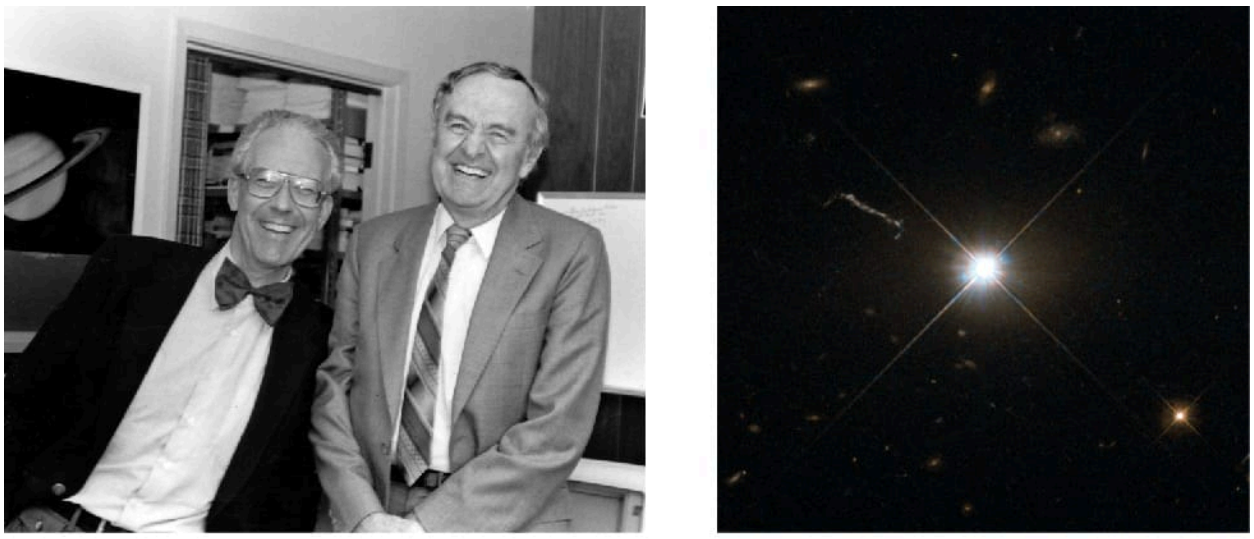
The answer came when astronomers obtained visible-light spectra of two of those faint “blue stars” that were strong sources of radio waves (Figure 27.2). Spectra of these radio “stars” only deepened the mystery: they had emission lines, but astronomers at first could not identify them with any known substance. By the 1960s, astronomers had a century of experience in identifying elements and compounds in the spectra of stars. Elaborate tables had been published showing the lines that each element would produce under a wide range of conditions. A “star” with unidentifiable lines in the ordinary visible light spectrum had to be something completely new.



Figure 27.2 : Typical Quasar. The arrow in this image marks the quasar known by its catalog number, PKS 1117-248. Note that nothing in this image distinguishes the quasar from an ordinary star. Its spectrum, however, shows that it is moving away from us at a speed of 36% the speed of light, or 67,000 miles per second. In contrast, the maximum speed observed for any star is only a few hundred miles per second. (credit: modification of work by WIYN Telescope, Kitt Peak National Observatory, NOAO)

In 1963 at Caltech’s Palomar Observatory, Maarten Schmidt (Figure 27.3) was puzzling over the spectrum of one of the radio stars, which was named 3C 273 because it was the 273rd entry in the third Cambridge catalog of radio sources (part (b) of Figure 27.3). There were strong emission lines in the spectrum, and Schmidt recognized that they had the same spacing between them as the

Balmer lines of hydrogen (see Radiation and Spectra). But the lines in 3C 273 were shifted far to the red of the wavelengths at which the Balmer lines are normally located. Indeed, these lines were at such long wavelengths that if the redshifts were attributed to the Doppler effect, 3C 273 was receding from us at a speed of 45,000 kilometers per second, or about 15% the speed of light! Since stars don't show Doppler shifts this large, no one had thought of considering high redshifts to be the cause of the strange spectra.



(a) (b)

Figure 27.3 : Quasar Pioneers and Quasar 3C 273. (a) Maarten Schmidt (left), who solved the puzzle of the quasar spectra in 1963, shares a joke in this 1987 photo with Allan Sandage, who took the first spectrum of a quasar. Sandage was also instrumental in measuring the value of Hubble's constant. (b) This is the first quasar for which a redshift was measured. The redshift showed that the light from it took about 2.5 billion years to reach us. Despite this great distance, it is still one of the quasars closest to the Milky Way Galaxy. Note also the faint streak going toward the upper left from the quasar. Some quasars, like 3C 273, eject super-fast jets of material. The jet from 3C 273 is about 200,000 light-years long. (credit a: modification of work by Andrew Fraknoi; credit b: modification of work by ESA/Hubble/NASA)

The puzzling emission lines in other star-like radio sources were then reexamined to see if they, too, might be well-known lines with large redshifts. This proved to be the case, but the other objects were found to be receding from us at even greater speeds. Their astounding speeds showed that the radio "stars" could not possibly be stars in our own Galaxy. Any true star moving at more than a few hundred kilometers per second would be able to overcome the gravitational pull of the Galaxy and completely escape from it. (As we shall see later in this chapter, astronomers eventually discovered that there was also more to these "stars" than just a point of light.)

It turns out that these high-velocity objects only look like stars because they are compact and very far away. Later, astronomers discovered objects with large redshifts that appear star-like but have no radio emission. Observations also showed that quasars were bright in the infrared and X-ray bands too, and not all these X-ray or infrared-bright quasars could be seen in either the radio or the visible-light bands of the spectrum. Today, all these objects are referred to as *quasi-stellar objects* (QSOs), or, as they are more popularly known, quasars. (The name was also soon appropriated by a manufacturer of home electronics.)

Link to Learning

Read an interview with Maarten Schmidt on the fiftieth anniversary of his insight about the spectrum of quasars and their redshifts.

Over a million quasars have now been discovered, and spectra are available for over a hundred thousand. All these spectra show redshifts, none show blueshifts, and their redshifts can be very large. Yet in a photo they look just like stars (Figure 27.4).

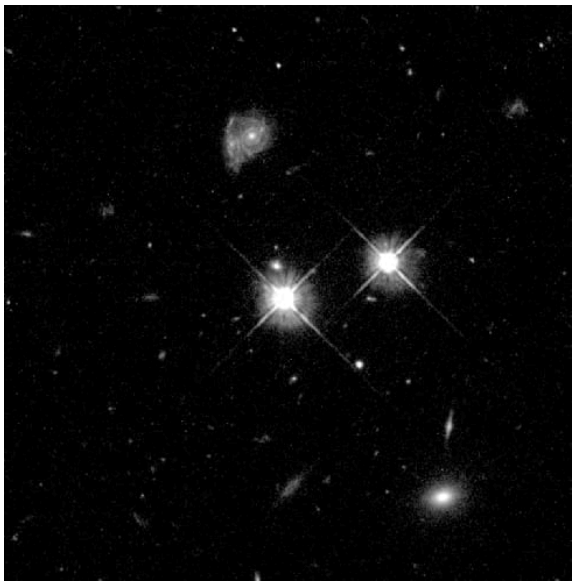


Figure 27.4 : Typical Quasar Imaged by the Hubble Space Telescope. One of these two bright "stars" in the middle is in our Galaxy, while the other is a quasar 9 billion light-years away. From this picture alone, there's no way to say which is which. (The quasar is the one in the center of the picture.) (credit: Charles Steidel (CIT)/NASA/ESA)

In the record-holding quasars, the first Lyman series line of hydrogen, with a laboratory wavelength of 121.5 nanometers in the ultraviolet portion of the spectrum, is shifted all the way through the visible region to the infrared. At such high redshifts, the simple formula for converting a Doppler shift to speed (Radiation and Spectra) must be modified to take into account the effects of the theory of relativity. If we apply the relativistic form of the Doppler shift formula, we find that these redshifts correspond to velocities of about 96% of the speed of light.

Example 27.1: Recession Speed of a Quasar

The formula for the Doppler shift, which astronomers denote by the letters *z*, is

$$z = \frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

where λ is the wavelength emitted by a source of radiation that is not moving, Δλ is the difference between that wavelength and the wavelength we measure, *v* is the speed with which the source moves away, and *c* (as usual) is the speed of light.

A line in the spectrum of a galaxy is at a wavelength of 393 nanometers (nm, or 10^{-9} m) when the source is at rest. Let's say the line is measured to be longer than this value (redshifted) by 7.86 nm. Then its redshift $z = \frac{7.86\,\text{nm}}{393\,\text{nm}} = 0.02$, so its speed away from us is 2% of the speed of light $\left(\frac{v}{c} = 0.02\right)$.

This formula is fine for galaxies that are relatively nearby and are moving away from us slowly in the expansion of the universe. But the quasars and distant galaxies we discuss in this chapter are moving away at speeds close to the speed of light. In that case, converting a Doppler shift (redshift) to a distance must include the effects of the special theory of relativity, which explains how measurements of space and time change when we see things moving at high speeds. The details of how this is done are way beyond the level of this text, but we can share with you the relativistic formula for the Doppler shift:

$$\frac{v}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1}$$

Let's do an example. Suppose a distant quasar has a redshift of 3. At what fraction of the speed of light is the quasar moving away?

Solution

We calculate the following:

$$\frac{v}{c} = \frac{(3+1)^2 - 1}{(3+1)^2 + 1} = \frac{16-1}{16+1} = \frac{15}{17} = 0.882$$

The quasar is thus receding from us at about 88% the speed of light.

Exercise 16.6.1

Several lines of hydrogen absorption in the visible spectrum have rest wavelengths of 410 nm, 434 nm, 486 nm, and 656 nm. In a spectrum of a distant galaxy, these same lines are observed to have wavelengths of 492 nm, 521 nm, 583 nm, and 787 nm respectively. What is the redshift of this galaxy? What is the recession speed of this galaxy?

Answer available at the end of this section.

Quasars Obey the Hubble Law

The first question astronomers asked was whether quasars obeyed the Hubble law and were really at the large distances implied by their redshifts. If they did not obey the rule that large redshift means large distance, then they could be much closer, and their luminosity could be a lot less. One straightforward way to show that quasars had to obey the Hubble law was to demonstrate that they were actually part of galaxies, and that their redshift was the same as the galaxy that hosted them. Since ordinary galaxies *do* obey the Hubble law, anything within them would be subject to the same rules.

Observations with the Hubble Space Telescope provided the strongest evidence showing that quasars are located at the centers of galaxies. Hints that this is true had been obtained with ground-based telescopes, but space observations were required to make a convincing case. The reason is that quasars can outshine their entire galaxies by factors of 10 to 100 or even more. When this light passes through Earth's atmosphere, it is blurred by turbulence and drowns out the faint light from the surrounding galaxy—much as the bright headlights from an oncoming car at night make it difficult to see anything close by.

The Hubble Space Telescope, however, is not affected by atmospheric turbulence and can detect the faint glow from some of the galaxies that host quasars (Figure 27.5). Quasars have been found in the cores of both spiral and elliptical galaxies, and each quasar has the same redshift as its host galaxy. A wide range of studies with the Hubble Space Telescope now clearly demonstrate that quasars are indeed far away. If so, they must be producing a truly impressive amount of energy to be detectable as points of light that are much brighter than their galaxy. Interestingly, many quasar host galaxies are found to be involved in a collision with a second galaxy, providing, as we shall see, an important clue to the source of their prodigious energy output.

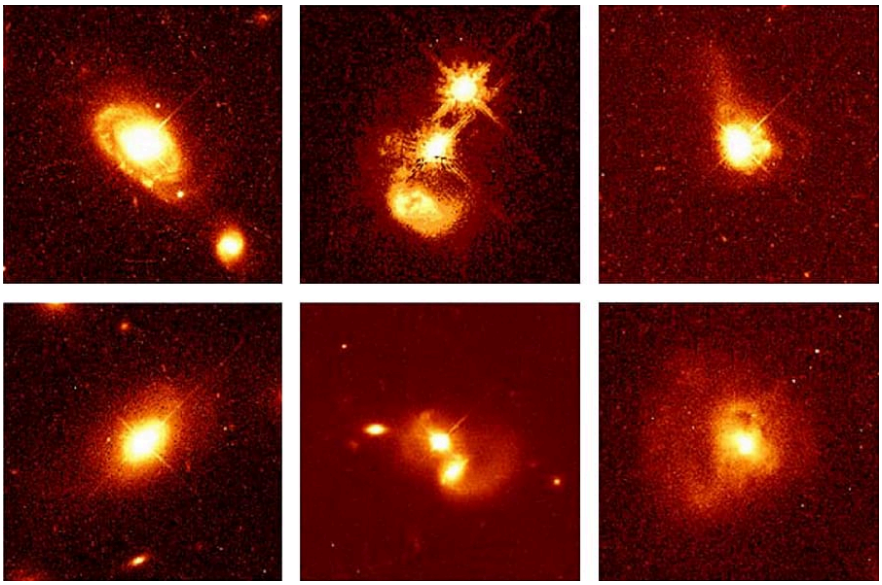


Figure 27.5 : Quasar Host Galaxies. The Hubble Space Telescope reveals the much fainter “host” galaxies around quasars. The top left image shows a quasar that lies at the heart of a spiral galaxy 1.4 billion light-years from Earth. The bottom left image shows a quasar that lies at the center of an elliptical galaxy some 1.5 billion light-years from us. The middle images show remote pairs of interacting galaxies, one of which harbors a quasar. Each of the right images shows long tails of gas and dust streaming away from a galaxy that contains a quasar. Such tails are produced when one galaxy collides with another. (credit: modification of work by John Bahcall, Mike Disney, NASA)

The Size of the Energy Source

Given their large distances, quasars have to be extremely luminous to be visible to us at all—far brighter than any normal galaxy. In visible light alone, most are far more energetic than the brightest elliptical galaxies. But, as we saw, quasars also emit energy at X-ray and ultraviolet wavelengths, and some are radio sources as well. When all their radiation is added together, some QSOs have total luminosities as large as a hundred trillion Suns (10^{14} L_{Sun}), which is 10 to 100 times the brightness of luminous elliptical galaxies.

Finding a mechanism to produce the large amount of energy emitted by a quasar would be difficult under any circumstances. But there is an additional problem. When astronomers began monitoring quasars carefully, they found that some vary in luminosity on time scales of months, weeks, or even, in some cases, days. This variation is irregular and can change the brightness of a quasar by a few tens of percent in both its visible light and radio output.

Think about what such a change in luminosity means. A quasar at its dimmest is still more brilliant than any normal galaxy. Now imagine that the brightness increases by 30% in a few weeks. Whatever mechanism is responsible must be able to release new energy at rates that stagger our imaginations. The most dramatic changes in quasar brightness are equivalent to the energy released by 100,000 billion Suns. To produce this much energy we would have to convert the total mass of about ten Earths into energy every minute.

Moreover, because the fluctuations occur in such short times, the part of a quasar that is varying must be smaller than the distance light travels in the time it takes the variation to occur—typically a few months. To see why this must be so, let’s consider a cluster of stars 10 light-years in diameter at a very large distance from Earth (see Figure 27.6, in which Earth is off to the right). Suppose every star in this cluster somehow brightens simultaneously and remains bright. When the light from this event arrives at Earth, we would first see the brighter light from stars on the near side; 5 years later we would see increased light from stars at the center. Ten years would pass before we detected more light from stars on the far side.

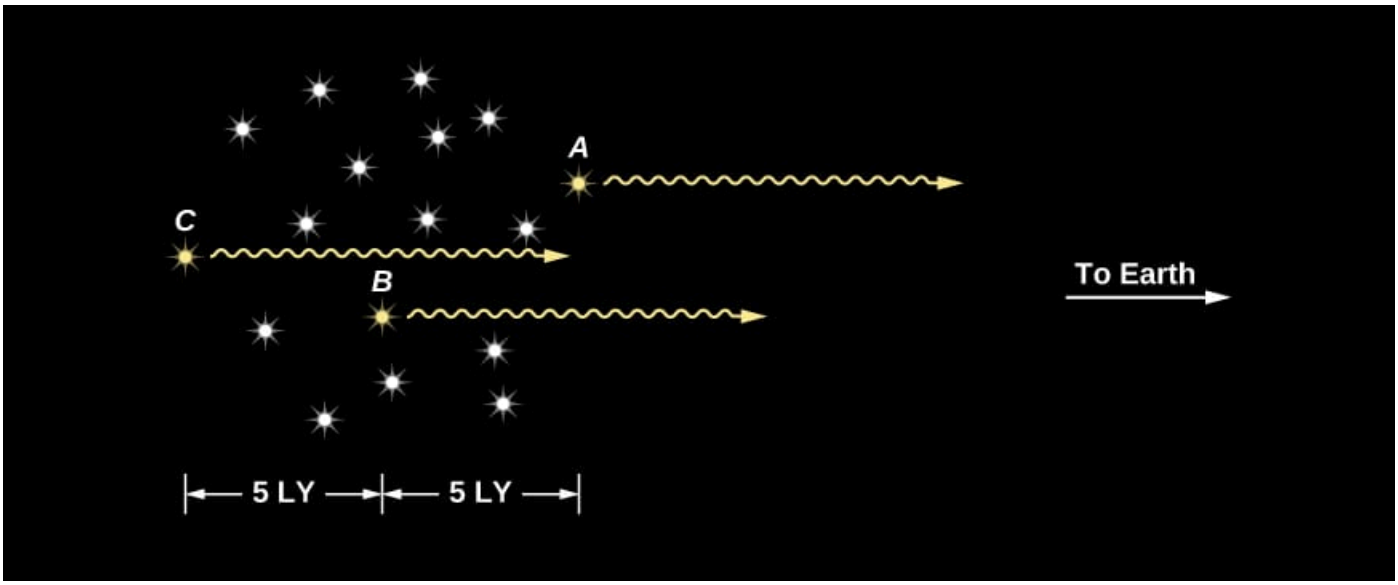


Figure 27.6 : How the Size of a Source Affects the Timescale of Its Variability. This diagram shows why light variations from a large region in space appear to last for an extended period of time as viewed from Earth. Suppose all the stars in this cluster, which is 10 light-years across, brighten simultaneously and instantaneously. From Earth, star *A* will appear to brighten 5 years before star *B*, which in turn will appear to brighten 5 years earlier than star *C*. It will take 10 years for an Earth observer to get the full effect of the brightening.

Even though all stars in the cluster brightened at the same time, the fact that the cluster is 10 light-years wide means that 10 years must elapse before the increased light from every part of the cluster reaches us. From Earth we would see the cluster get brighter and brighter, as light from more and more stars began to reach us. Not until 10 years after the brightening began would we see the cluster reach maximum brightness. In other words, if an extended object suddenly flares up, it will seem to brighten over a period of time equal to the time it takes light to travel across the object from its far side.

We can apply this idea to brightness changes in quasars to estimate their diameters. Because quasars typically vary (get brighter and dimmer) over periods of a few months, the region where the energy is generated can be no larger than a few light-months across. If it were larger, it would take longer than a few months for the light from the far side to reach us.

How large is a region of a few light-months? Pluto, usually the outermost (dwarf) planet in our solar system, is about 5.5 light-hours from us, while the nearest star is 4 light-years away. Clearly a region a few light months across is tiny relative to the size of the entire Galaxy. And some quasars vary even more rapidly, which means their energy is generated in an even smaller region. Whatever mechanism powers the quasars must be able to generate more energy than that produced by an entire galaxy in a volume of space that, in some cases, is not much larger than our solar system.

Earlier Evidence

Even before the discovery of quasars, there had been hints that something very strange was going on in the centers of at least some galaxies. Back in 1918, American astronomer Heber Curtis used the large Lick Observatory telescope to photograph the galaxy Messier 87 in the constellation Virgo. On that photograph, he saw what we now call a jet coming from the center, or nucleus, of the galaxy (Figure 27.7). This jet literally and figuratively pointed to some strange activity going on in that galaxy nucleus. But he had no idea what it was. No one else knew what to do with this space oddity either.

The random factoid that such a central jet existed lay around for a quarter century, until Carl Seyfert, a young astronomer at Mount Wilson Observatory, also in California, found half a dozen galaxies with extremely bright nuclei that were almost stellar, rather than fuzzy in appearance like most galaxy nuclei. Using spectroscopy, he found that these nuclei contain gas moving at up to two percent the speed of light. That may not sound like much, but it is 6 million miles per hour, and more than 10 times faster than the typical motions of stars in galaxies.

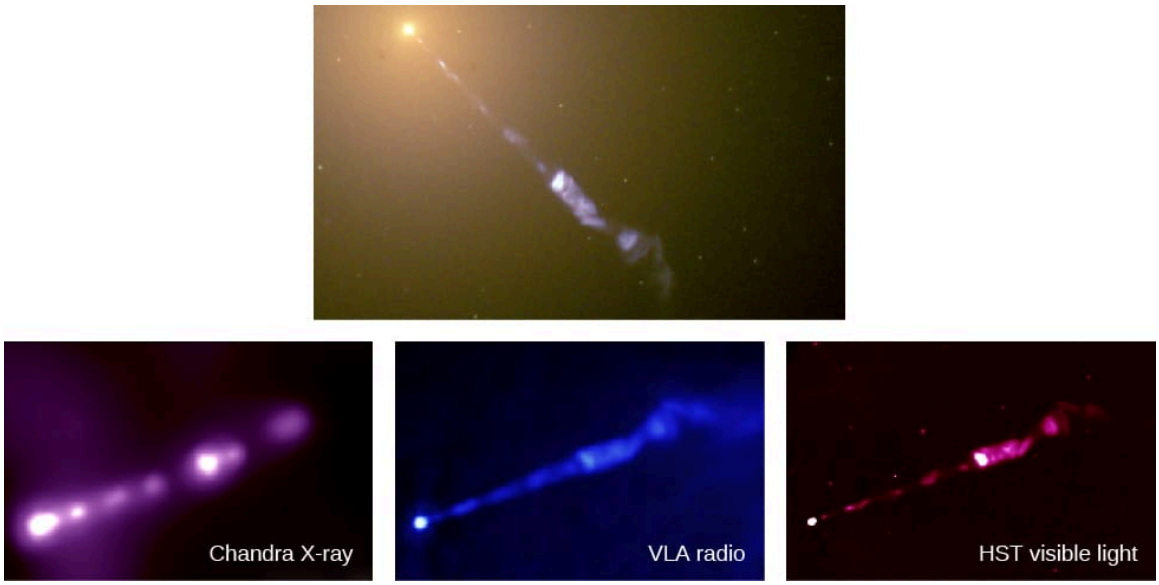


Figure 27.7 : M87 Jet. Streaming out like a cosmic searchlight from the center of the galaxy, M87 is one of nature's most amazing phenomena, a huge jet of electrons and other particles traveling at nearly the speed of light. In this Hubble Space Telescope image, the blue of the jet contrasts with the yellow glow from the combined light of billions of unseen stars and yellow, point-like globular clusters that make up the galaxy (at the upper left). As we shall see later in this chapter, the jet, which is several thousand light-years long, originates in a disk of superheated gas swirling around a giant black hole at the center of M87. The light that we see is produced by electrons twisting along magnetic field lines in the jet, a process known as synchrotron radiation, which gives the jet its bluish tint. The jet in M87 can be observed in X-ray, radio, and visible light, as shown in the bottom three images. At the extreme left of each bottom image, we see the bright galactic nucleus harboring a supermassive black hole. (credit top: modification of work by NASA, The Hubble Heritage Team(STScI/AURA); credit bottom: modification of work by X-ray: H. Marshall (MIT), et al., CXC, NASA; Radio: F. Zhou, F. Owen (NRAO), J. Biretta (STScI); Optical: E. Perlman (UMBC), et al.)

After decades of study, astronomers identified many other strange objects beyond our Milky Way Galaxy; they populate a whole "zoo" of what are now called active galaxies or active galactic nuclei (AGN). Astronomers first called them by many different names, depending on what sorts of observations discovered each category, but now we know that we are always looking at the same basic mechanism. What all these galaxies have in common is some activity in their nuclei that produces an enormous amount of energy in a very small volume of space. In the next section, we describe a model that explains all these galaxies with strong central activity—both the AGNs and the QSOs.

Link to Learning

To see a jet for yourself, check out a time-lapse video of the jet ejected from NGC 3862.

Answers

Exercise 16.6.1

Because this is the same galaxy, we could pick any one of the four wavelengths and calculate how much it has shifted. If we use a rest wavelength of 410 nm and compare it to the shifted wavelength of 492 nm, we see that

$$z = \frac{\Delta\lambda}{\lambda} = \frac{(492\,\text{nm} - 410\,\text{nm})}{410\,\text{nm}} = \frac{82\,\text{nm}}{410\,\text{nm}} = 0.20$$

In the classical view, this galaxy is receding at 20% of the speed of light; however, at 20% of the speed of light, relativistic effects are starting to become important. So, using the relativistic Doppler equation, we compute the true recession rate as

$$\frac{v}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1} = \frac{(0.2+1)^2 - 1}{(0.2+1)^2 + 1} = \frac{1.44 - 1}{1.44 + 1} = \frac{0.44}{2.44} = 0.18$$

Therefore, the actual recession speed is only 18% of the speed of light. While this may not initially seem like a big difference from the classical measurement, there is already an 11% deviation between the classical and the relativistic solutions; and at greater recession speeds, the divergence between the classical and relativistic speeds increases rapidly!

This page titled 16.6: Quasars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 27.2: Quasars by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.7: Supermassive Black Holes- What Quasars really are

Learning Objectives

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

- Describe the characteristics common to all quasars
- Justify the claim that supermassive black holes are the source of the energy emitted by quasars (and AGNs)
- Explain how a quasar's energy is produced

In order to find a common model for quasars (and their cousins, the AGNs), let's first list the common characteristics we have been describing—and add some new ones:

- Quasars are hugely powerful, emitting more power in radiated light than all the stars in our Galaxy combined.
- Quasars are tiny, about the size of our solar system (to astronomers, that is really small!).
- Some quasars are observed to be shooting out pairs of straight jets at close to the speed of light, in a tight beam, to distances far beyond the galaxies they live in. These jets are themselves powerful sources of radio and gamma-ray radiation.
- Because quasars put out so much power from such a small region, they can't be powered by nuclear fusion the way stars are; they must use some process that is far more efficient.
- As we shall see later in this chapter, quasars were much more common when the universe was young than they are today. That means they must have been able to form in the first billion years or so after the universe began to expand.

The readers of this text are in a much better position than the astronomers who discovered quasars in the 1960s to guess what powers the quasars. That's because the key idea in solving the puzzle came from observations of the black holes we discussed in Black Holes and Curved Spacetime. The discovery of the first stellar mass black hole in the binary system Cygnus X-1 was announced in 1971, several years after the discovery of quasars. Proof that there is a black hole at the center of our own Galaxy came even later. Back when astronomers first began trying to figure out what powered quasars, black holes were simply one of the more exotic predictions of the general theory of relativity that still waited to be connected to the real world.

It was only as proof of the existence of black holes accumulated over several decades that it became clearer that only supermassive black holes could account for all the observed properties of quasars and AGNs. As we saw in The Milky Way Galaxy, our own Galaxy has a black hole in its center, and the energy is emitted from a small central region. While our black hole doesn't have the mass or energy of the quasar black holes, the mechanism that powers them is similar. The evidence now shows that most—and probably all—elliptical galaxies and all spirals with nuclear bulges have black holes at their centers. The amount of energy emitted by material near the black hole depends on two things: the mass of the black hole and the amount of matter that is falling into it.

If a black hole with a billion Suns' worth of mass inside (10^9 M_{Sun}) accretes (gathers) even a relatively modest amount of additional material—say, about 10 M_{Sun} per year—then (as we shall see) it can, in the process, produce as much energy as a thousand normal galaxies. This is enough to account for the total energy of a quasar. If the mass of the black hole is smaller than a billion solar masses or the accretion rate is low, then the amount of energy emitted can be much smaller, as it is in the case of the Milky Way.

Link to Learning

Watch a video of an artist's impression of matter accreting around a supermassive black hole.

Observational Evidence for Black Holes

In order to prove that a black hole is present at the center of a galaxy, we must demonstrate that so much mass is crammed into so small a volume that no normal objects—massive stars or clusters of stars—could possibly account for it (just as we did for the black hole in the Milky Way). We already know from observations (discussed in Black Holes and Curved Spacetime) that an

accreting black hole is surrounded by a hot *accretion disk* with gas and dust that swirl around the black hole before it falls in.

If we assume that the energy emitted by quasars is also produced by a hot accretion disk, then, as we saw in the previous section, the size of the disk must be given by the time the quasar energy takes to vary. For quasars, the emission in visible light varies on typical time scales of 5 to 2000 days, limiting the size of the disk to that many light-days.

In the X-ray band, quasars vary even more rapidly, so the light travel time argument tells us that this more energetic radiation is generated in an even smaller region. Therefore, the mass around which the accretion disk is swirling must be confined to a space that is even smaller. If the quasar mechanism involves a great deal of mass, then the only astronomical object that can confine a lot of mass into a very small space is a black hole. In a few cases, it turns out that the X-rays are emitted from a region just a few times the size of the black hole event horizon.

The next challenge, then, is to "weigh" this central mass in a quasar. In the case of our own Galaxy, we used observations of the orbits of stars very close to the galactic center, along with Kepler's third law, to estimate the mass of the central black hole (The Milky Way Galaxy). In the case of distant galaxies, we cannot measure the orbits of individual stars, but we can measure the orbital speed of the gas in the rotating accretion disk. The Hubble Space Telescope is especially well suited to this task because it is above the blurring of Earth's atmosphere and can obtain spectra very close to the bright central regions of active galaxies. The Doppler effect is then used to measure radial velocities of the orbiting material and so derive the speed with which it moves around.

One of the first galaxies to be studied with the Hubble Space Telescope is our old favorite, the giant elliptical M87. Hubble Space Telescope images showed that there is a disk of hot (10,000 K) gas swirling around the center of M87 (Figure 27.8). It was surprising to find hot gas in an elliptical galaxy because this type of galaxy is usually devoid of gas and dust. But the discovery was extremely useful for pinning down the existence of the black hole. Astronomers measured the Doppler shift of spectral lines emitted by this gas, found its speed of rotation, and then used the speed to derive the amount of mass inside the disk—applying Kepler's third law.

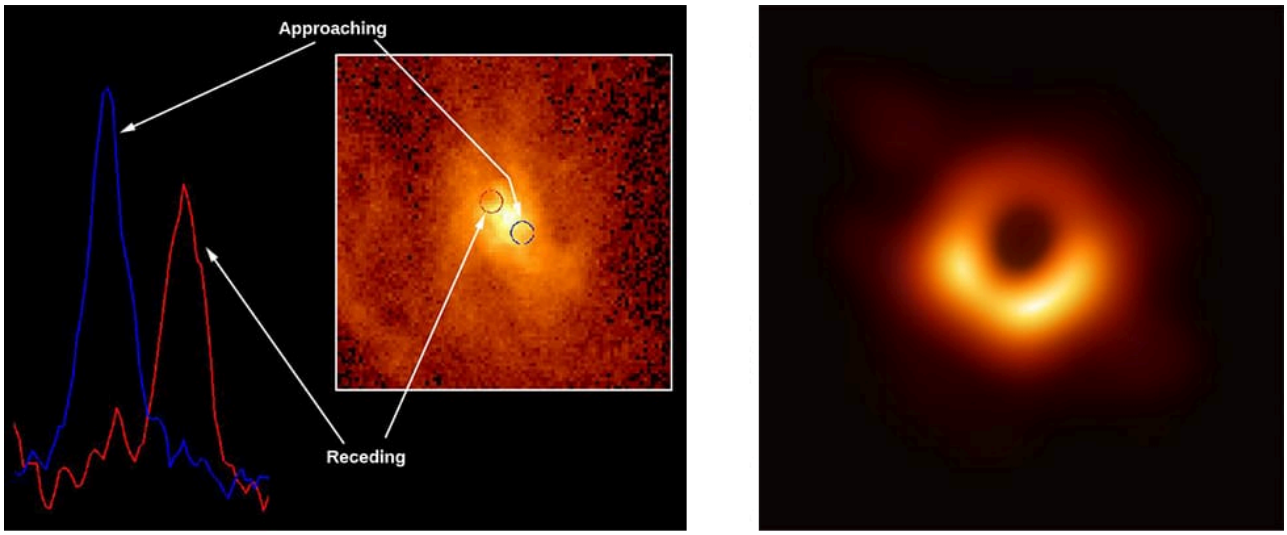


Figure 27.8 : Evidence for a Black Hole at the Center of M87. (a) The disk of whirling gas at right was discovered at the center of the giant elliptical galaxy M87 with the Hubble Space Telescope. Observations made on opposite sides of the disk show that one side is approaching us (the spectral lines are blueshifted by the Doppler effect) while the other is receding (lines redshifted), a clear indication that the disk is rotating. The rotation speed is about 550 kilometers per second or 1.2 million miles per hour. Such a high rotation speed is evidence that there is a very massive black hole at the center of M87. (b) By 2019, the Event Horizon Telescope, a collaborative venture among radio telescopes spanning our planet, was able to produce the first radio image of the shadow of the M87 black hole on the accretion disk surrounding it. This was taken with radio waves, so color is arbitrary. The shadow is about 2 ½ times as big as the event horizon of the supermassive black hole. (credit (a): modification of work by Holland Ford, STScI/JHU; Richard Harms, Linda Dressel, Ajay K. Kochhar, Applied Research Corp.; Zlatan Tsvetanov, Arthur Davidsen, Gerard Kriss, Johns Hopkins; Ralph Bohlin, George Hartig, STScI; Bruce Margon, University of Washington in Seattle; NASA; credit (b): modification of work by European Southern Observatory (ESO), Event Horizon Telescope Collaboration)

Modern estimates show that there is a mass of at least 3.5 billion M_{Sun} concentrated in a tiny region at the very center of M87. So much mass in such a small volume of space must be a black hole. Let's stop for a moment and take in this figure: a single

black hole that has swallowed enough material to make 3.5 billion stars like the Sun. Few astronomical measurements have ever led to so mind-boggling a result. What a strange environment the neighborhood of such a supermassive black hole must be.

Another example is shown in Figure 27.9. Here, we see a disk of dust and gas that surrounds a 300-million-M_{Sun} black hole in the center of an elliptical galaxy. (The bright spot in the center is produced by the combined light of stars that have been pulled close together by the gravitational force of the black hole.) The mass of the black hole was again derived from measurements of the rotational speed of the disk. The gas in the disk is moving around at 155 kilometers per second at a distance of only 186 light-years from its center. Given the pull of the mass at the center, we expect that the whole dust disk should be swallowed by the black hole in several billion years.

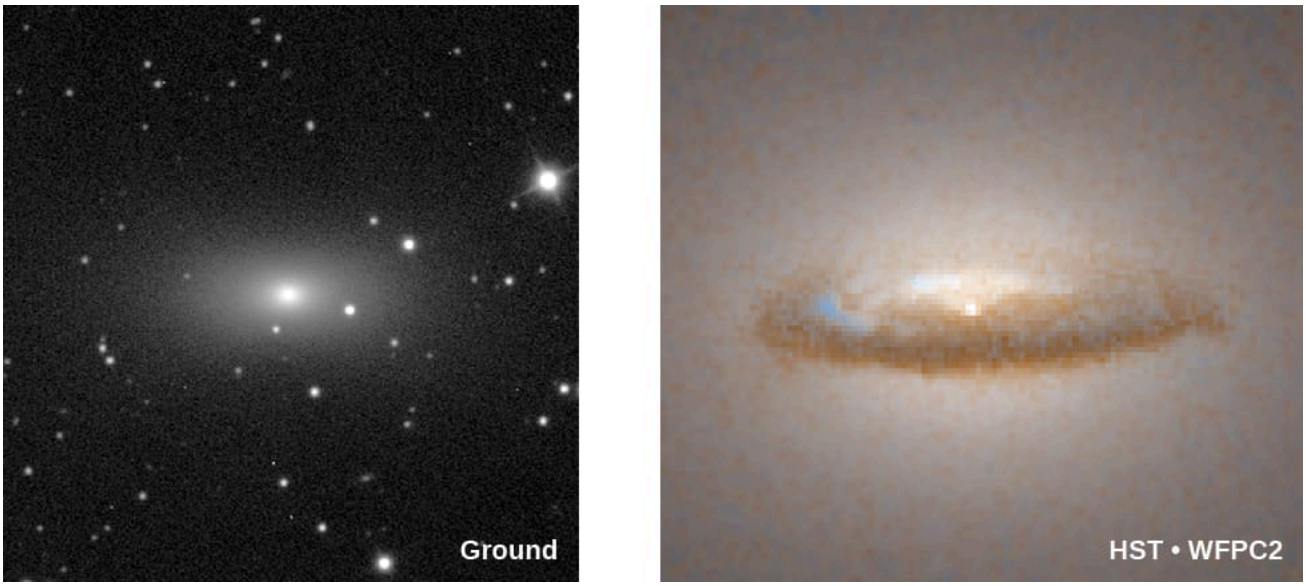


Figure 27.9 : Another Galaxy with a Black-Hole Disk. The ground-based image shows an elliptical galaxy called NGC 7052 located in the constellation of Vulpecula, almost 200 million light-years from Earth. At the galaxy's center (right) is a dust disk roughly 3700 light-years in diameter. The disk rotates like a giant merry-go-round: gas in the inner part (186 light-years from the center) whirls around at a speed of 155 kilometers per second (341,000 miles per hour). From these measurements and Kepler's third law, it is possible to estimate that the disk is orbiting around a central black hole with a mass of 300 million Suns. (credit: modification of work by Roeland P. van der Marel (STScI), Frank C. van den Bosch (University of Washington), NASA)

But do we *have* to accept black holes as the only explanation of what lies at the center of these galaxies? What else could we put in such a small space other than a giant black hole? The alternative is stars. But to explain the masses in the centers of galaxies without a black hole we need to put at least a million stars in a region the size of the solar system. To fit, they would have be only 2 star diameters apart. Collisions between stars would happen all the time. And these collisions would lead to mergers of stars, and very soon the one giant star that they form would collapse into a black hole. So there is really no escape: only a black hole can fit so much mass into so small a space.

As we saw earlier, observations now show that all the galaxies with a spherical concentration of stars—either elliptical galaxies or spiral galaxies with nuclear bulges (see the chapter on Galaxies)—harbor one of these giant black holes at their centers. Among them is our neighbor spiral galaxy, the Andromeda galaxy, M31. The masses of these central black holes range from a just under a million up to at least 30 billion times the mass of the Sun. Several black holes may be even more massive, but the mass estimates have large uncertainties and need verification. We call these black holes "supermassive" to distinguish them from the much smaller black holes that form when some stars die (see The Death of Stars). So far, the most massive black holes from stars—those detected through gravitational waves detected by LIGO—have masses of about 40 solar masses.

Energy Production around a Black Hole

By now, you may be willing to entertain the idea that huge black holes lurk at the centers of active galaxies. But we still need to answer the question of how such a black hole can account for one of the most powerful sources of energy in the universe. As we saw in Black Holes and Curved Spacetime, a black hole itself can radiate no energy. Any energy we detect from it must come from material very close to the black hole, but not inside its event horizon.

In a galaxy, a central black hole (with its strong gravity) attracts matter—stars, dust, and gas—orbiting in the dense nuclear regions. This matter spirals in toward the spinning black hole and forms an accretion disk of material around it. As the material spirals ever closer to the black hole, it accelerates and becomes compressed, heating up to temperatures of millions of degrees. Such hot matter can radiate prodigious amounts of energy as it falls in toward the black hole.

To convince yourself that falling into a region with strong gravity can release a great deal of energy, imagine dropping a printed version of your astronomy textbook out the window of the ground floor of the library. It will land with a thud, and maybe give a surprised pigeon a nasty bump, but the energy released by its fall will not be very great. Now take the same book up to the fifteenth floor of a tall building and drop it from there. For anyone below, astronomy could suddenly become a deadly subject; when the book hits, it does so with a great deal of energy.

Dropping things from far away into the much stronger gravity of a black hole is much more effective in turning the energy released by infall into other forms of energy. Just as the falling book can heat up the air, shake the ground, or produce sound energy that can be heard some distance away, so the energy of material falling toward a black hole can be converted to significant amounts of electromagnetic radiation.

What a black hole has to work with is not textbooks but streams of infalling gas. If a dense blob of gas moves through a thin gas at high speed, it heats up as it slows by friction. As it slows down, kinetic (motion) energy is turned into heat energy. Just like a spaceship reentering the atmosphere (Figure 27.10), gas approaching a black hole heats up and glows where it meets other gas. But this gas, as it approaches the event horizon, reaches speeds of 10% the speed of light and more. It therefore gets far, far hotter than a spaceship, which reaches no more than about 1500 K. Indeed, gas near a supermassive black hole reaches a temperature of about 150,000 K, about 100 times hotter than a spaceship returning to Earth. It can even get so hot—millions of degrees—that it radiates X-rays.

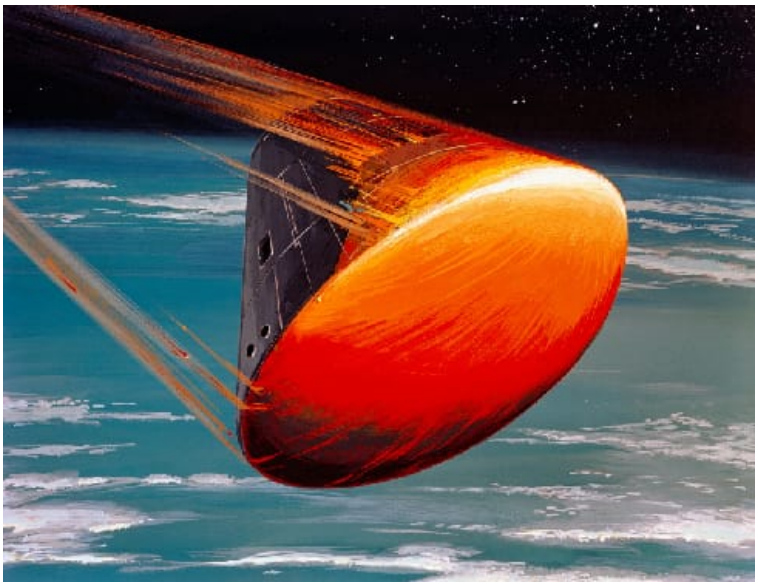


Figure 27.10 : Friction in Earth's Atmosphere. In this artist's impression, the rapid motion of a spacecraft (the Apollo mission reentry capsule) through the atmosphere compresses and heats the air ahead of it, which heats the spacecraft in turn until it glows red hot. Pushing on the air slows down the spacecraft, turning the kinetic energy of the spacecraft into heat. Fast-moving gas falling into a quasar heats up in a similar way. (credit: modification of work by NASA)

The amount of energy that can be liberated this way is enormous. Einstein showed that mass and energy are interchangeable with his famous formula $E = mc^2$ (see The Sun: A Nuclear Powerhouse). A hydrogen bomb releases just 1% of that energy, as does a star. Quasars are much more efficient than that. The energy released falling to the event horizon of a black hole can easily reach 10% or, in the extreme theoretical limit, 32%, of that energy. (Unlike the hydrogen atoms in a bomb or a star, the gas falling into the black hole is not actually losing mass from its atoms to free up the energy; the energy is produced just because the gas is falling closer and closer to the black hole.) This huge energy release explains how a tiny volume like the region around a black hole can release as much power as a whole galaxy. But to radiate all that energy, instead of just falling inside the event horizon with barely a peep, the hot gas must take the time to swirl around the star in the accretion disk and emit some of its energy.

Most black holes don't show any signs of quasar emission. We call them "quiescent." But, like sleeping dragons, they can be woken up by being roused with a fresh supply of gas. Our own Milky Way black hole is currently quiescent, but it may have been a

quasar just a few million years ago (Figure 27.11). Two giant bubbles that extend 25,000 light-years above and below the galactic center are emitting gamma rays. Were these produced a few million years ago when a significant amount of matter fell into the black hole at the center of the galaxy? Astronomers are still working to understand what remarkable event might have formed these enormous bubbles.

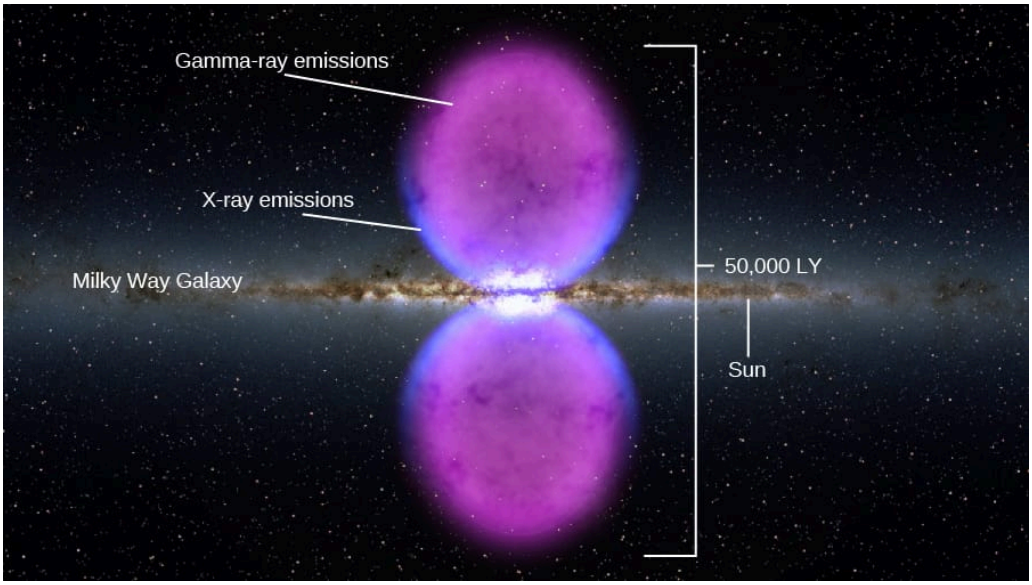


Figure 27.11 : Fermi Bubbles in the Galaxy. Giant bubbles shining in gamma-ray light lie above and below the center of the Milky Way Galaxy, as seen by the Fermi satellite. (The gamma-ray and X-ray image is superimposed on a visible-light image of the inner parts of our Galaxy.) The bubbles may be evidence that the supermassive black hole at the center of our Galaxy was a quasar a few million years ago. (credit: modification of work by NASA's Goddard Space Flight Center)

The physics required to account for the exact way in which the energy of infalling material is converted to radiation near a black hole is far more complicated than our simple discussion suggests. To understand what happens in the "rough and tumble" region around a massive black hole, astronomers and physicists must resort to computer simulations (and they require supercomputers, fast machines capable of awesome numbers of calculations per second). The details of these models are beyond the scope of our book, but they support the basic description presented here.

Radio Jets

So far, our model seems to explain the central energy source in quasars and active galaxies. But, as we have seen, there is more to quasars and other active galaxies than the point-like energy source. They can also have long jets that glow with radio waves, light, and sometimes even X-rays, and that extend far beyond the limits of the parent galaxy. Can we find a way for our black hole and its accretion disk to produce these jets of energetic particles as well?

Many different observations have now traced these jets to within 3 to 30 light-years of the parent quasar or galactic nucleus. While the black hole and accretion disk are typically smaller than 1 light-year, we nevertheless presume that if the jets come this close, they probably originate in the vicinity of the black hole. Another characteristic of the jets we need to explain is that they contain matter moving close to the speed of light.

Why are energetic electrons and other particles near a supermassive black hole ejected into jets, and often into two oppositely directed jets, rather than in all directions? Again, we must use theoretical models and supercomputer simulations of what happens when a lot of material whirls inward in a crowded black hole accretion disk. Although there is no agreement on exactly how jets form, it has become clear that any material escaping from the neighborhood of the black hole has an easier time doing so *perpendicular to* the disk.

In some ways, the inner regions of black hole accretion disks resemble a baby that is just learning to eat by herself. As much food as goes into the baby's mouth can sometimes wind up being spit out in various directions. In the same way, some of the material whirling inward toward a black hole finds itself under tremendous pressure and orbiting with tremendous speed. Under such conditions, simulations show that a significant amount of material can be flung outward—not back along the disk, where more material is crowding in, but above and below the disk. If the disk is thick (as it tends to be when a lot of material falls in

quickly), it can channel the outrushing material into narrow beams perpendicular to the disk (Figure 27.12).

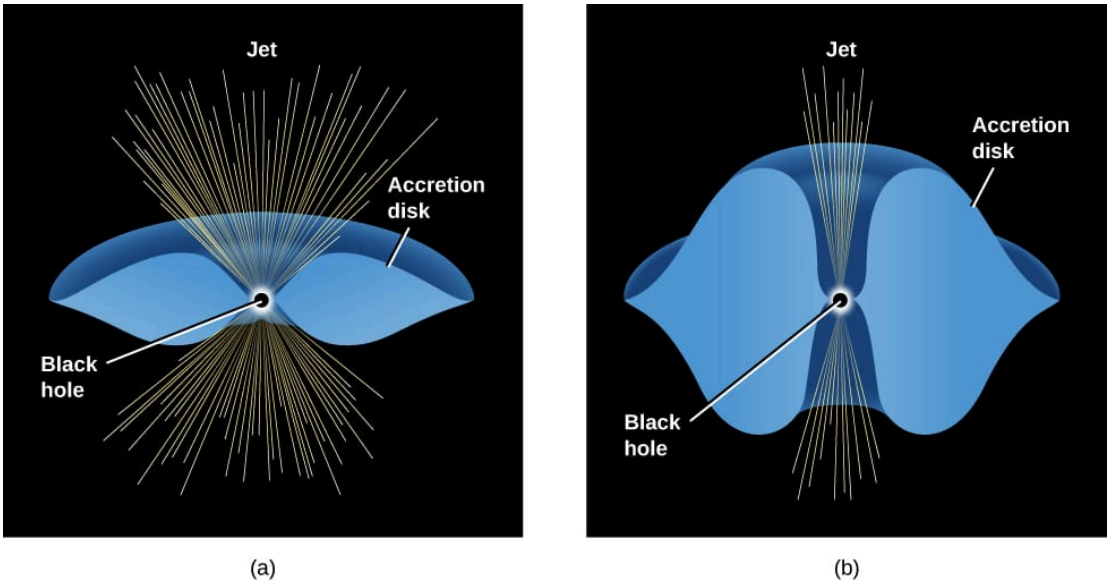


Figure 27.12 : Models of Accretion Disks. These schematic drawings show what accretion disks might look like around large black holes for (a) a thin accretion disk and (b) a “fat” disk—the type needed to account for channeling the outflow of hot material into narrow jets oriented perpendicular to the disk.

Figure 27.13 shows observations of an elliptical galaxy that behaves in exactly this way. At the center of this active galaxy, there is a ring of dust and gas about 400 light-years in diameter, surrounding a 1.2-billion-M_{Sun} black hole. Radio observations show that two jets emerge in a direction perpendicular to the ring, just as the model predicts.

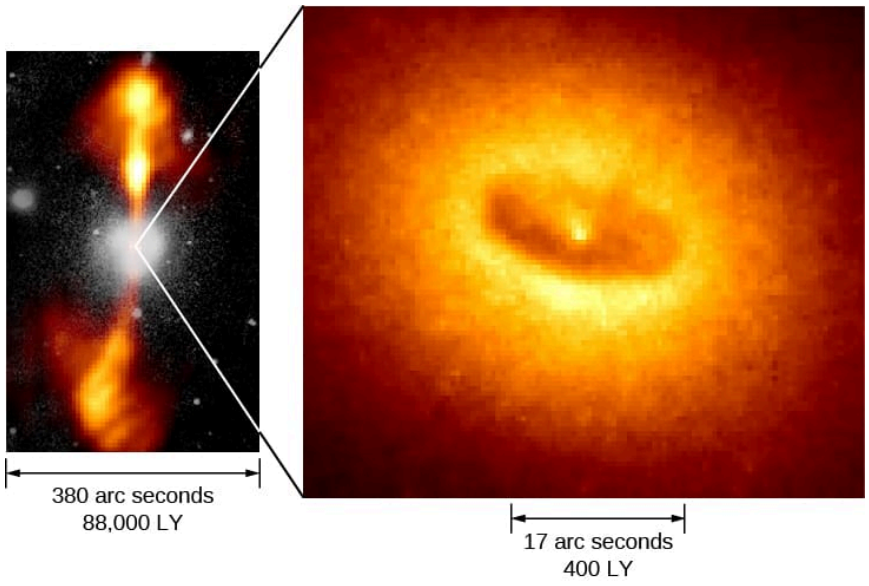


Figure 27.13 : Jets and Disk in an Active Galaxy. The picture on the left shows the active elliptical galaxy NGC 4261, which is located in the Virgo Cluster at a distance of about 100 million light-years. The galaxy itself—the white circular region in the center—is shown the way it looks in visible light, while the jets are seen at radio wavelengths. A Hubble Space Telescope image of the central portion of the galaxy is shown on the right. It contains a ring of dust and gas about 800 light-years in diameter, surrounding a supermassive black hole. Note that the jets emerge from the galaxy in a direction perpendicular to the plane of the ring. (credit: modification of work by ESA/HST)

Making Connections: Quasars and the Attitudes of Astronomers

The discovery of quasars in the early 1960s was the first in a series of surprises astronomers had in store. Within another decade they would find neutron stars (in the form of pulsars), the first hints of black holes (in binary X-ray sources), and even the radio echo of the Big Bang itself. Many more new discoveries lay ahead.

As Maarten Schmidt reminisced in 1988, "This had, I believe, a profound impact on the conduct of those practicing astronomy. Before the 1960s, there was much authoritarianism in the field. New ideas expressed at meetings would be instantly judged by senior astronomers and rejected if too far out." We saw a good example of this in the trouble Chandrasekhar had in finding acceptance for his ideas about the death of stars with cores greater than 1.4 M_{Sun} (see the feature box on Subrahmanyan Chandrasekhar).

"The discoveries of the 1960s," Schmidt continued, "were an embarrassment, in the sense that they were totally unexpected and could not be evaluated immediately. In reaction to these developments, an attitude has evolved where even outlandish ideas in astronomy are taken seriously. Given our lack of solid knowledge in extragalactic astronomy, this is probably to be preferred over authoritarianism."[1]

That is not to say that astronomers (being human) don't continue to have prejudices and preferences. For example, a small group of astronomers who thought that the redshifts of quasars were not connected with their distances (which was definitely a minority opinion) often felt excluded from meetings or from access to telescopes in the 1960s and 1970s. It's not so clear that they actually *were ex*cluded, as much as that they felt the very difficult pressure of knowing that most of their colleagues strongly disagreed with them. As it turned out, the evidence—which must ultimately decide all scientific questions—was not on their side either.

But today, as better instruments bring solutions to some problems and starkly illuminate our ignorance about others, the entire field of astronomy seems more open to discussing unusual ideas. Of course, before any hypotheses become accepted, they must be tested—again and again—against the evidence that nature itself reveals. Still, the many strange proposals published about what dark matter might be (see The Evolution and Distribution of Galaxies) attest to the new openness that Schmidt described.

With this black hole model, we have come a long way toward understanding the quasars and active galaxies that seemed very mysterious only a few decades ago. As often happens in astronomy, a combination of better instruments (making better observations) and improved theoretical models enabled us to make significant progress on a puzzling aspect of the cosmos.

Footnotes

- 1M. Schmidt, "The Discovery of Quasars," in *Modern Cosmology in Retrospect*, ed. B. Bertotti et al. (Cambridge University Press, 1990).

This page titled 16.7: Supermassive Black Holes- What Quasars really are is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 27.3: Supermassive Black Holes- What Quasars really are by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

16.8: Quasars as Probes of Evolution in the Universe

Learning Objectives

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

- Trace the rise and fall of quasars over cosmic time
- Describe some of the ways in which galaxies and black holes influence each other's growth
- Describe some ways the first black holes may have formed
- Explain why some black holes are not producing quasar emission but rather are quiescent

The quasars' brilliance and large distance make them ideal probes of the far reaches of the universe and its remote past. Recall that when first introducing quasars, we mentioned that they generally tend to be far away. When we see extremely distant objects, we are seeing them as they were long ago. Radiation from a quasar 8 billion light-years away is telling us what that quasar and its environment were like 8 billion years ago, much closer to the time that the galaxy that surrounds it first formed. Astronomers have now detected light emitted from quasars that were already formed only a few hundred million years after the universe began its expansion 13.8 billion years ago. Thus, they give us a remarkable opportunity to learn about the time when large structures were first assembling in the cosmos.

The Evolution of Quasars

Quasars provide compelling evidence that we live in an evolving universe—one that changes with time. They tell us that astronomers living billions of years ago would have seen a universe that is very different from the universe of today. Counts of the number of quasars at different redshifts (and thus at different times in the evolution of the universe) show us how dramatic these changes are (Figure 27.14). We now know that the number of quasars was greatest at the time when the universe was only 20% of its present age.

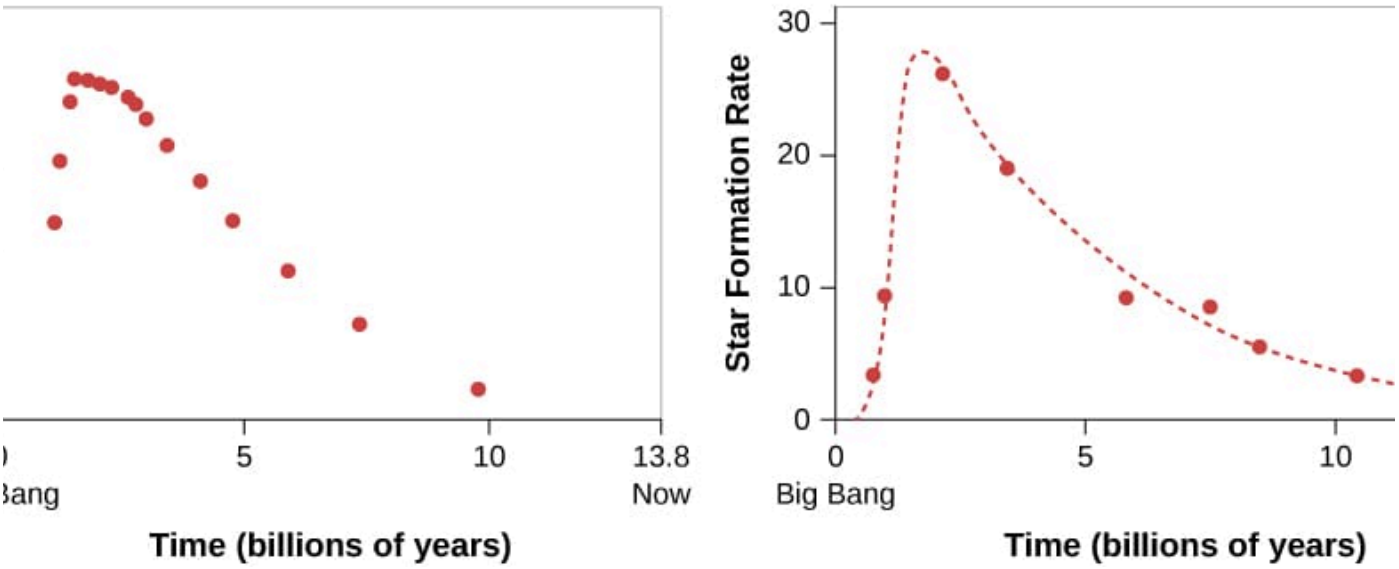


Figure 27.14 : Relative Number of Quasars and Rate at Which Stars Formed as a Function of the Age of the Universe. An age of 0 on the plots corresponds to the beginning of the universe; an age of 13.8 corresponds to the present time. Both the number of quasars and the rate of star formation were at a peak when the universe was about 20% as old as it is now.

As you can see, the drop-off in the numbers of quasars as time gets nearer to the present day is quite abrupt. Observations also show that the emission from the accretion disks around the most massive black holes peaks early and then fades. The most powerful quasars are seen only at early times. In order to explain this result, we make use of our model of the energy source of the quasars—namely that quasars are black holes with enough fuel to make a brilliant accretion disk right around them.

The fact that there were more quasars long ago (far away) than there are today (nearby) could be explained if there was more material available to be accreted by black holes early in the history of the universe. You might say that the quasars were more active when their black holes had fuel for their "energy-producing engines." If that fuel was mostly consumed in the first few billion years after the universe began its expansion, then later in its life, a "hungry" black hole would have very little left with which to light up the galaxy's central regions.

In other words, if matter in the accretion disk is continually being depleted by falling into the black hole or being blown out from the galaxy in the form of jets, then a quasar can continue to radiate only as long as new gas is available to replenish the accretion disk.

In fact, there *was* more gas around to be accreted early in the history of the universe. Back then, most gas had not yet collapsed to form stars, so there was more fuel available for both the feeding of black holes and the forming of new stars. Much of that fuel was subsequently consumed in the formation of stars during the first few billion years after the universe began its expansion. Later in its life, a galaxy would have little left to feed a hungry black hole or to form more new stars. As we see from Figure 27.14, both star formation and black hole growth peaked together when the universe was about 2 billion years old. Ever since, both have been in sharp decline. We are late to the party of the galaxies and have missed some of the early excitement.

Observations of nearer galaxies (seen later in time) indicate that there is another source of fuel for the central black holes—the collision of galaxies. If two galaxies of similar mass collide and merge, or if a smaller galaxy is pulled into a larger one, then gas and dust from one may come close enough to the black hole in the other to be devoured by it and so provide the necessary fuel. Astronomers have found that collisions were also much more common early in the history of the universe than they are today. There were more small galaxies in those early times because over time, as we shall see (in The Evolution and Distribution of Galaxies), small galaxies tend to combine into larger ones. Again, this means that we would expect to see more quasars long ago (far away) than we do today (nearby)—as we in fact do.

Codependence of Black Holes and Galaxies

Once black hole masses began to be measured reliably in the late 1990s, they posed an enigma. It looked as though the mass of the central black hole depended on the mass of the galaxy. The black holes in galaxies always seem to be just 1/200 the mass of the galaxy they live in. This result is shown schematically in Figure 27.15, and some of the observations are plotted in Figure 27.16.

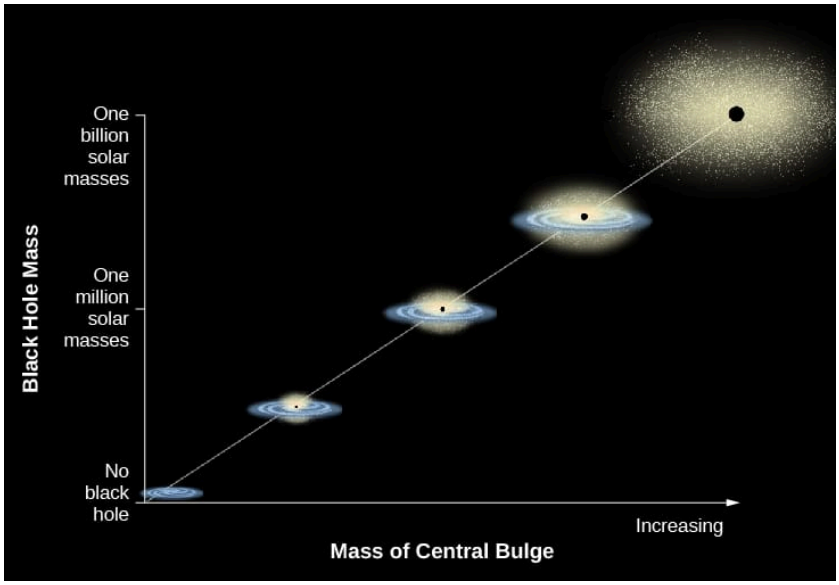


Figure 27.15 : Relationship between Black Hole Mass and the Mass of the Host Galaxy. Observations show that there is a close correlation between the mass of the black hole at the center of a galaxy and the mass of the spherical distribution of stars that surrounds the black hole. That spherical distribution may be in the form of either an elliptical galaxy or the central bulge of a spiral galaxy. (credit: modification of work by K. Cordes, S. Brown (STScI))

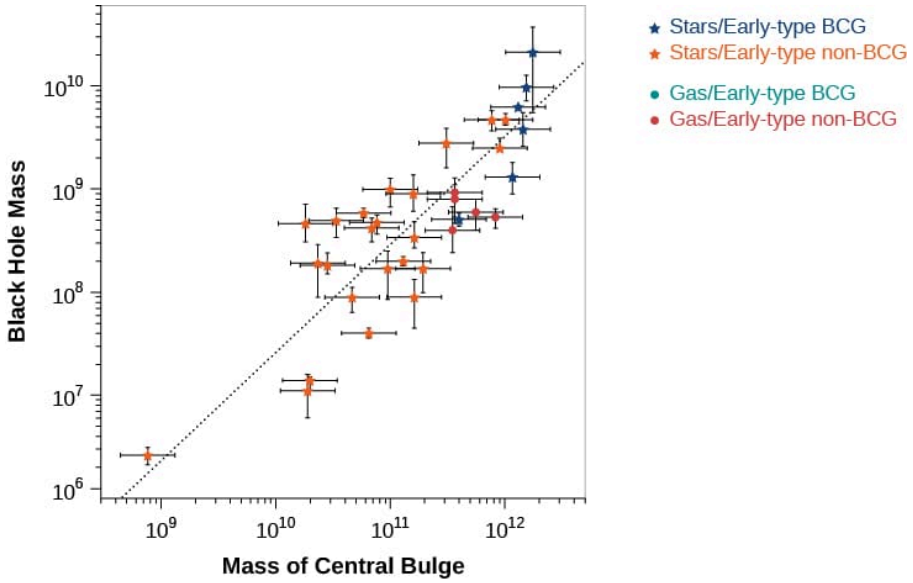


Figure 27.16 : Correlation between the Mass of the Central Black Hole and the Mass Contained within the Bulge of Stars Surrounding the Black Hole, Using Data from Real Galaxies. The black hole always turns out to be about 1/200 the mass of the stars surrounding it. The horizontal and vertical bars surrounding each point show the uncertainty of the measurement. (credit: modification of work by Nicholas J. McConnell, Chung-Pei Ma, "Revisiting the Scaling Relations of Black Hole Masses and Host Galaxy Properties," *The Astrophysical Journal*, 764:184 (14 pp.), February 20, 2013.)

Somehow black hole mass and the mass of the surrounding bulge of stars are connected. But why does this correlation exist? Unfortunately, astronomers do not yet know the answer to this question. We do know, however, that the black hole can influence the rate of star formation in the galaxy, and that the properties of the surrounding galaxy can influence how fast the black hole grows. Let's see how these processes work.

How a Galaxy Can Influence a Black Hole in Its Center

Let's look first at how the surrounding galaxy might influence the growth and size of the black hole. Without large quantities of fresh "food," the surroundings of black holes glow only weakly as bits of local material spiral inward toward the black hole. So somehow large amounts of gas have to find their way to the black hole from the galaxy in order to feed the quasar and make it grow and give off the energy to be noticed. Where does this "food" for the black hole come from originally and how might it be replenished? The jury is still out, but the options are pretty clear.

One obvious source of fuel for the black hole is matter from the host galaxy itself. Galaxies start out with large amounts of interstellar gas and dust, and at least some of this interstellar matter is gradually converted into stars as the galaxy evolves. On the other hand, as stars go through their lives and die, they lose mass all the time into the space between them, thereby returning some of the gas and dust to the interstellar medium. We expect to find more gas and dust in the central regions early in a galaxy's life than later on, when much of it has been converted into stars. Any of the interstellar matter that ventures too close to the black hole may be accreted by it. This means that we would expect that the number and luminosity of quasars powered in this way would decline with time. And as we have seen, that is just what we find.

Today both *elliptical galaxies* and the *nuclear bulges of spiral galaxies* have very little raw material left to serve as a source of fuel for the black hole. And most of the giant black holes in nearby galaxies, including the one in our own Milky Way, are now dark and relatively quiet—mere shadows of their former selves. So that fits with our observations.

We should note that even if you have a quiescent supermassive black hole, a star in the area could occasionally get close to it. Then the powerful tidal forces of the black hole can pull the whole star apart into a stream of gas. This stream quickly forms an accretion disk that gives off energy in the normal way and makes the black hole region into a temporary quasar. However, the material will fall into the black hole after only a few weeks or months. The black hole then goes back into its lurking, quiescent state, until another victim wanders by.

This sort of "cannibal" event happens only once every 100,000 years or so in a typical galaxy. But we can monitor millions of galaxies in the sky, so a few of these "tidal disruption events" are found each year (Figure 27.17). However, these individual events,

dramatic as they are, are too rare to account for the huge masses of the central black holes.

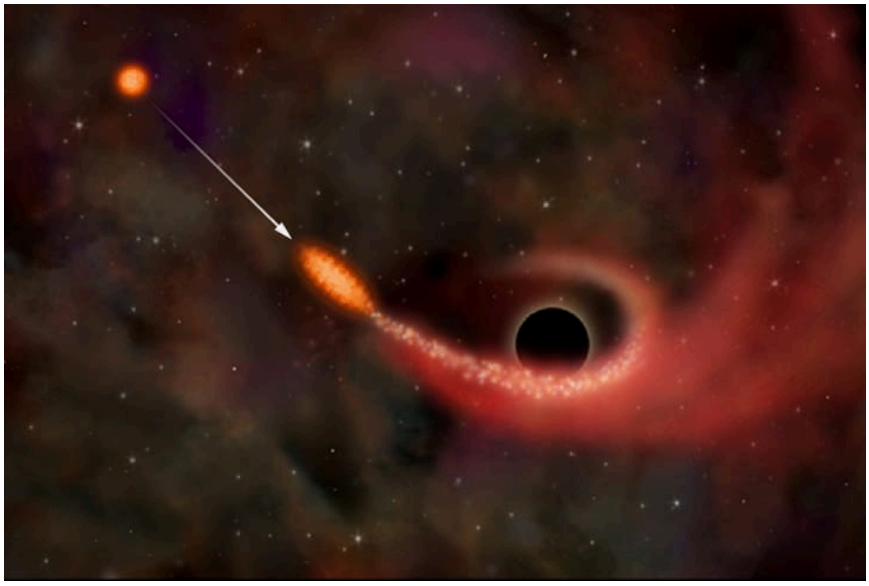


Figure 27.17 : A Black Hole Snacks on a Star. This artist's impression shows three stages of a star (red) swinging too close to a giant black hole (black circle). The star starts off (top left) in its normal spherical shape, then begins to be pulled into a long football shape by tides raised by the black hole (center). When the star gets closer still, the tides become stronger than the gravity holding the star together, and it breaks up into a streamer (right). Much of the star's matter forms a temporary accretion disk that lights up as a quasar for a few weeks or months. (credit: modification of work by NASA/CXC/M. Weiss)

Another source of fuel for the black hole is the collision of its host galaxy with another galaxy. Some of the brightest galaxies turn out, when a detailed picture is taken, to be pairs of colliding galaxies. And most of them have quasars inside them, not easily visible to us because they are buried by enormous amounts of dust and gas.

A collision between two cars creates quite a mess, pushing parts out of their regular place. In the same way, if two galaxies collide and merge, then gas and dust (though not so much the stars) can get pushed out of their regular orbits. Some may veer close enough to the black hole in one galaxy or the other to be devoured by it and so provide the necessary fuel to power a quasar. As we saw, galaxy collisions and mergers happened most frequently when the universe was young and probably help account for the fact that quasars were most common when the universe was only about 20% of its current age.

Collisions in today's universe are less frequent, but they do happen. Once a galaxy reaches the size of the Milky Way, most of the galaxies it merges with will be much smaller galaxies—*dwarf galaxies* (see the chapter on Galaxies). These don't disrupt the big galaxy much, but they can supply some additional gas to its black hole.

By the way, if two galaxies, each of which contains a black hole, collide, then the two black holes may merge and form an even larger black hole (Figure 27.18). In this process they will emit a burst of gravitational waves. One of the main goals of the European Space Agency's planned LISA (Laser Interferometer Space Antenna) mission is to detect the gravitational wave signals from the merging of supermassive black holes.

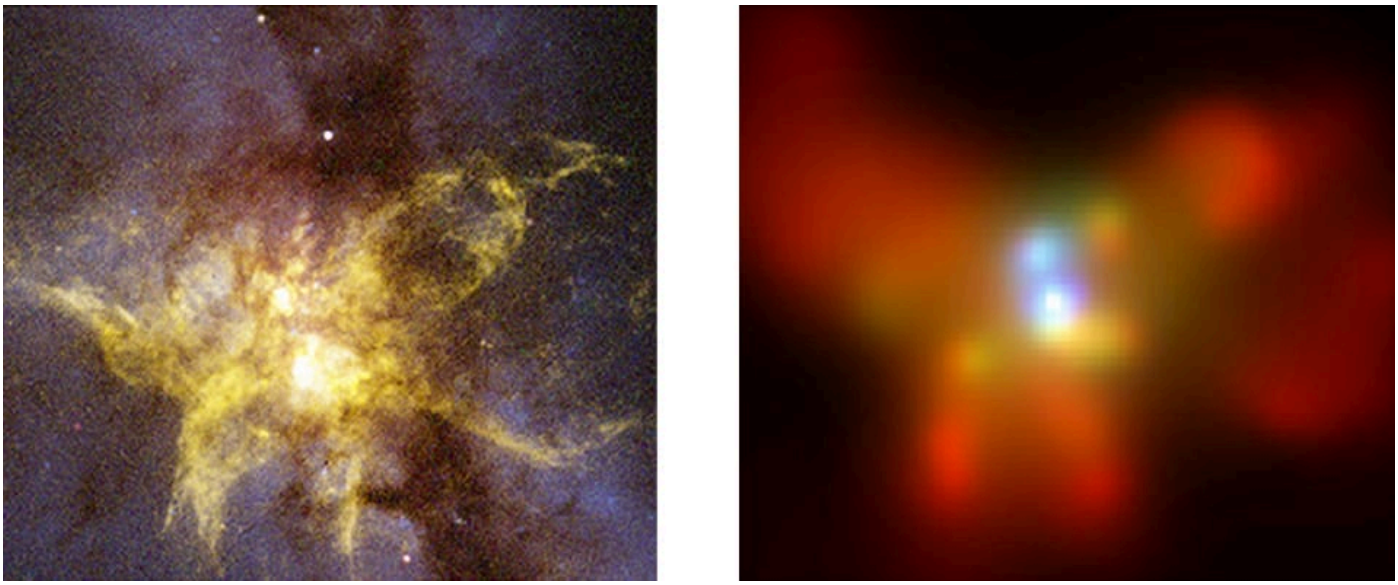


Figure 27.18 : Colliding Galaxies with Two Black Holes. We compare Hubble Space Telescope visible-light (left) and Chandra X-ray (right) images of the central regions of NGC 6240, a galaxy about 400 million light-years away. It is a prime example of a galaxy in which stars are forming, evolving, and exploding at an exceptionally rapid rate due to a relatively recent merger (30 million years ago). The Chandra image shows two bright X-ray sources, each produced by hot gas surrounding a black hole. Over the course of the next few hundred million years, the two supermassive black holes, which are about 3000 light-years apart, will drift toward each other and merge to form an even larger black hole. This detection of a binary black hole supports the idea that black holes can grow to enormous masses in the centers of galaxies by merging with nearby galaxies. (credit left: modification of work by NASA/CXC/MPE/S.Komossa et al; credit right: NASA/STScI/R. P. van der Marel, J. Gerssen)

Link to Learning

Watch two galaxies collide to form a supermassive black hole.

How Does the Black Hole Influence the Formation of Stars in the Galaxy?

We have seen that the material in galaxies can influence the growth of the black hole. The black hole in turn can also influence the galaxy in which it resides. It can do so in three ways: through its jets, through winds of particles that manage to stream away from the accretion disk, and through radiation from the accretion disk. As they stream away from the black hole, all three can either promote star formation by compressing the surrounding gas and dust—or instead suppress star formation by heating the surrounding gas and shredding molecular clouds, thereby inhibiting or preventing star formation. The outflowing energy can even be enough to halt the accretion of new material and starve the black hole of fuel. Astronomers are still trying to evaluate the relative importance of these effects in determining the overall evolution of galactic bulges and the rates of star formation.

In summary, we have seen how galaxies and supermassive black holes can each influence the evolution of the other: the galaxy supplies fuel to the black hole, and the quasar can either support or suppress star formation. The balance of these processes probably helps account for the correlation between black hole and bulge masses, but there are as yet no theories that explain quantitatively and in detail why the correlation between black hole and bulge masses is as tight as it is or why the black hole mass is always about 1/200 times the mass of the bulge.

The Birth of Black Holes and Galaxies

While the connection between quasars and galaxies is increasingly clear, the biggest puzzle of all—namely, how the supermassive black holes in galaxies got started—remains unsolved. Observations show that they existed when the universe was very young. One dramatic example is the discovery of a quasar that was already shining when the universe was only 700 million years old. What does it take to create a large black hole so quickly? A related problem is that in order to eventually build black holes containing more than 2 billion solar masses, it is necessary to have giant “seed” black holes with masses at least 2000 times the mass of the Sun—and they must somehow have been created shortly after the expansion of the universe began.

Astronomers are now working actively to develop models for how these seed black holes might have formed. Theories suggest that galaxies formed from collapsing clouds of dark matter and gas. Some of the gas formed stars, but perhaps some of the gas settled to the center where it became so concentrated that it formed a black hole. If this happened, the black hole could form right away—although this requires that the gas should not be rotating very much initially.

A more likely scenario is that the gas will have some angular momentum (rotation) that will prevent direct collapse to a black hole. In that case, the very first generation of stars will form, and some of them, according to calculations, will have masses hundreds of times that of the Sun. When these stars finish burning hydrogen, just a few million years later, the supernovae they end with will create black holes a hundred or so times the mass of the Sun. These can then merge with others or accrete the rich gas supply available at these early times.

The challenge is growing these smaller black holes quickly enough to make the much larger black holes we see a few hundred million years later. It turns out to be difficult because there are limits on how fast they can accrete matter. These should make sense to you from what we discussed earlier in the chapter. If the rate of accretion becomes too high, then the energy streaming outward from the black hole's accretion disk will become so strong as to blow away the infalling matter.

What if, instead, a collapsing gas cloud doesn't form a black hole directly or break up and form a group of regular stars, but stays together and makes one fairly massive star embedded within a dense cluster of thousands of lower mass stars and large quantities of dense gas? The massive star will have a short lifetime and will soon collapse to become a black hole. It can then begin to attract the dense gas surrounding it. But calculations show that the gravitational attraction of the many nearby stars will cause the black hole to zigzag randomly within the cluster and will prevent the formation of an accretion disk. If there is no accretion disk, then matter can fall freely into the black hole from all directions. Calculations suggest that under these conditions, a black hole even as small as 10 times the mass of the Sun could grow to more than 10 billion times the mass of the Sun by the time the universe is a billion years old.

Scientists are exploring other ideas for how to form the seeds of supermassive black holes, and this remains a very active field of research. Whatever mechanism caused the rapid formation of these supermassive black holes, they do give us a way to observe the youthful universe when it was only about five percent as old as it is now.

Link to Learning

Take a look at some new results from the Chandra X-ray Observatory about the formation of supermassive black holes in the early universe.

This page titled 16.8: Quasars as Probes of Evolution in the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 27.4: Quasars as Probes of Evolution in the Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

17: THE BIG BANG AND THE ORIGIN OF THE COSMOS

17.1: The Beginning of the Universe
17.2: The Cosmic Microwave Background
17.3: The Inflationary Universe

This page titled 17: The Big Bang and the Origin of the Cosmos is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

17.1: The Beginning of the Universe

Learning Objectives

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

- Describe what the universe was like during the first few minutes after it began to expand
- Explain how the first new elements were formed during the first few minutes after the Big Bang
- Describe how the contents of the universe change as the temperature of the universe decreases

The best evidence we have today indicates that the first galaxies did not begin to form until a few hundred million years after the Big Bang. What were things like before there were galaxies and space had not yet stretched very significantly? Amazingly, scientists have been able to calculate in some detail what was happening in the universe in the first few minutes after the Big Bang.

The History of the Idea

It is one thing to say the universe had a beginning (as the equations of general relativity imply) and quite another to describe that beginning. The Belgian priest and cosmologist Georges Lemaître was probably the first to propose a specific model for the Big Bang itself (Figure 29.11). He envisioned all the matter of the universe starting in one great bulk he called the *primeval atom*, which then broke into tremendous numbers of pieces. Each of these pieces continued to fragment further until they became the present atoms of the universe, created in a vast nuclear fission. In a popular account of his theory, Lemaître wrote, “The evolution of the world could be compared to a display of fireworks just ended—some few red wisps, ashes, and smoke. Standing on a well-cooled cinder, we see the slow fading of the suns and we try to recall the vanished brilliance of the origin of the worlds.”



Figure 29.11 : Abbé Georges Lemaître (1894–1966). This Belgian cosmologist studied theology at Mechelen and mathematics and physics at the University of Leuven. It was there that he began to explore the expansion of the universe and postulated its explosive beginning. He actually predicted Hubble’s law 2 years before its verification, and he was the first to consider seriously the physical processes by which the universe began.

Link to Learning

View a short video about the work of Lemaître, considered by some to be the father of the Big Bang theory.

Physicists today know much more about nuclear physics than was known in the 1920s, and they have shown that the primeval fission model cannot be correct. Yet Lemaître's vision was in some respects quite prophetic. We still believe that everything was together at the beginning; it was just not in the form of matter we now know. Basic physical principles tell us that when the universe was much denser, it was also much hotter, and that it cools as it expands, much as gas cools when sprayed from an aerosol can.

By the 1940s, scientists knew that fusion of hydrogen into helium was the source of the Sun's energy. Fusion requires high temperatures, and the early universe must have been hot. Based on these ideas, American physicist George Gamow (Figure 29.12) suggested a universe with a different kind of beginning that involved nuclear fusion instead of fission. Ralph Alpher worked out the details for his PhD thesis, and the results were published in 1948. (Gamow, who had a quirky sense of humor, decided at the last minute to add the name of physicist Hans Bethe to their paper, so that the coauthors on this paper about the beginning of things would be Alpher, Bethe, and Gamow, a pun on the first three letters of the Greek alphabet: alpha, beta, and gamma.) Gamow's universe started with fundamental particles that built up the heavy elements by fusion in the Big Bang.

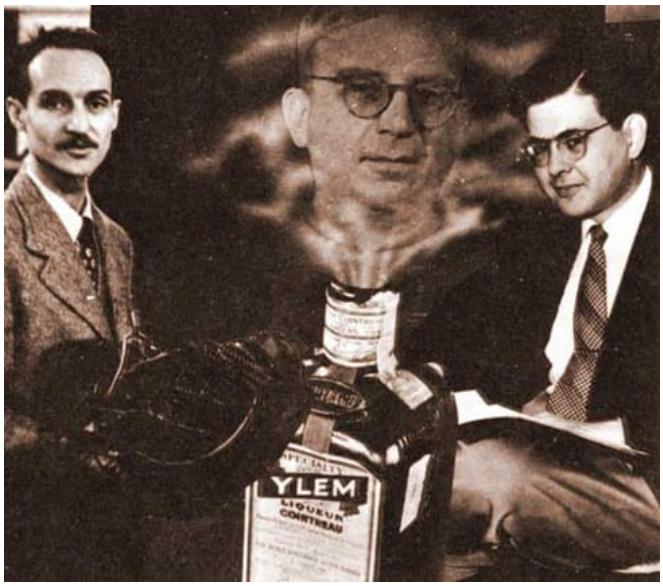


Figure 29.12 : George Gamow and Collaborators. This composite image shows George Gamow emerging like a genie from a bottle of ylem, a Greek term for the original substance from which the world formed. Gamow revived the term to describe the material of the hot Big Bang. Flanking him are Robert Herman (left) and Ralph Alpher (right), with whom he collaborated in working out the physics of the Big Bang. (The modern composer Karlheinz Stockhausen was inspired by Gamow's ideas to write a piece of music called *Ylem*, in which the players actually move away from the stage as they perform, simulating the expansion of the universe.)

Gamow's ideas were close to our modern view, except we now know that the early universe remained hot enough for fusion for only a short while. Thus, only the three lightest elements—hydrogen, helium, and a small amount of lithium—were formed in appreciable abundances at the beginning. The heavier elements formed later in stars. Since the 1940s, many astronomers and physicists have worked on a detailed theory of what happened in the early stages of the universe.

The First Few Minutes

Let's start with the first few minutes following the Big Bang. Three basic ideas hold the key to tracing the changes that occurred during the time just after the universe began. The first, as we have already mentioned, is that the universe cools as it expands. Figure 29.13 shows how the temperature changes with the passage of time. Note that a huge span of time, from a tiny fraction of a second to billions of years, is summarized in this diagram. In the first fraction of a second, the universe was unimaginably hot. By the time 0.01 second had elapsed, the temperature had dropped to 100 billion (10^{11}) K. After about 3 minutes, it had fallen to about 1 billion (10^{9}) K, still some 70 times hotter than the interior of the Sun. After a few hundred thousand years, the temperature was down to a mere 3000 K, and the universe has continued to cool since that time.

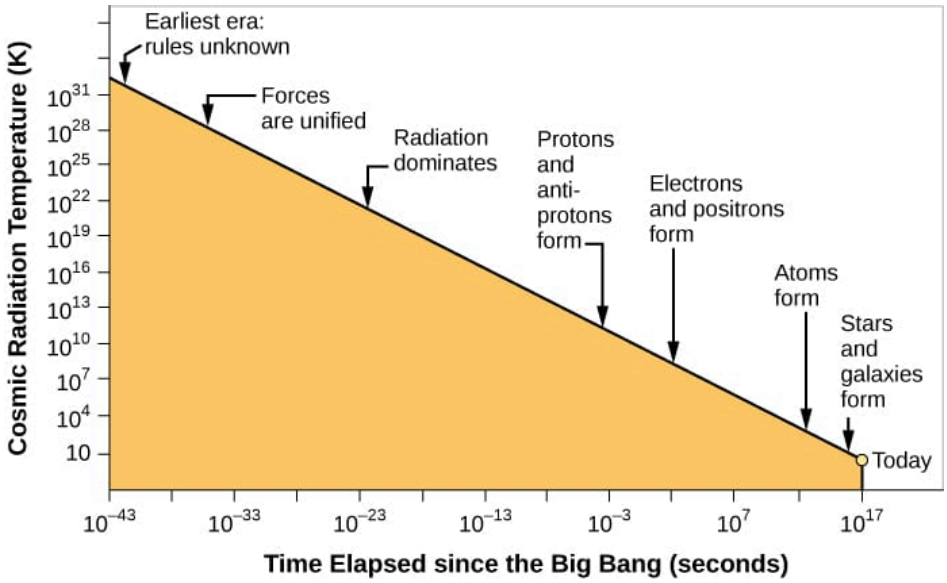


Figure 29.13 : Temperature of the Universe. This graph shows how the temperature of the universe varies with time as predicted by the standard model of the Big Bang. Note that both the temperature (vertical axis) and the time in seconds (horizontal axis) change over vast scales on this compressed diagram.

All of these temperatures but the last are derived from theoretical calculations since (obviously) no one was there to measure them directly. As we shall see in the next section, however, we have actually detected the feeble glow of radiation emitted at a time when the universe was a few hundred thousand years old. We can measure the characteristics of that radiation to learn what things were like long ago. Indeed, the fact that we have found this ancient glow is one of the strongest arguments in favor of the Big Bang model.

The second step in understanding the evolution of the universe is to realize that at very early times, it was so hot that it contained mostly radiation (and not the matter that we see today). The photons that filled the universe could collide and produce material particles; that is, under the conditions just after the Big Bang, energy could turn into matter (and matter could turn into energy). We can calculate how much mass is produced from a given amount of energy by using Einstein's formula Math input error

The idea that energy could turn into matter in the universe at large is a new one for many students, since it is not part of our everyday experience. That's because, when we compare the universe today to what it was like right after the Big Bang, we live in cold, hard times. The photons in the universe today typically have far-less energy than the amount required to make new matter. In the discussion on the source of the Sun's energy in The Sun: A Nuclear Powerhouse, we briefly mentioned that when subatomic particles of matter and *antimatter* collide, they turn into pure energy. But the reverse, energy turning into matter and antimatter, is equally possible. This process has been observed in particle accelerators around the world. If we have enough energy, under the right circumstances, new particles of matter (and antimatter) are indeed created —and the conditions were right during the first few minutes after the expansion of the universe began.

Our third key point is that the hotter the universe was, the more energetic were the photons available to make matter and antimatter (see Figure 29.13). To take a specific example, at a temperature of 6 billion (6×10^9) K, the collision of two typical photons can create an electron and its antimatter counterpart, a positron. If the temperature exceeds 10^{14} K, much more massive protons and antiprotons can be created.

The Evolution of the Early Universe

Keeping these three ideas in mind, we can trace the evolution of the universe from the time it was about 0.01 second old and had a temperature of about 100 billion K. Why not begin at the very beginning? There are as yet no theories that allow us penetrate to a time before about 10^{-43} second (this number is a decimal point followed by 42 zeros and then a one). It is so small that we cannot relate it to anything in our everyday experience. When the universe was that young, its density was so high that the theory of general relativity is not adequate to describe it, and even the concept of time breaks down.

Scientists, by the way, have been somewhat more successful in describing the universe when it was older than 10^{-43} second but

still less than about 0.01 second old. We will take a look at some of these ideas later in this chapter, but for now, we want to start with somewhat more familiar situations.

By the time the universe was 0.01 second old, it consisted of a soup of matter and radiation; the matter included protons and neutrons, leftovers from an even younger and hotter universe. Each particle collided rapidly with other particles. The temperature was no longer high enough to allow colliding photons to produce neutrons or protons, but it was sufficient for the production of electrons and positrons (Figure 29.14). There was probably also a sea of exotic subatomic particles that would later play a role as dark matter. All the particles jiggled about on their own; it was still much too hot for protons and neutrons to combine to form the nuclei of atoms.

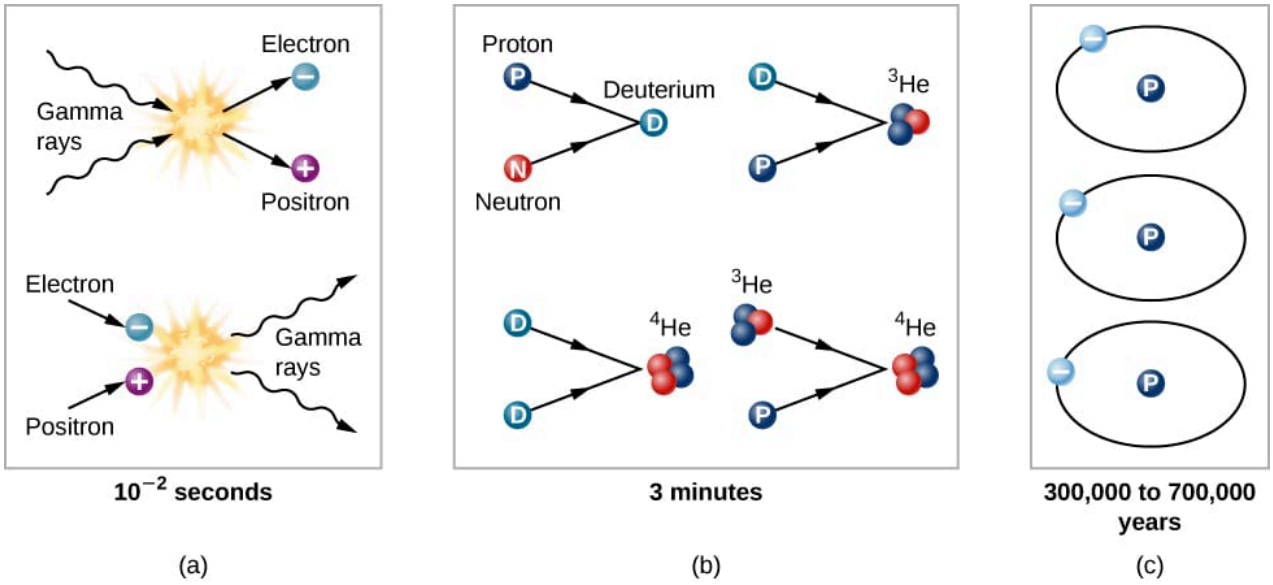


Figure 29.14 : Particle Interactions in the Early Universe. (a) In the first fractions of a second, when the universe was very hot, energy was converted into particles and antiparticles. The reverse reaction also happened: a particle and antiparticle could collide and produce energy. (b) As the temperature of the universe decreased, the energy of typical photons became too low to create matter. Instead, existing particles fused to create such nuclei as deuterium and helium. (c) Later, it became cool enough for electrons to settle down with nuclei and make neutral atoms. Most of the universe was still hydrogen.

Think of the universe at this time as a seething cauldron, with photons colliding and interchanging energy, and sometimes being destroyed to create a pair of particles. The particles also collided with one another. Frequently, a matter particle and an antimatter particle met and turned each other into a burst of gamma-ray radiation.

Among the particles created in the early phases of the universe was the ghostly neutrino (see The Sun: A Nuclear Powerhouse), which today interacts only very rarely with ordinary matter. In the crowded conditions of the very early universe, however, neutrinos ran into so many electrons and positrons that they experienced frequent interactions despite their “antisocial” natures.

By the time the universe was a little more than 1 second old, the density had dropped to the point where neutrinos no longer interacted with matter but simply traveled freely through space. In fact, these neutrinos should now be all around us. Since they have been traveling through space unimpeded (and hence unchanged) since the universe was 1 second old, measurements of their properties would offer one of the best tests of the Big Bang model. Unfortunately, the very characteristic that makes them so useful—the fact that they interact so weakly with matter that they have survived unaltered for all but the first second of time—also renders them unable to be measured, at least with present techniques. Perhaps someday someone will devise a way to capture these elusive messengers from the past.

Atomic Nuclei Form

When the universe was about 3 minutes old and its temperature was down to about 900 million K, protons and neutrons could combine. At higher temperatures, these atomic nuclei had immediately been blasted apart by interactions with high-energy photons and thus could not survive. But at the temperatures and densities reached between 3 and 4 minutes after the beginning, deuterium (a proton and neutron) lasted long enough that collisions could convert some of it into helium, (Figure 29.14). In essence, the entire universe was acting the way centers of stars do today—fusing new elements from simpler components. In

addition, a little bit of element 3, lithium, could also form.

This burst of cosmic fusion was only a brief interlude, however. By 4 minutes after the Big Bang, more helium was having trouble forming. The universe was still expanding and cooling down. After the formation of helium and some lithium, the temperature had dropped so low that the fusion of helium nuclei into still-heavier elements could not occur. No elements beyond lithium could form in the first few minutes. That 4-minute period was the end of the time when the entire universe was a fusion factory. In the cool universe we know today, the fusion of new elements is limited to the centers of stars and the explosions of supernovae.

Still, the fact that the Big Bang model allows the creation of a good deal of helium is the answer to a long-standing mystery in astronomy. Put simply, there is just too much helium in the universe to be explained by what happens inside stars. All the generations of stars that have produced helium since the Big Bang cannot account for the quantity of helium we observe. Furthermore, even the oldest stars and the most distant galaxies show significant amounts of helium. These observations find a natural explanation in the synthesis of helium by the Big Bang itself during the first few minutes of time. We estimate that *10 times more helium* was manufactured in the first 4 minutes of the universe than in all the generations of stars during the succeeding 10 to 15 billion years.

Learning from Deuterium

We can learn many things from the way the early universe made atomic nuclei. It turns out that all of the deuterium (a hydrogen nucleus with a neutron in it) in the universe was formed during the first 4 minutes. In stars, any region hot enough to fuse two protons to form a deuterium nucleus is also hot enough to change it further—either by destroying it through a collision with an energetic photon or by converting it into helium through nuclear reactions.

The amount of deuterium that can be produced in the first 4 minutes of creation depends on the density of the universe at the time deuterium was formed. If the density were relatively high, nearly all the deuterium would have been converted into helium through interactions with protons, just as it is in stars. If the density were relatively low, then the universe would have expanded and thinned out rapidly enough that some deuterium would have survived. The amount of deuterium we see today thus gives us a clue to the density of the universe when it was about 4 minutes old. Theoretical models can relate the density then to the density now; thus, measurements of the abundance of deuterium today can give us an estimate of the current density of the universe.

The measurements of deuterium indicate that the present-day density of ordinary matter—protons and neutrons—is about 5×10^{-28} kg/m^3. Deuterium can only provide an estimate of the density of ordinary matter because the abundance of deuterium is determined by the particles that interact to form it, namely protons and neutrons alone. From the abundance of deuterium, we know that not enough protons and neutrons are present, by a factor of about 20, to produce a critical-density universe.

We do know, however, that there are dark matter particles that add to the overall matter density of the universe, which is then higher than what is calculated for ordinary matter alone. Because dark matter particles do not affect the production of deuterium, measurement of the deuterium abundance cannot tell us how much dark matter exists. Dark matter is made of some exotic kind of particle, not yet detected in any earthbound laboratory. It is definitely not made of protons and neutrons like the readers of this book.

This page titled 17.1: The Beginning of the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 29.4: The Beginning of the Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

17.2: The Cosmic Microwave Background

Learning Objectives

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

- Explain why we can observe the afterglow of the hot, early universe
- Discuss the properties of this afterglow as we see it today, including its average temperature and the size of its temperature fluctuations
- Describe open, flat, and curved universes and explain which type of universe is supported by observations
- Summarize our current knowledge of the basic properties of the universe including its age and contents

The description of the first few minutes of the universe is based on theoretical calculations. It is crucial, however, that a scientific theory should be testable. What predictions does it make? And do observations show those predictions to be accurate? One success of the theory of the first few minutes of the universe is the correct prediction of the amount of helium in the universe.

Another prediction is that a significant milestone in the history of the universe occurred about 380,000 years after the Big Bang. Scientists have directly observed what the universe was like at this early stage, and these observations offer some of the strongest support for the Big Bang theory. To find out what this milestone was, let's look at what theory tells us about what happened during the first few hundred thousand years after the Big Bang.

The fusion of helium and lithium was completed when the universe was about 4 minutes old. The universe then continued to resemble the interior of a star in some ways for a few hundred thousand years more. It remained hot and opaque, with radiation being scattered from one particle to another. It was still too hot for electrons to "settle down" and become associated with a particular nucleus; such free electrons are especially effective at scattering photons, thus ensuring that no radiation ever got very far in the early universe without having its path changed. In a way, the universe was like an enormous crowd right after a popular concert; if you get separated from a friend, even if he is wearing a flashing button, it is impossible to see through the dense crowd to spot him. Only after the crowd clears is there a path for the light from his button to reach you.

The Universe Becomes Transparent

Not until a few hundred thousand years after the Big Bang, when the temperature had dropped to about 3000 K and the density of atomic nuclei to about 1000 per cubic centimeter, did the electrons and nuclei manage to combine to form stable atoms of hydrogen and helium (Figure 29.14). With no free electrons to scatter photons, the universe became transparent for the first time in cosmic history. From this point on, matter and radiation interacted much less frequently; we say that they *decoupled* from each other and evolved separately. Suddenly, electromagnetic radiation could really travel, and it has been traveling through the universe ever since.

Discovery of the Cosmic Background Radiation

If the model of the universe described in the previous section is correct, then—as we look far outward in the universe and thus far back in time—the first "afterglow" of the hot, early universe should still be detectable. Observations of it would be very strong evidence that our theoretical calculations about how the universe evolved are correct. As we shall see, we have indeed detected the radiation emitted at this photon decoupling time, when radiation began to stream freely through the universe without interacting with matter (Figure 29.15).

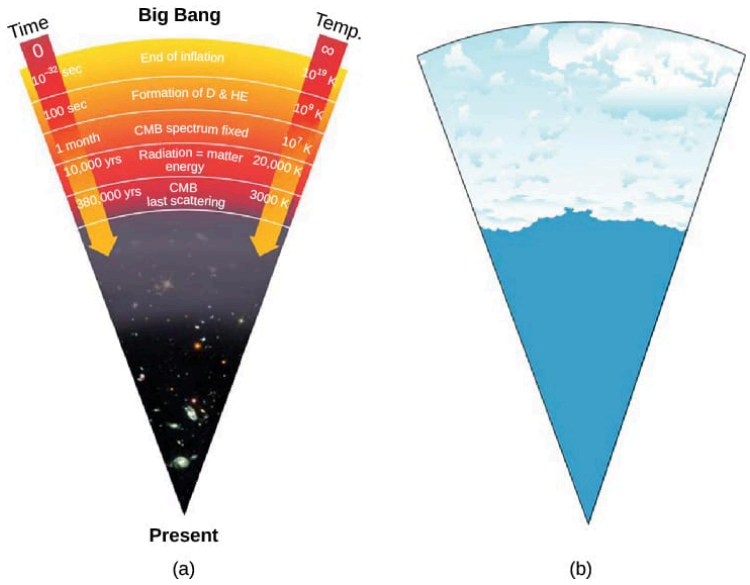


Figure 29.15 : Cosmic Microwave Background and Clouds Compared. (a) Early in the universe, photons (electromagnetic energy) were scattering off the crowded, hot, charged particles and could not get very far without colliding with another particle. But after electrons and photons settled into neutral atoms, there was far less scattering, and photons could travel over vast distances. The universe became transparent. As we look out in space and back in time, we can't see back beyond this time. (b) This is similar to what happens when we see clouds in Earth's atmosphere. Water droplets in a cloud scatter light very efficiently, but clear air lets light travel over long distances. So as we look up into the atmosphere, our vision is blocked by the cloud layers and we can't see beyond them. (credit: modification of work by NASA)

The detection of this afterglow was initially an accident. In the late 1940s, Ralph Alpher and Robert Herman, working with George Gamow, realized that just before the universe became transparent, it must have been radiating like a blackbody at a temperature of about 3000 K—the temperature at which hydrogen atoms could begin to form. If we could have seen that radiation just after neutral atoms formed, it would have resembled radiation from a reddish star. It was as if a giant fireball filled the whole universe.

But that was nearly 14 billion years ago, and, in the meantime, the scale of the universe has increased a thousand fold. This expansion has increased the wavelength of the radiation by a factor of 1000 (see Figure 29.7). According to Wien's law, which relates wavelength and temperature, the expansion has correspondingly lowered the temperature by a factor of 1000 (see the chapter on Radiation and Spectra). The cosmic background behaves like a blackbody and should therefore have a spectrum that obeys Wien's Law.

Alpher and Herman predicted that the glow from the fireball should now be at radio wavelengths and should resemble the radiation from a blackbody at a temperature only a few degrees above absolute zero. Since the fireball was everywhere throughout the universe, the radiation left over from it should also be everywhere. If our eyes were sensitive to radio wavelengths, the whole sky would appear to glow very faintly. However, our eyes can't see at these wavelengths, and at the time Alpher and Herman made their prediction, there were no instruments that could detect the glow. Over the years, their prediction was forgotten.

In the mid-1960s, in Holmdel, New Jersey, Arno Penzias and Robert Wilson of AT&T's Bell Laboratories had built a delicate microwave antenna (Figure 29.16) to measure astronomical sources, including supernova remnants like Cassiopeia A (see the chapter on The Death of Stars). They were plagued with some unexpected background noise, just like faint static on a radio, which they could not get rid of. The puzzling thing about this radiation was that it seemed to be coming from all directions at once. This is very unusual in astronomy: after all, most radiation has a specific direction where it is strongest—the direction of the Sun, or a supernova remnant, or the disk of the Milky Way, for example.



Figure 29.16 : Robert Wilson (left) and Arno Penzias (right). These two scientists are standing in front of the horn-shaped antenna with which they discovered the cosmic background radiation. The photo was taken in 1978, just after they received the Nobel Prize in physics.

Penzias and Wilson at first thought that any radiation appearing to come from all directions must originate from inside their telescope, so they took everything apart to look for the source of the noise. They even found that some pigeons had roosted inside the big horn-shaped antenna and had left (as Penzias delicately put it) “a layer of white, sticky, dielectric substance coating the inside of the antenna.” However, nothing the scientists did could reduce the background radiation to zero, and they reluctantly came to accept that it must be real, and it must be coming from space.

Penzias and Wilson were not cosmologists, but as they began to discuss their puzzling discovery with other scientists, they were quickly put in touch with a group of astronomers and physicists at Princeton University (a short drive away). These astronomers had—as it happened—been redoing the calculations of Alpher and Herman from the 1940s and also realized that the radiation from the decoupling time should be detectable as a faint afterglow of radio waves. The different calculations of what the observed temperature would be for this cosmic microwave background (CMB)[2] were uncertain, but all predicted less than 40 K.

Penzias and Wilson found the distribution of intensity at different radio wavelengths to correspond to a temperature of 3.5 K. This is very cold—closer to absolute zero than most other astronomical measurements—and a testament to how much space (and the waves within it) has stretched. Their measurements have been repeated with better instruments, which give us a reading of 2.73 K. So Penzias and Wilson came very close. Rounding this value, scientists often refer to “the 3-degree microwave background.”

Many other experiments on Earth and in space soon confirmed the discovery by Penzias and Wilson: The radiation was indeed coming from all directions (it was isotropic) and matched the predictions of the Big Bang theory with remarkable precision. Penzias and Wilson had inadvertently observed the glow from the primeval fireball. They received the Nobel Prize for their work in 1978. And just before his death in 1966, Lemaître learned that his “vanished brilliance” had been discovered and confirmed.

Link to Learning

You may enjoy watching *Three Degrees*, a 26-minute video from Bell Labs about Penzias and Wilson’s discovery of the cosmic background radiation (with interesting historical footage).

Properties of the Cosmic Microwave Background

One issue that worried astronomers is that Penzias and Wilson were measuring the background radiation filling space through Earth’s atmosphere. What if that atmosphere is a source of radio waves or somehow affected their measurements? It would be better to measure something this important from space.

The first accurate measurements of the CMB were made with a satellite orbiting Earth. Named the Cosmic Background Explorer

(COBE), it was launched by NASA in November 1989. The data it received quickly showed that the CMB closely matches that expected from a blackbody with a temperature of 2.73 K (Figure 29.17). This is exactly the result expected if the CMB was indeed redshifted radiation emitted by a hot gas that filled all of space shortly after the universe began.

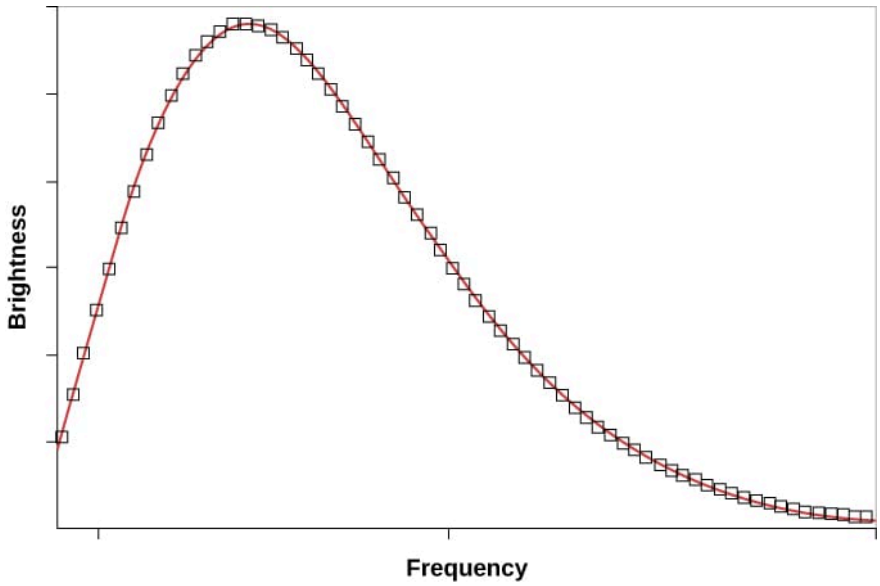


Figure 29.17 : Cosmic Background Radiation. The solid line shows how the intensity of radiation should change with wavelength for a blackbody with a temperature of 2.73 K. The boxes show the intensity of the cosmic background radiation as measured at various wavelengths by COBE's instruments. The fit is perfect. When this graph was first shown at a meeting of astronomers, they gave it a standing ovation.

The first important conclusion from measurements of the CMB, therefore, is that the universe we have today has indeed evolved from a hot, uniform state. This observation also provides direct support for the general idea that we live in an evolving universe, since the universe is cooler today than it was in the beginning.

Small Differences in the CMB

It was known even before the launch of COBE that the CMB is extremely *isotropic*. In fact, its uniformity in every direction is one of the best confirmations of the cosmological principle— that the universe is homogenous and isotropic.

According to our theories, however, the temperature could not have been *perfectly* uniform when the CMB was emitted. After all, the CMB is radiation that was scattered from the particles in the universe at the time of decoupling. If the radiation were completely smooth, then all those particles must have been distributed through space absolutely evenly. Yet it is those particles that have become all the galaxies and stars (and astronomy students) that now inhabit the cosmos. Had the particles been completely smoothly distributed, they could not have formed all the large-scale structures now present in the universe—the clusters and superclusters of galaxies discussed in the last few chapters.

The early universe must have had tiny density fluctuations from which such structures could evolve. Regions of higher-than-average density would have attracted additional matter and eventually grown into the galaxies and clusters that we see today. It turns out that these denser regions would appear to us to be colder spots, that is, they would have lower-than-average temperatures.

The reason that temperature and density are related can be explained this way. At the time of decoupling, photons in a slightly denser portion of space had to expend some of their energy to escape the gravitational force exerted by the surrounding gas. In losing energy, the photons became slightly colder than the overall average temperature at the time of decoupling. Vice versa, photons that were located in a slightly less dense portion of space lost less energy upon leaving it than other photons, thus appearing slightly hotter than average. Therefore, if the seeds of present-day galaxies existed at the time that the CMB was emitted, we should see some slight variations in the CMB temperature as we look in different directions in the sky.

Scientists working with the data from the COBE satellite did indeed detect very subtle temperature differences—about 1 part in 100,000—in the CMB. The regions of lower-than-average temperature come in a variety of sizes, but even the smallest of the colder areas detected by COBE is far too large to be the precursor of an individual galaxy, or even a supercluster of galaxies. This

is because the COBE instrument had “blurry vision” (poor resolution) and could only measure large patches of the sky. We needed instruments with “sharper vision.”

The most detailed measurements of the CMB have been obtained by two satellites launched more recently than COBE. The results from the first of these satellites, the Wilkinson Microwave Anisotropy Probe (WMAP) spacecraft, were published in 2003. In 2015, measurements from the Planck satellite extended the WMAP measurements to even-higher spatial resolution and lower noise (Figure 29.18).

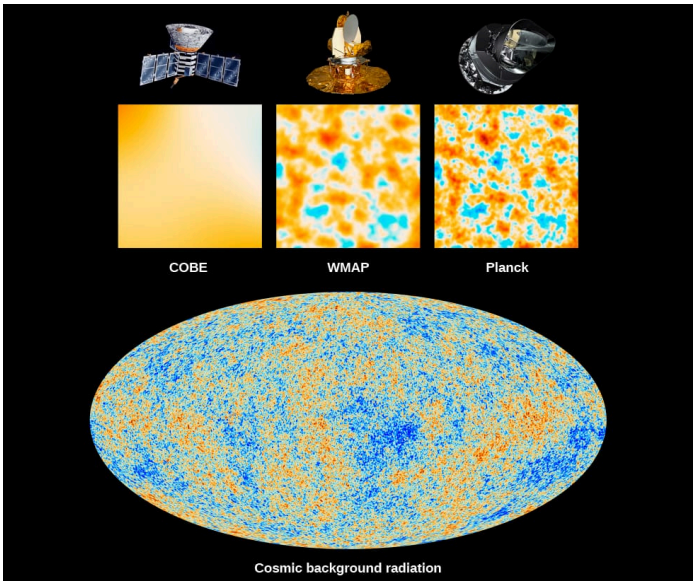


Figure 29.18 : CMB Observations. This comparison shows how much detail can be seen in the observations of three satellites used to measure the CMB. The CMB is a snapshot of the oldest light in our universe, imprinted on the sky when the universe was just about 380,000 years old. The first spacecraft, launched in 1989, is NASA’s Cosmic Background Explorer, or COBE. WMAP was launched in 2001, and Planck was launched in 2009. The three panels show 10-square-degree patches of all-sky maps. This cosmic background radiation image (bottom) is an all-sky map of the CMB as observed by the Planck mission. The colors in the map represent different temperatures: red for warmer and blue for cooler. These tiny temperature fluctuations correspond to regions of slightly different densities, representing the seeds of all future structures: the stars, galaxies, and galaxy clusters of today. (credit top: modification of work by NASA/JPL-Caltech/ESA; credit bottom: modification of work by ESA and the Planck Collaboration)

Theoretical calculations show that the sizes of the hot and cold spots in the CMB depend on the geometry of the universe and hence on its total density. (It’s not at all obvious that it should do so, and it takes some pretty fancy calculations—way beyond the level of our text—to make the connection, but having such a dependence is very useful.) The total density we are discussing here includes both the amount of mass in the universe and the mass equivalent of the dark energy. That is, we must add together mass and energy: ordinary matter, dark matter, and the dark energy that is speeding up the expansion.

To see why this works, remember (from the chapter on Black Holes and Curved Spacetime) that with his theory of general relativity, Einstein showed that matter can curve space and that the amount of curvature depends on the amount of matter present. Therefore, the total amount of matter in the universe (including dark matter and the equivalent matter contribution by dark energy), determines the overall geometry of space. Just like the geometry of space around a black hole has a curvature to it, so the entire universe may have a curvature. Let’s take a look at the possibilities (Figure 29.19).

If the density of matter is higher than the critical density, the universe will eventually collapse. In such a closed universe, two initially parallel rays of light will eventually meet. This kind of geometry is referred to as spherical geometry. If the density of matter is less than critical, the universe will expand forever. Two initially parallel rays of light will diverge, and this is referred to as hyperbolic geometry. In a critical-density universe, two parallel light rays never meet, and the expansion comes to a halt only at some time infinitely far in the future. We refer to this as a flat universe, and the kind of Euclidean geometry you learned in high school applies in this type of universe.

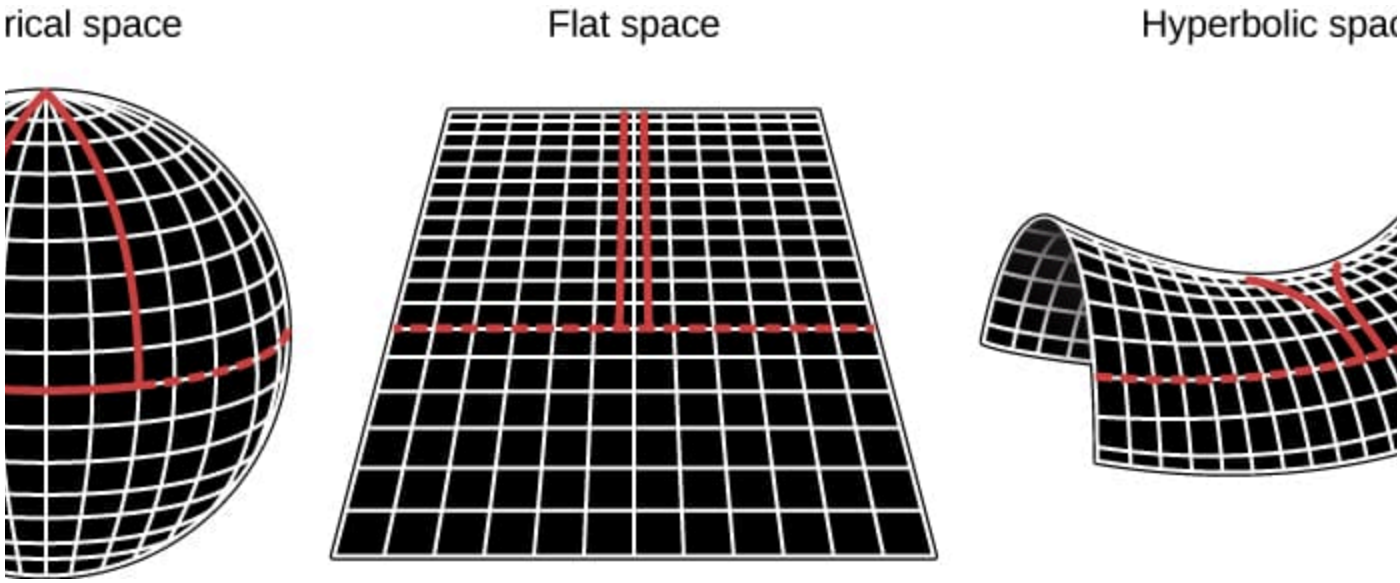


Figure 29.19 : Picturing Space Curvature for the Entire Universe. The density of matter and energy determines the overall geometry of space. If the density of the universe is greater than the critical density, then the universe will ultimately collapse and space is said to be *closed* like the surface of a sphere. If the density exactly equals the critical density, then space is *flat* like a sheet of paper; the universe will expand forever, with the rate of expansion coming to a halt infinitely far in the future. If the density is less than critical, then the expansion will continue forever and space is said to be *open* and negatively curved like the surface of a saddle (where more space than you expect opens up as you move farther away). Note that the red lines in each diagram show what happens in each kind of space—they are initially parallel but follow different paths depending on the curvature of space. Remember that these drawings are trying to show how space for the entire universe is "warped"—this can't be seen locally in the small amount of space that we humans occupy.

If the density of the universe is equal to the critical density, then the hot and cold spots in the CMB should typically be about a degree in size. If the density is greater than critical, then the typical sizes will be larger than one degree. If the universe has a density less than critical, then the structures will appear smaller. In Figure 29.20, you can see the differences easily. WMAP and Planck observations of the CMB confirmed earlier experiments that we do indeed live in a flat, critical-density universe.

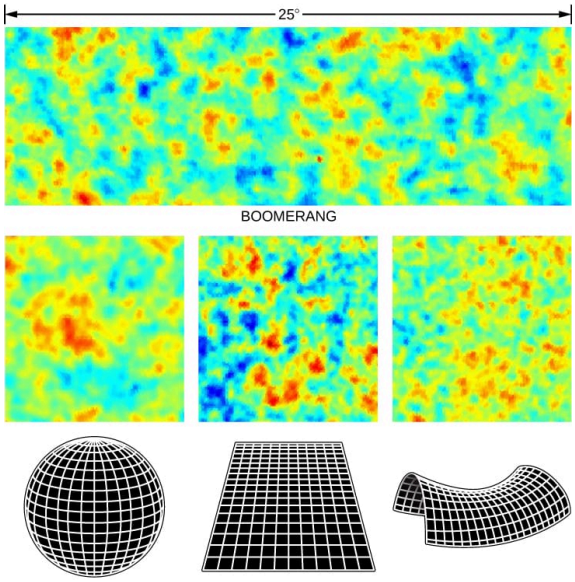


Figure 29.20 : Comparison of CMB Observations with Possible Models of the Universe. Cosmological simulations predict that if our universe has critical density, then the CMB images will be dominated by hot and cold spots of around one degree in size (bottom center). If, on the other hand, the density is higher than critical (and the universe will ultimately collapse), then the images' hot and cold spots will appear larger than one degree (bottom left). If the density of the universe is less than critical (and the expansion will continue forever), then the structures will appear smaller (bottom right). As the measurements show, the universe is at critical density. The measurements shown were made by a balloon-borne instrument called BOOMERanG (Balloon Observations of Millimetric Extragalactic Radiation and Geophysics), which was flown in Antarctica. Subsequent satellite observations by WMAP and Planck confirm the BOOMERanG result. (credit: modification of work by NASA)

Key numbers from an analysis of the Planck data give us the best values currently available for some of the basic properties of the universe:

- Age of universe: 13.799 ± 0.038 billion years (Note: That means we know the age of the universe to within 38 million years. Amazing!)
- Hubble constant: 67.31 ± 0.96 kilometers/second/million parsecs (in the units we've been using, 20.65 kilometers/second/million light-years)
- Fraction of universe's content that is "dark energy": $68.5\% \pm 1.3\%$
- Fraction of the universe's content that is matter: $31.5\% \pm 1.3\%$

Note that this value for the Hubble constant is slightly smaller than the value of 70 kilometers/second/million parsecs that we have adopted in this book. In fact, the value derived from measurements of redshifts is 73 kilometers/second/million parsecs. So precise is modern cosmology these days that scientists are working hard to resolve this discrepancy. The fact that the difference between these two independent measurements is so small is actually a remarkable achievement. Only a few decades ago, astronomers were arguing about whether the Hubble constant was around 50 kilometers/second/million parsecs or 100 kilometers/second/million parsecs.

Analysis of Planck data also shows that ordinary matter (mainly protons and neutrons) makes up 4.9% of the total density. Dark matter plus normal matter add up to 31.5% of the total density. Dark energy contributes the remaining 68.5%. The age of the universe at decoupling—that is, when the CMB was emitted—was 380,000 years.

Perhaps the most surprising result from the high-precision measurements by WMAP and the even higher-precision measurements from Planck is that there were no surprises. The model of cosmology with ordinary matter at about 5%, dark matter at about 25%, and dark energy about 70% has survived since the late 1990s when cosmologists were forced in that direction by the supernovae data. In other words, the very strange universe that we have been describing, with only about 5% of its contents being made up of the kinds of matter we are familiar with here on Earth, really seems to be the universe we live in.

After the CMB was emitted, the universe continued to expand and cool off. By 400 to 500 million years after the Big Bang, the very first stars and galaxies had already formed. Deep in the interiors of stars, matter was reheated, nuclear reactions were ignited, and the more gradual synthesis of the heavier elements that we have discussed throughout this book began.

We conclude this quick tour of our model of the early universe with a reminder. You must not think of the Big Bang as a *localized* explosion *in space*, like an exploding superstar. There were no boundaries and there was no single site where the explosion happened. It was an explosion *of space* (and time and matter and energy) that happened everywhere in the universe. All matter and energy that exist today, including the particles of which you are made, came from the Big Bang. We were, and still are, in the midst of a Big Bang; it is all around us.

Footnotes

- 2Recall that microwaves are in the radio region of the electromagnetic spectrum.

This page titled 17.2: The Cosmic Microwave Background is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 29.5: The Cosmic Microwave Background by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

17.3: The Inflationary Universe

Learning Objectives

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

- Describe two important properties of the universe that the simple Big Bang model cannot explain
- Explain why these two characteristics of the universe can be accounted for if there was a period of rapid expansion (inflation) of the universe just after the Big Bang
- Name the four forces that control all physical processes in the universe

The hot Big Bang model that we have been describing is remarkably successful. It accounts for the expansion of the universe, explains the observations of the CMB, and correctly predicts the abundances of the light elements. As it turns out, this model also predicts that there should be exactly three types of neutrinos in nature, and this prediction has been confirmed by experiments with high-energy accelerators. We can't relax just yet, however. This standard model of the universe doesn't explain *all* the observations we have made about the universe as a whole.

Problems with the Standard Big Bang Model

There are a number of characteristics of the universe that can only be explained by considering further what might have happened before the emission of the CMB. One problem with the standard Big Bang model is that it does not explain why the density of the universe is equal to the critical density. The mass density could have been, after all, so low and the effects of dark energy so high that the expansion would have been too rapid to form any galaxies at all. Alternatively, there could have been so much matter that the universe would have already begun to contract long before now. Why is the universe balanced so precisely on the knife edge of the critical density?

Another puzzle is the remarkable *uniformity* of the universe. The temperature of the CMB is the same to about 1 part in 100,000 everywhere we look. This sameness might be expected if all the parts of the visible universe were in contact at some point in time and had the time to come to the same temperature. In the same way, if we put some ice into a glass of lukewarm water and wait a while, the ice will melt and the water will cool down until they are the same temperature.

However, if we accept the standard Big Bang model, all parts of the visible universe were *not* in contact at any time. The fastest that information can go from one point to another is the speed of light. There is a maximum distance that light can have traveled from any point since the time the universe began—that's the distance light could have covered since then. This distance is called that point's *horizon distance* because anything farther away is "below its horizon"—unable to make contact with it. One region of space separated by more than the horizon distance from another has been completely isolated from it through the entire history of the universe.

If we measure the CMB in two opposite directions in the sky, we are observing regions that were significantly beyond each other's horizon distance at the time the CMB was emitted. We can see both regions, but *they* can never have seen each other. Why, then, are their temperatures so precisely the same? According to the standard Big Bang model, they have never been able to exchange information, and there is no reason they should have identical temperatures. (It's a little like seeing the clothes that all the students wear at two schools in different parts of the world become identical, without the students ever having been in contact.) The only explanation we could suggest was simply that the universe somehow *started out* being absolutely uniform (which is like saying all students were born liking the same clothes). Scientists are always uncomfortable when they must appeal to a special set of initial conditions to account for what they see.

The Inflationary Hypothesis

Some physicists suggested that these fundamental characteristics of the cosmos—its flatness and uniformity—can be explained if shortly after the Big Bang (and before the emission of the CMB), the universe experienced a sudden increase in size. A model universe in which this rapid, early expansion occurs is called an inflationary universe. The inflationary universe is identical to the Big Bang universe for all time after the first 10^{-30} second. Prior to that, the model suggests that there was a brief period of

extraordinarily rapid expansion or inflation, during which the scale of the universe increased by a factor of about 10^{50} times more than predicted by standard Big Bang models (Figure 29.25).

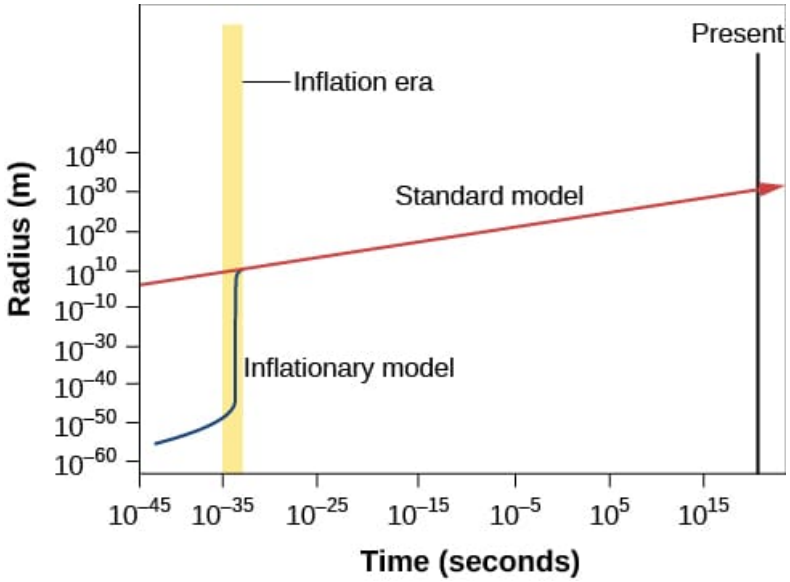


Figure 29.25 : Expansion of the Universe. This graph shows how the scale factor of the observable universe changes with time for the standard Big Bang model (red line) and for the inflationary model (blue line). (Note that the time scale at the bottom is extremely compressed.) During inflation, regions that were very small and in contact with each other are suddenly blown up to be much larger and outside each other's horizon distance. The two models are the same for all times after 10^{-30} second.

Prior to (and during) inflation, all the parts of the universe that we can now see were so small and close to each other that they *could* exchange information, that is, the horizon distance included all of the universe that we can now observe. Before (and during) inflation, there was adequate time for the observable universe to homogenize itself and come to the same temperature. Then, inflation expanded those regions tremendously, so that many parts of the universe are now beyond each other's horizon.

Another appeal of the inflationary model is its prediction that the density of the universe should be exactly equal to the critical density. To see why this is so, remember that curvature of spacetime is intimately linked to the density of matter. If the universe began with some curvature of its spacetime, one analogy for it might be the skin of a balloon. The period of inflation was equivalent to blowing up the balloon to a tremendous size. The universe became so big that from our vantage point, no curvature should be visible (Figure 29.26). In the same way, Earth's surface is so big that it looks flat to us no matter where we are. Calculations show that a universe with no curvature is one that is at critical density. Universes with densities either higher or lower than the critical density would show marked curvature. But we saw that the observations of the CMB in Figure 29.18, which show that the universe has critical density, rule out the possibility that space is significantly curved.

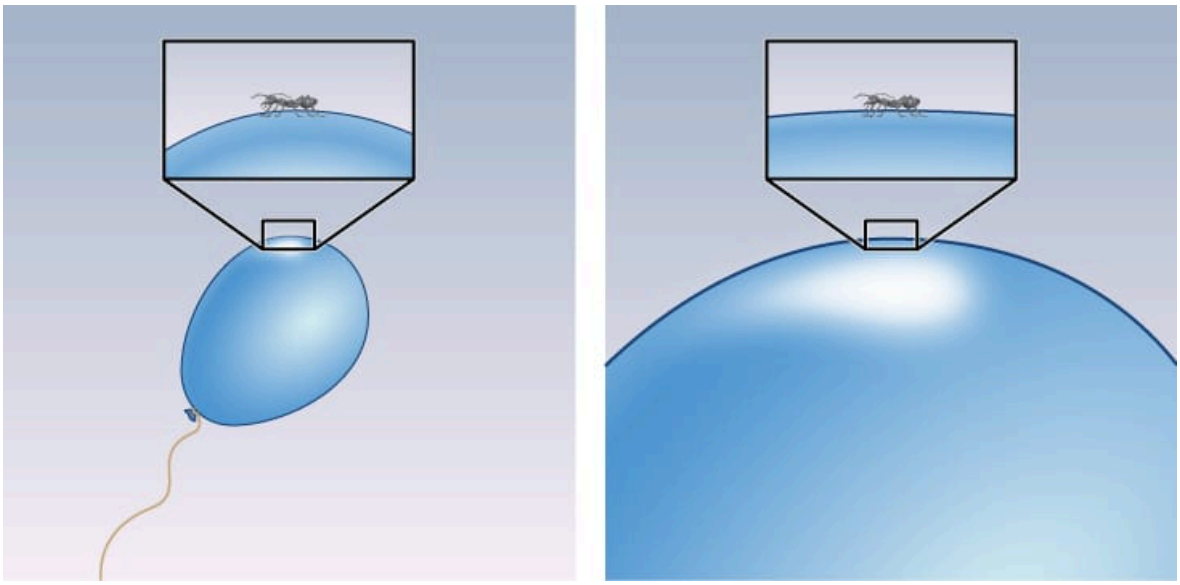


Figure 29.26 : Analogy for Inflation. During a period of rapid inflation, a curved balloon grows so large that to any local observer it looks flat. The inset shows the geometry from the ant's point of view.

Grand Unified Theories

While inflation is an intriguing idea and widely accepted by researchers, we cannot directly observe events so early in the universe. The conditions at the time of inflation were so extreme that we cannot reproduce them in our laboratories or high-energy accelerators, but scientists have some ideas about what the universe might have been like. These ideas are called grand unified theories or GUTs.

In GUT models, the forces that we are familiar with here on Earth, including gravity and electromagnetism, behaved very differently in the extreme conditions of the early universe than they do today. In physical science, the term *force* is used to describe anything that can change the motion of a particle or body. One of the remarkable discoveries of modern science is that all known physical processes can be described through the action of just four forces: gravity, electromagnetism, the strong nuclear force, and the weak nuclear force (Table 29.3).

Table 29.3: The Forces of Nature

Force	Relative Strength Today	Range of Action	Important Applications
Gravity	1	Whole universe	Motions of planets, stars, galaxies
Electromagnetism	10^{36}	Whole universe	Atoms, molecules, electricity, magnetic fields
Weak nuclear force	10^{33}	10^{-17} meters	Radioactive decay
Strong nuclear force	10^{38}	10^{-15} meters	The existence of atomic nuclei

Gravity is perhaps the most familiar force, and certainly appears strong if you jump off a tall building. However, the force of gravity between two elementary particles—say two protons—is by far the weakest of the four forces. Electromagnetism—which includes both magnetic and electrical forces, holds atoms together, and produces the electromagnetic radiation that we use to study the universe—is much stronger, as you can see in Table 29.3. The weak nuclear force is only weak in comparison to its strong "cousin," but it is in fact much stronger than gravity.

Both the weak and strong nuclear forces differ from the first two forces in that they act only over very small distances—those comparable to the size of an atomic nucleus or less. The weak force is involved in radioactive decay and in reactions that result in the production of neutrinos. The strong force holds protons and neutrons together in an atomic nucleus.

Physicists have wondered why there are four forces in the universe—why not 300 or, preferably, just one? An important hint comes from the name *electromagnetic force*. For a long time, scientists thought that the forces of electricity and magnetism were

separate, but James Clerk Maxwell (see the chapter on Radiation and Spectra) was able to *unify* these forces—to show that they are aspects of the same phenomenon. In the same way, many scientists (including Einstein) have wondered if the four forces we now know could also be unified. Physicists have actually developed GUTs that unify three of the four forces (but not gravity).

In these theories, the strong, weak, and electromagnetic forces are not three independent forces but instead are different manifestations or aspects of what is, in fact, a single force. The theories predict that at high enough temperatures, there would be only one force. At lower temperatures (like the ones in the universe today), however, this single force has changed into three different forces (Figure 29.27). Just as different gases or liquids freeze at different temperatures, we can say that the different forces "froze out" of the unified force at different temperatures. Unfortunately, the temperatures at which the three forces acted as one force are so high that they cannot be reached in any laboratory on Earth. Only the early universe, at times prior to 10^{-35} second, was hot enough to unify these forces.

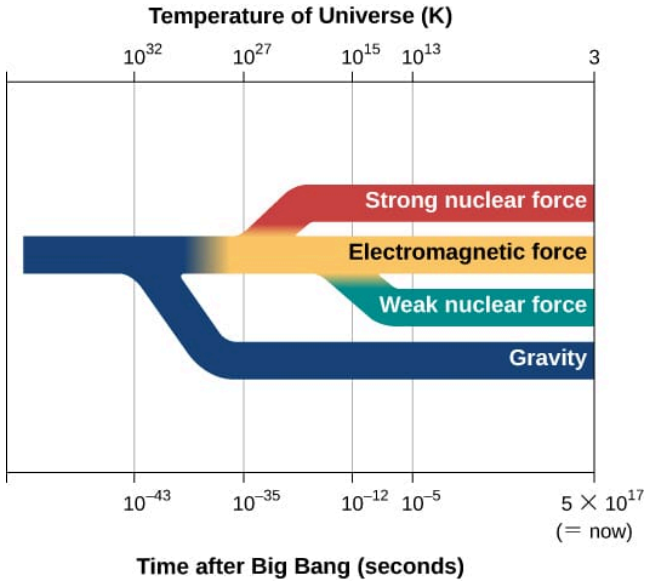


Figure 29.27 : Four Forces That Govern the Universe. The behavior of the four forces depends on the temperature of the universe. This diagram (inspired by some grand unified theories) shows that at very early times when the temperature of the universe was very high, all four forces resembled one another and were indistinguishable. As the universe cooled, the forces took on separate and distinctive characteristics.

Many physicists think that gravity was also unified with the three other forces at still higher temperatures, and scientists have tried to develop a theory that combines all four forces. For example, in string theory, the point-like particles of matter that we have discussed in this book are replaced by one-dimensional objects called strings. In this theory, infinitesimal strings, which have length but not height or width, are the building blocks used to construct all the forms of matter and energy in the universe. These strings exist in 11-dimensional space (not the 4-dimensional spacetime with which we are familiar). The strings vibrate in the various dimensions, and depending on how they vibrate, they are seen in our world as matter or gravity or light. As you can imagine, the mathematics of string theory is very complex, and the theory remains untested by experiments. Even the largest particle accelerators on Earth do not achieve high enough energy to show whether string theory applies to the real world.

String theory is interesting to scientists because it is currently the only approach that seems to have the potential of combining all four forces to produce what physicists have termed the Theory of Everything.[3] Theories of the earliest phases of the universe must take both quantum mechanics and gravity into account, but at the simplest level, gravity and quantum mechanics are incompatible. General relativity, our best theory of gravity, says that the motions of objects can be predicted exactly. Quantum mechanics says you can only calculate the probability (chance) that an object will do something. String theory is an attempt to resolve this paradox. The mathematics that underpins string theory is elegant and beautiful, but it remains to be seen whether it will make predictions that can be tested by observations in yet-to-be-developed, high-energy accelerators on Earth or by observations of the early universe.

The earliest period in the history of the universe from time zero to 10^{-43} second is called the Planck time. The universe was unimaginably hot and dense, and theorists believe that at this time, quantum effects of gravity dominated physical interactions—and, as we have just discussed, we have no tested theory of quantum gravity. Inflation is hypothesized to have occurred somewhat later, when the universe was between perhaps 10^{-35} and 10^{-33} second old and the temperature was 10^{27} to 10^{28} K. This rapid expansion took place when three forces (electromagnetic, strong, and weak) are thought to have been unified,

and this is when GUTs are applicable.

After inflation, the universe continued to expand (but more slowly) and to cool. An important milestone was reached when the temperature was down to 10^{15} K and the universe was 10^{-10} second old. Under these conditions, all four forces were separate and distinct. High-energy particle accelerators can achieve similar conditions, and so theories of the history of the universe from this point on have a sound basis in experiments.

As yet, we have no direct evidence of what the conditions were during the inflationary epoch, and the ideas presented here are speculative. Researchers are trying to devise some experimental tests. For example, the quantum fluctuations in the very early universe would have caused variations in density and produced gravitational waves that may have left a detectable imprint on the CMB. Detection of such an imprint will require observations with equipment whose sensitivity is improved from what we have today. Ultimately, however, it may provide confirmation that we live in a universe that once experienced an epoch of rapid inflation.

If you are typical of the students who read this book, you may have found this brief discussion of dark matter, inflation, and cosmology a bit frustrating. We have offered glimpses of theories and observations, but have raised more questions than we have answered. What is dark matter? What is dark energy? Inflation explains the observations of flatness and uniformity of the university, but did it actually happen? These ideas are at the forefront of modern science, where progress almost always leads to new puzzles, and much more work is needed before we can see clearly. Bear in mind that less than a century has passed since Hubble demonstrated the existence of other galaxies. The quest to understand just how the universe of galaxies came to be will keep astronomers busy for a long time to come.

Footnotes

- 3This name became the title of a film about physicist Stephen Hawking in 2014.

This page titled 17.3: The Inflationary Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 29.7: The Inflationary Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

18: DARK MATTER AND DARK ENERGY

18.1: The Challenge of Dark Matter
18.2: The Age of the Universe
18.3: A Model of the Universe
18.4: What Is the Universe Really Made Of?

This page titled 18: Dark Matter and Dark Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

18.1: The Challenge of Dark Matter

Learning Objectives

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

- Explain how astronomers know that the solar system contains very little dark matter
- Summarize the evidence for dark matter in most galaxies
- Explain how we know that galaxy clusters are dominated by dark matter
- Relate the presence of dark matter to the average mass-to-light ratio of huge volumes of space containing many galaxies

So far this chapter has focused almost entirely on matter that radiates electromagnetic energy—stars, planets, gas, and dust. But, as we have pointed out in several earlier chapters (especially The Milky Way Galaxy), it is now clear that galaxies contain large amounts of dark matter as well. There is much more dark matter, in fact, than matter we can see—which means it would be foolish to ignore the effect of this unseen material in our theories about the structure of the universe. (As many a ship captain in the polar seas found out too late, the part of the iceberg visible above the ocean's surface was not necessarily the only part he needed to pay attention to.) Dark matter turns out to be extremely important in determining the evolution of galaxies and of the universe as a whole.

The idea that much of the universe is filled with dark matter may seem like a bizarre concept, but we can cite a historical example of "dark matter" much closer to home. In the mid-nineteenth century, measurements showed that the planet Uranus did not follow exactly the orbit predicted from Newton's laws if one added up the gravitational forces of all the known objects in the solar system. Some people worried that Newton's laws may simply not work so far out in our solar system. But the more straightforward interpretation was to attribute Uranus' orbital deviations to the gravitational effects of a new planet that had not yet been seen. Calculations showed where that planet had to be, and Neptune was discovered just about in the predicted location.

In the same way, astronomers now routinely determine the location and amount of dark matter in galaxies by measuring its gravitational effects on objects we can see. And, by measuring the way that galaxies move in clusters, scientists have discovered that dark matter is also distributed among the galaxies in the clusters. Since the environment surrounding a galaxy is important in its development, dark matter must play a central role in galaxy evolution as well. Indeed, it appears that dark matter makes up most of the matter in the universe. But what *is* dark matter? What is it made of? We'll look next at the search for dark matter and the quest to determine its nature.

Dark Matter in the Local Neighborhood

Is there dark matter in our own solar system? Astronomers have examined the orbits of the known planets and of spacecraft as they journey to the outer planets and beyond. No deviations have been found from the orbits predicted on the basis of the masses of objects already discovered in our solar system and the theory of gravity. We therefore conclude that there is no evidence that there are large amounts of dark matter nearby.

Astronomers have also looked for evidence of dark matter in the region of the Milky Way Galaxy that lies within a few hundred light-years of the Sun. In this vicinity, most of the stars are restricted to a thin disk. It is possible to calculate how much mass the disk must contain in order to keep the stars from wandering far above or below it. The total matter that must be in the disk is less than twice the amount of luminous matter. This means that no more than half of the mass in the region near the Sun can be dark matter.

Dark Matter in and around Galaxies

In contrast to our local neighborhood near the Sun and solar system, there is (as we saw in The Milky Way Galaxy) ample evidence strongly suggesting that about 90% of the mass in the entire galaxy is in the form of a halo of dark matter. In other words, there is apparently about nine times more dark matter than visible matter. Astronomers have found some stars in the outer regions of the Milky Way beyond its bright disk, and these stars are revolving very rapidly around its center. The mass contained in all the stars and all the interstellar matter we can detect in the galaxy does not exert enough gravitational force to explain how those

fast-moving stars remain in their orbits and do not fly away. Only by having large amounts of unseen matter could the galaxy be holding on to those fast-moving outer stars. The same result is found for other spiral galaxies as well.

Figure 28.23 is an example of the kinds of observations astronomers are making, for the Triangulum galaxy, a member of our Local Group. The observed rotation of spiral galaxies like Andromeda is usually seen in plots, known as *rotation curves,* that show velocity versus distance from the galaxy center. Such plots suggest that the dark matter is found in a large halo surrounding the luminous parts of each galaxy. The radius of the halos around the Milky Way and Andromeda may be as large as 300,000 light-years, much larger than the visible size of these galaxies.

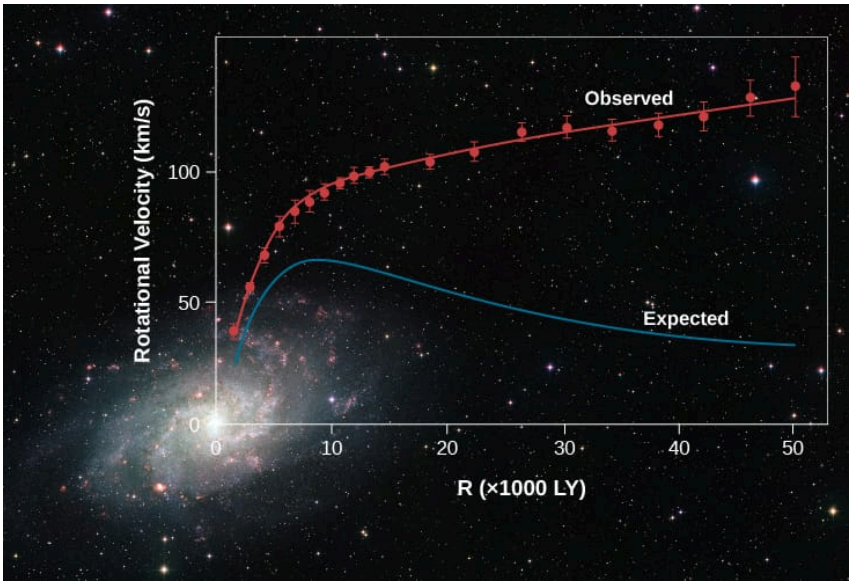


Figure 28.23 : Rotation Indicates Dark Matter. We see the Milky Way's neighbor, the Triangulum galaxy, with a graph that shows the velocity at which stars and clouds of gas orbit the galaxy at different distances from the center (red line). As is true of the Milky Way, the rotational velocity (or orbital speed) does not decrease with distance from the center, which is what you would expect if an assembly of objects rotates around a common center. A calculation (blue line) based on the total mass visible as stars, gas, and dust predicts that the velocity should be much lower at larger distances from the center. The discrepancy between the two curves implies the presence of a halo of massive dark matter extending outside the boundary of the luminous matter. This dark matter causes everything in the galaxy to orbit faster than the observed matter alone could explain. (credit background: modification of work by ESO)

Dark Matter in Clusters of Galaxies

Galaxies in clusters also move around: they orbit the cluster's center of mass. It is not possible for us to follow a galaxy around its entire orbit because that typically takes about a billion years. It is possible, however, to measure the velocities with which galaxies in a cluster are moving, and then estimate what the total mass in the cluster must be to keep the individual galaxies from flying out of the cluster. The observations indicate that the mass of the galaxies alone cannot keep the cluster together—some other gravity must again be present. The total amount of dark matter in clusters exceeds by more than ten times the luminous mass contained within the galaxies themselves, indicating that dark matter exists between galaxies as well as inside them.

There is another approach we can take to measuring the amount of dark matter in clusters of galaxies. As we saw, the universe is expanding, but this expansion is not perfectly uniform, thanks to the interfering hand of gravity. Suppose, for example, that a galaxy lies outside but relatively close to a rich cluster of galaxies. The gravitational force of the cluster will tug on that neighboring galaxy and slow down the rate at which it moves away from the cluster due to the expansion of the universe.

Consider the Local Group of galaxies, lying on the outskirts of the Virgo Supercluster. The mass concentrated at the center of the Virgo Cluster exerts a gravitational force on the Local Group. As a result, the Local Group is moving away from the center of the Virgo Cluster at a velocity a few hundred kilometers per second slower than the Hubble law predicts. By measuring such deviations from a smooth expansion, astronomers can estimate the total amount of mass contained in large clusters.

There are two other very useful methods for measuring the amount of dark matter in galaxy clusters, and both of them have produced results in general agreement with the method of measuring galaxy velocities: gravitational lensing and X-ray emission. Let's take a look at both.

As Albert Einstein showed in his theory of general relativity, the presence of mass bends the surrounding fabric of spacetime. Light follows those bends, so very massive objects can bend light significantly. You saw examples of this in the Astronomy Basics feature box Gravitational Lensing in the previous section. Visible galaxies are not the only possible gravitational lenses. Dark matter can also reveal its presence by producing this effect. Figure 28.24 shows a galaxy cluster that is acting like a gravitational lens; the streaks and arcs you see on the picture are lensed images of more distant galaxies. Gravitational lensing is well enough understood that astronomers can use the many ovals and arcs seen in this image to calculate detailed maps of how much matter there is in the cluster and how that mass is distributed. The result from studies of many such gravitational lens clusters shows that, like individual galaxies, galaxy clusters contain more than ten times as much dark matter as luminous matter.



Figure 28.24 : Cluster Abell 2218. This view from the Hubble Space Telescope shows the massive galaxy cluster Abell 2218 at a distance of about 2 billion light-years. Most of the yellowish objects are galaxies belonging to the cluster. But notice the numerous long, thin streaks, many of them blue; those are the distorted and magnified images of even more distant background galaxies, gravitationally lensed by the enormous mass of the intervening cluster. By carefully analyzing the lensed images, astronomers can construct a map of the dark matter that dominates the mass of the cluster. (credit: modification of work by NASA, ESA, and Johan Richard (Caltech))

The third method astronomers use to detect and measure dark matter in galaxy clusters is to image them in the light of X-rays. When the first sensitive X-ray telescopes were launched into orbit around Earth in the 1970s and trained on massive galaxy clusters, it was quickly discovered that the clusters emit copious X-ray radiation (see Figure 28.25). Most stars do not emit much X-ray radiation, and neither does most of the gas or dust between the stars inside galaxies. What could be emitting the X-rays seen from virtually all massive galaxy clusters?

It turns out that just as galaxies have gas distributed between their stars, clusters of galaxies have gas distributed between their galaxies. The particles in these huge reservoirs of gas are not just sitting still; rather, they are constantly moving, zooming around under the influence of the cluster's immense gravity like mini planets around a giant sun. As they move and bump against each other, the gas heats up hotter and hotter until, at temperatures as high as 100 million K, it shines brightly at X-ray wavelengths. The more mass the cluster has, the faster the motions, the hotter the gas, and the brighter the X-rays. Astronomers calculate that the mass present to induce those motions must be about ten times the mass they can see in the clusters, including all the galaxies and all the gas. Once again, this is evidence that the galaxy clusters are seen to be dominated by dark matter.



Figure 28.25 : X-Ray Image of a Galaxy Cluster. This composite image shows the galaxy cluster Abell 1689 at a distance of 2.3 billion light-years. The finely detailed views of the galaxies, most of them yellow, are in visible and near-infrared light from the Hubble Space Telescope, while the diffuse purple haze shows X-rays as seen by Chandra X-ray Observatory. The abundant X-rays, the gravitationally lensed images (thin curving arcs) of background galaxies, and the measured velocities of galaxies in the clusters all show that the total mass of Abell 1689—most of it dark matter—is about 10^{15} solar masses. (credit: modification of work by NASA/ESA/JPL-Caltech/Yale/CNRS)

Mass-to-Light Ratio

We described the use of the mass-to-light ratio to characterize the matter in galaxies or clusters of galaxies in Properties of Galaxies. For systems containing mostly old stars, the mass-to-light ratio is typically 10 to 20, where mass and light are measured in units of the Sun's mass and luminosity. A mass-to-light ratio of 100 or more is a signal that a substantial amount of dark matter is present. Table 28.1 summarizes the results of measurements of mass-to-light ratios for various classes of objects. Very large mass-to-light ratios are found for all systems of galaxy size and larger, indicating that dark matter is present in all these types of objects. This is why we say that dark matter apparently makes up most of the total mass of the universe.

Table 28.1: Mass-To-Light Ratios

Type of Object	Mass-to-Light Ratio
Sun	1
Matter in vicinity of Sun	2
Mass in Milky Way within 80,000 light-years of the center	10
Small groups of galaxies	50–150
Rich clusters of galaxies	250–300

The clustering of galaxies can be used to derive the total amount of mass in a given region of space, while visible radiation is a good indicator of where the luminous mass is. Studies show that the dark matter and luminous matter are very closely associated. The dark matter halos do extend beyond the luminous boundaries of the galaxies that they surround. However, where there are large clusters of galaxies, you will also find large amounts of dark matter. Voids in the galaxy distribution are also voids in the distribution of dark matter.

What Is the Dark Matter?

How do we go about figuring out what the dark matter consists of? The technique we might use depends on its composition. Let's consider the possibility that some of the dark matter is made up of normal particles: protons, neutrons, and electrons. Suppose these particles were assembled into black holes, brown dwarfs, or white dwarfs. If the black holes had no accretion disks, they

would be invisible to us. White and brown dwarfs do emit some radiation but have such low luminosities that they cannot be seen at distances greater than a few thousand light-years.

We can, however, look for such compact objects because they can act as gravitational lenses. (See the Astronomy Basics feature box Gravitational Lensing.) Suppose the dark matter in the halo of the Milky Way were made up of black holes, brown dwarfs, and white dwarfs. These objects have been whimsically dubbed MACHOs (MAssive Compact Halo Objects). If an invisible MACHO passes directly between a distant star and Earth, it acts as a gravitational lens, focusing the light from the distant star. This causes the star to appear to brighten over a time interval of a few hours to several days before returning to its normal brightness. Since we can't predict when any given star might brighten this way, we have to monitor huge numbers of stars to catch one in the act. There are not enough astronomers to keep monitoring so many stars, but today's automated telescopes and computer systems can do it for us.

Research teams making observations of millions of stars in the nearby galaxy called the Large Magellanic Cloud have reported several examples of the type of brightening expected if MACHOs are present in the halo of the Milky Way (Figure 28.26). However, there are not enough MACHOs in the halo of the Milky Way to account for the mass of the dark matter in the halo.



Figure 28.26 : Large and Small Magellanic Clouds. Here, the two small galaxies we call the Large Magellanic Cloud and Small Magellanic Cloud can be seen above the auxiliary telescopes for the Very Large Telescope Array on Cerro Paranal in Chile. You can see from the number of stars that are visible that this is a very dark site for doing astronomy. (credit: ESO/J. Colosimo)

This result, along with a variety of other experiments, leads us to conclude that the types of matter we are familiar with can make up only a tiny portion of the dark matter. Another possibility is that dark matter is composed of some new type of particle—one that researchers are now trying to detect in laboratories here on Earth (see The Big Bang).

The kinds of dark matter particles that astronomers and physicists have proposed generally fall into two main categories: hot and cold dark matter. The terms *hot* and *cold* don't refer to true temperatures, but rather to the average velocities of the particles, analogous to how we might think of particles of air moving in your room right now. In a cold room, the air particles move more slowly on average than in a warm room.

In the early universe, if dark matter particles easily moved fast and far compared to the lumps and bumps of ordinary matter that eventually became galaxies and larger structures, we call those particles hot dark matter. In that case, smaller lumps and bumps would be smeared out by the particle motions, meaning fewer small galaxies would get made.

On the other hand, if the dark matter particles moved slowly and covered only small distances compared to the sizes of the lumps in the early universe, we call that cold dark matter. Their slow speeds and energy would mean that even the smaller lumps of ordinary matter would survive to grow into small galaxies. By looking at when galaxies formed and how they evolve, we can use observations to distinguish between the two kinds of dark matter. So far, observations seem most consistent with models based on cold dark matter.

Solving the dark matter problem is one of the biggest challenges facing astronomers. After all, we can hardly understand the

evolution of galaxies and the long-term history of the universe without understanding what its most massive component is made of. For example, we need to know just what role dark matter played in starting the higher-density “seeds” that led to the formation of galaxies. And since many galaxies have large halos made of dark matter, how does this affect their interactions with one another and the shapes and types of galaxies that their collisions create?

Astronomers armed with various theories are working hard to produce models of galaxy structure and evolution that take dark matter into account in just the right way. Even though we don’t know what the dark matter is, we do have some clues about how it affected the formation of the very first galaxies. As we will see in The Big Bang, careful measurements of the microwave radiation left over after the Big Bang have allowed astronomers to set very tight limits on the actual sizes of those early seeds that led to the formation of the large galaxies that we see in today’s universe. Astronomers have also measured the relative numbers and distances between galaxies and clusters of different sizes in the universe today. So far, most of the evidence seems to weigh heavily in favor of cold dark matter, and most current models of galaxy and large-scale structure formation use cold dark matter as their main ingredient.

As if the presence of dark matter—a mysterious substance that exerts gravity and outweighs all the known stars and galaxies in the universe but does not emit or absorb light—were not enough, there is an even more baffling and equally important constituent of the universe that has only recently been discovered: we have called it dark energy in parallel with dark matter. We will say more about it and explore its effects on the evolution of the universe in The Big Bang. For now, we can complete our inventory of the contents of the universe by noting that it appears that the entire universe contains some mysterious energy that pushes spacetime apart, taking galaxies and the larger structures made of galaxies along with it. Observations show that dark energy becomes more and more important relative to gravity as the universe ages. As a result, the expansion of the universe is accelerating, and this acceleration seems to be happening mostly since the universe was about half its current age.

What we see when we peer out into the universe—the light from trillions of stars in hundreds of billions of galaxies wrapped in intricate veils of gas and dust—is therefore actually only a sprinkling of icing on top of the cake: as we will see in The Big Bang, when we look outside galaxies and clusters of galaxies at the universe as a whole, astronomers find that for every gram of luminous normal matter, such as protons, neutrons, electrons, and atoms in the universe, there are about 4 grams of nonluminous normal matter, mainly intergalactic hydrogen and helium. There are about 27 grams of dark matter, and the energy equivalent (remember Einstein’s famous $E = mc^2$) of about 68 grams of dark energy. Dark matter, and (as we will see) even more so dark energy, are dramatic demonstrations of what we have tried to emphasize throughout this book: science is always a “progress report,” and we often encounter areas where we have more questions than answers.

Let’s next put together all these clues to trace the life history of galaxies and large-scale structure in the universe. What follows is the current consensus, but research in this field is moving rapidly, and some of these ideas will probably be modified as new observations are made.

This page titled 18.1: The Challenge of Dark Matter is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax.

- 28.5: The Challenge of Dark Matter by OpenStax is licensed CC BY 4.0.

18.2: The Age of the Universe

Learning Objectives

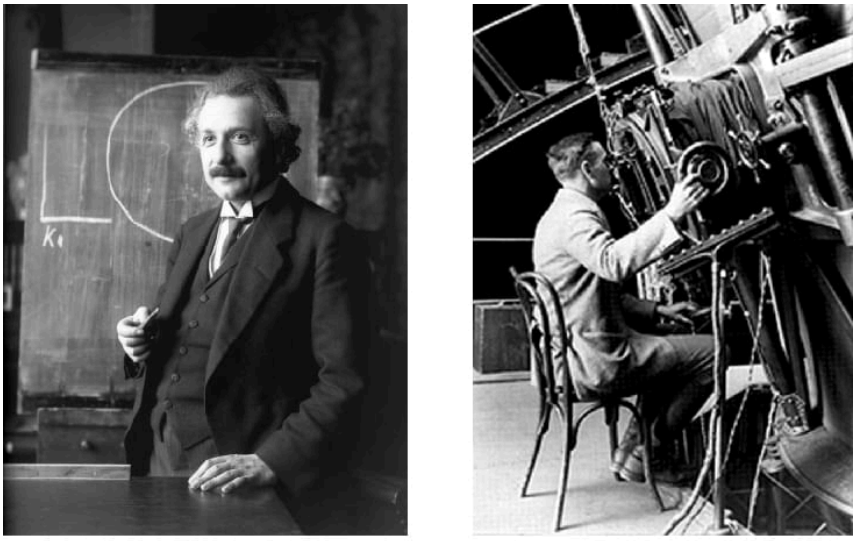
By the end of this section, you will be able to:

- Describe how we estimate the age of the universe
- Explain how changes in the rate of expansion over time affect estimates of the age of the universe
- Describe the evidence that dark energy exists and that the rate of expansion is currently accelerating
- Describe some independent evidence for the age of the universe that is consistent with the age estimate based on the rate of expansion

To explore the history of the universe, we will follow the same path that astronomers followed historically—beginning with studies of the nearby universe and then probing ever-more-distant objects and looking further back in time.

The realization that the universe changes with time came in the 1920s and 1930s when measurements of the redshifts of a large sample of galaxies became available. With hindsight, it is surprising that scientists were so shocked to discover that the universe is expanding. In fact, our theories of gravity demand that the universe must be either expanding or contracting. To show what we mean, let's begin with a universe of finite size—say a giant ball of a thousand galaxies. All these galaxies attract each other because of their gravity. If they were initially stationary, they would inevitably begin to move closer together and eventually collide. They could avoid this collapse only if for some reason they happened to be moving away from each other at high speeds. In just the same way, only if a rocket is launched at high enough speed can it avoid falling back to Earth.

The problem of what happens in an infinite universe is harder to solve, but Einstein (and others) used his theory of general relativity (which we described in Black Holes and Curved Spacetime) to show that even infinite universes cannot be static. Since astronomers at that time did not yet know the universe was expanding (and Einstein himself was philosophically unwilling to accept a universe in motion), he changed his equations by introducing an arbitrary new term (we might call it a fudge factor) called the cosmological constant. This constant represented a hypothetical force of repulsion that could balance gravitational attraction on the largest scales and permit galaxies to remain at fixed distances from one another. That way, the universe could remain still.



(a) (b)

Figure 29.2 : Einstein and Hubble. (a) Albert Einstein is shown in a 1921 photograph. (b) Edwin Hubble at work in the Mt. Wilson Observatory.

About a decade later, Hubble, and his coworkers reported that the universe is expanding, so that no mysterious balancing force is needed. (We discussed this in the chapter on Galaxies.) Einstein is reported to have said that the introduction of the cosmological constant was "the biggest blunder of my life." As we shall see later in this chapter, however, relatively recent observations indicate

that the expansion is *accelerating*. Observations are now being carried out to determine whether this acceleration is consistent with a cosmological constant. In a way, it may turn out that Einstein was right after all.

Link to Learning

View this web exhibit on the history of our thinking about cosmology, with images and biographies, from the American Institute of Physics Center for the History of Physics.

The Hubble Time

If we had a movie of the expanding universe and ran the film *backward*, what would we see? The galaxies, instead of moving apart, would move *together* in our movie—getting closer and closer all the time. Eventually, we would find that all the matter we can see today was once concentrated in an infinitesimally small volume. Astronomers identify this time with the *beginning of the universe*. The explosion of that concentrated universe at the beginning of time is called the Big Bang (not a bad term, since you can't have a bigger bang than one that creates the entire universe). But when did this bang occur?

We can make a reasonable estimate of the time since the universal expansion began. To see how astronomers do this, let's begin with an analogy. Suppose your astronomy class decides to have a party (a kind of "Big Bang") at someone's home to celebrate the end of the semester. Unfortunately, everyone is celebrating with so much enthusiasm that the neighbors call the police, who arrive and send everyone away at the same moment. You get home at 2 a.m., still somewhat upset about the way the party ended, and realize you forgot to look at your watch to see what time the police got there. But you use a map to measure that the distance between the party and your house is 40 kilometers. And you also remember that you drove the whole trip at a steady speed of 80 kilometers/hour (since you were worried about the police cars following you). Therefore, the trip must have taken:

$$\text{time} = \frac{\text{distance}}{\text{velocity}} = \blacksquare = 0.5\ \text{hours}$$

So the party must have broken up at 1:30 a.m.

No humans were around to look at their watches when the universe began, but we can use the same technique to estimate when the galaxies began moving away from each other. (Remember that, in reality, it is space that is expanding, not the galaxies that are moving through static space.) If we can measure how far apart the galaxies are now, and how fast they are moving, we can figure out how long a trip it's been.

Let's call the age of the universe measured in this way T_0. Let's first do a simple case by assuming that the expansion has been at a constant rate ever since the expansion of the universe began. In this case, the time it has taken a galaxy to move a distance, *d*, away from the Milky Way (remember that at the beginning the galaxies were all together in a very tiny volume) is (as in our example)

$$T_0 = d/$$

where *v* is the velocity of the galaxy. If we can measure the speed with which galaxies are moving away, and also the distances between them, we can establish how long ago the expansion began.

Making such measurements should sound very familiar. This is just what Hubble and many astronomers after him needed to do in order to establish the Hubble law and the Hubble constant. We learned in Galaxies that a galaxy's distance and its velocity in the expanding universe are related by

$$V =$$

where *H* is the Hubble constant. Combining these two expressions gives us

$$T_0 = \frac{d}{v} = \frac{d}{(H \times d)} =$$

We see, then, that the work of calculating this time was already done for us when astronomers measured the Hubble constant. The age of the universe estimated in this way turns out to be just the *reciprocal of the Hubble constant* (that is, 1/*H*). This age estimate is sometimes called the Hubble time. For a Hubble constant of 20 kilometers/second per million light-years, the Hubble

time is about 15 billion years. (By the way, the unit used by astronomers for the Hubble constant is kilometers/second per million parsecs. In these units, the Hubble constant is equal to about 70 kilometers/second per million parsecs, again with an uncertainty of about 5%.)

To make numbers easier to remember, we have done some rounding here. Estimates for the Hubble constant are actually closer to 21 or 22 kilometers/second per million light-years, which would make the age closer to 14 billion years. But there is still about a 5% uncertainty in the Hubble constant, which means the age of the universe estimated in this way is also uncertain by about 5%.

To put these uncertainties in perspective, however, you should know that 50 years ago, the uncertainty was a factor of 2. Remarkable progress toward pinning down the Hubble constant has been made in the last couple of decades.

The Role of Deceleration

The Hubble time is the right age for the universe only if the expansion rate has been constant throughout the time since the expansion of the universe began. Continuing with our end-of-the-semester-party analogy, this is equivalent to assuming that you traveled home from the party at a constant rate, when in fact this may not have been the case. At first, mad about having to leave, you may have driven fast, but then as you calmed down—and thought about police cars on the highway—you may have begun to slow down until you were driving at a more socially acceptable speed (such as 80 kilometers/hour). In this case, given that you were driving faster at the beginning, the trip home would have taken less than a half-hour.

In the same way, in calculating the Hubble time, we have assumed that the expansion rate has been constant throughout all of time. It turns out that this is not a good assumption. Earlier in their thinking about this, astronomers expected that the rate of expansion should be slowing down. We know that matter creates gravity, whereby all objects pull on all other objects. The mutual attraction between galaxies was expected to slow the expansion as time passed. This means that, if gravity were the only force acting (a big *if*, as we shall see in the next section), then the rate of expansion must have been faster in the past than it is today. In this case, we would say the universe has been *decelerating* since the beginning.

How much it has decelerated depends on the importance of gravity in slowing the expansion. If the universe were nearly empty, the role of gravity would be minor. Then the deceleration would be close to zero, and the universe would have been expanding at a constant rate. But in a universe with any significant density of matter, the pull of gravity means that the rate of expansion should be slower now than it used to be. If we use the current rate of expansion to estimate how long it took the galaxies to reach their current separations, we will overestimate the age of the universe—just as we may have overestimated the time it took for you to get home from the party.

A Universal Acceleration

Astronomers spent several decades looking for evidence that the expansion was decelerating, but they were not successful. What they needed were 1) larger telescopes so that they could measure the redshifts of more distant galaxies and 2) a very luminous *standard bulb* (or standard candle), that is, some astronomical object with known luminosity that produces an enormous amount of energy and can be observed at distances of a billion light-years or more.

Recall that we discussed standard bulbs in the chapter on Galaxies. If we compare how luminous a standard bulb is supposed to be and how dim it actually looks in our telescopes, the difference allows us to calculate its distance. The redshift of the galaxy such a bulb is in can tell us how fast it is moving in the universe. So we can measure its distance and motion independently.

These two requirements were finally met in the 1990s. Astronomers showed that supernovae of type Ia (see The Death of Stars), with some corrections based on the shapes of their light curves, are standard bulbs. This type of supernova occurs when a white dwarf accretes enough material from a companion star to exceed the Chandrasekhar limit and then collapses and explodes. At the time of maximum brightness, these dramatic supernovae can briefly outshine the galaxies that host them, and hence, they can be observed at very large distances. Large 8- to 10-meter telescopes can be used to obtain the spectra needed to measure the redshifts of the host galaxies (Figure 29.3).

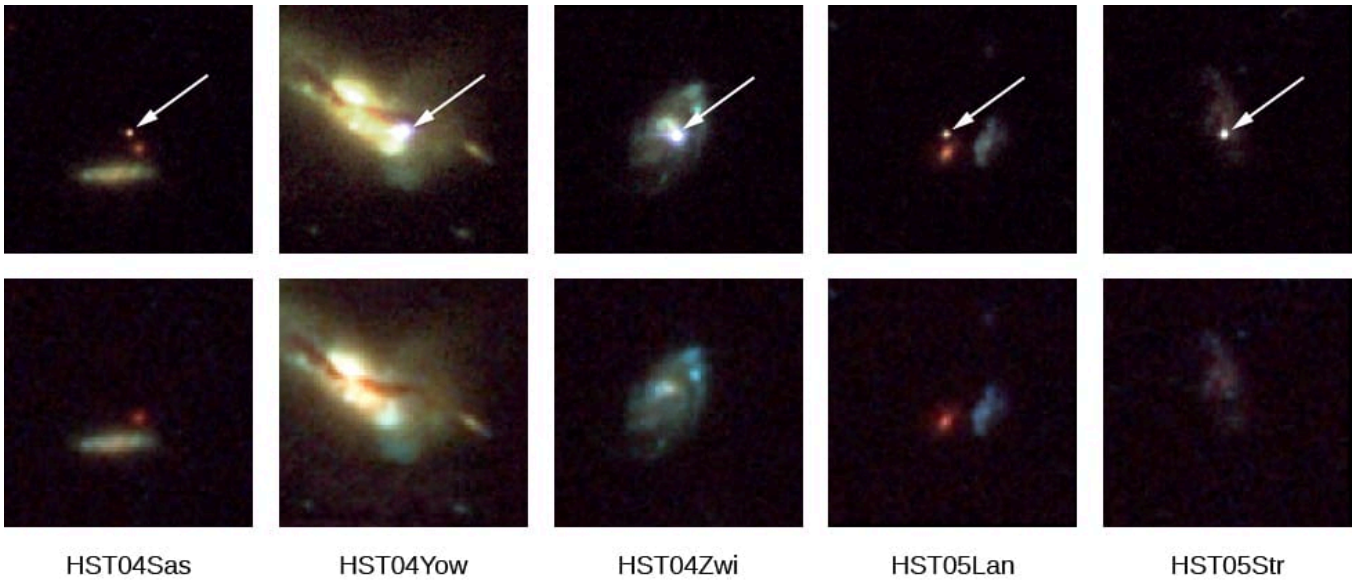


Figure 29.3 : Five Supernovae and Their Host Galaxies. The top row shows each galaxy and its supernova (arrow). The bottom row shows the same galaxies either before or after the supernovae exploded. (credit: modification of work by NASA, ESA, and A. Riess (STScI))

The result of painstaking, careful study of these supernovae in a range of galaxies, carried out by two groups of researchers, was published in 1998. It was shocking—and so revolutionary that their discovery received the 2011 Nobel Prize in Physics. What the researchers found was that these type Ia supernovae in distant galaxies were fainter than expected from Hubble's law, given the measured redshifts of their host galaxies. In other words, distances estimated from the supernovae used as standard bulbs disagreed with the distances measured from the redshifts.

If the universe were decelerating, we would expect the far-away supernovae to be *brighter* than expected. The slowing down would have kept them closer to us. Instead, they were *fainter*, which at first seemed to make no sense.

Before accepting this shocking development, astronomers first explored the possibility that the supernovae might not really be as useful as standard bulbs as they thought. Perhaps the supernovae appeared too faint because dust along our line of sight to them absorbed some of their light. Or perhaps the supernovae at large distances were for some reason intrinsically less luminous than nearby supernovae of type Ia.

A host of more detailed observations ruled out these possibilities. Scientists then had to consider the alternative that the distance estimated from the redshift was incorrect. Distances derived from redshifts assume that the Hubble constant has been truly constant for all time. We saw that one way it might not be constant is that the expansion is slowing down. But suppose neither assumption is right (steady speed or slowing down.)

Suppose, instead, that the universe is *accelerating*. If the universe is expanding faster now than it was billions of years ago, our motion away from the distant supernovae has sped up since the explosion occurred, sweeping us farther away from them. The light of the explosion has to travel a greater distance to reach us than if the expansion rate were constant. The farther the light travels, the fainter it appears. This conclusion would explain the supernova observations in a natural way, and this has now been substantiated by many additional observations over the last couple of decades. It really seems that *the expansion of the universe is accelerating*, a notion so unexpected that astronomers at first resisted considering it.

How can the expansion of the universe be speeding up? If you want to accelerate your car, you must supply energy by stepping on the gas. Similarly, energy must be supplied to accelerate the expansion of the universe. The discovery of the acceleration was shocking because scientists still have no idea what the source of the energy is. Scientists call whatever it is dark energy, which is a clear sign of how little we understand it.

Note that this new component of the universe is not the dark matter we talked about in earlier chapters. Dark energy is something else that we have also not yet detected in our laboratories on Earth.

What is dark energy? One possibility is that it is the cosmological constant, which is an energy associated with the vacuum of "empty" space itself. Quantum mechanics (the intriguing theory of how things behave at the atomic and subatomic levels) tells us that the source of this vacuum energy might be tiny elementary particles that flicker in and out of existence everywhere throughout

the universe. Various attempts have been made to calculate how big the effects of this vacuum energy should be, but so far these attempts have been unsuccessful. In fact, the order of magnitude of theoretical estimates of the vacuum energy based on the quantum mechanics of matter and the value required to account for the acceleration of the expansion of the universe differ by an incredible factor of at least 10^{120} (that is a 1 followed by 120 zeros)! Various other theories have been suggested, but the bottom line is that, although there is compelling evidence that dark energy exists, we do not yet know the source of that energy.

Whatever the dark energy turns out to be, we should note that the discovery that the rate of expansion has not been constant since the beginning of the universe complicates the calculation of the age of the universe. Interestingly, the acceleration seems not to have started with the Big Bang. During the first several billion years after the Big Bang, when galaxies were close together, gravity was strong enough to slow the expansion. As galaxies moved farther apart, the effect of gravity weakened. Several billion years after the Big Bang, dark energy took over, and the expansion began to accelerate (Figure 29.4).

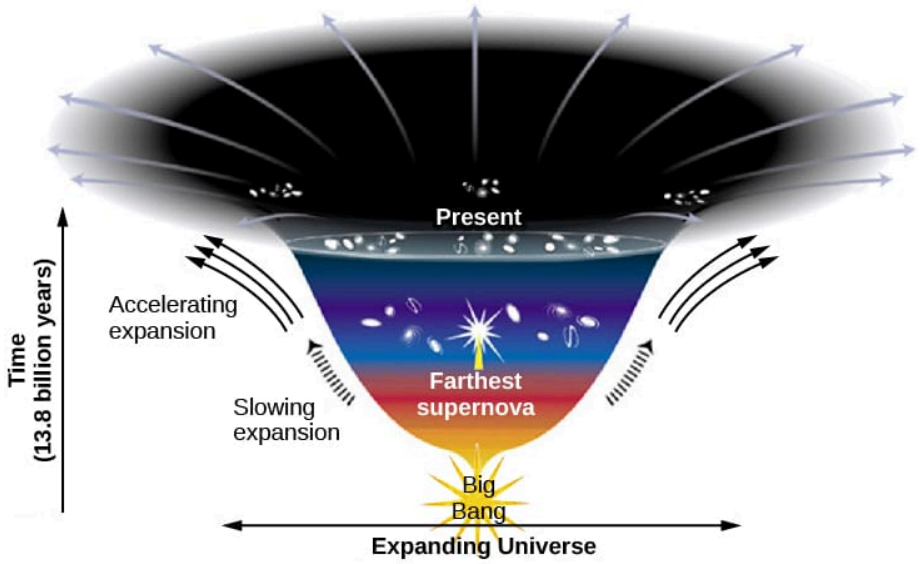


Figure 29.4 : Changes in the Rate of Expansion of the Universe Since Its Beginning 13.8 Billion Years Ago. The more the diagram spreads out horizontally, the faster the change in the velocity of expansion. After a period of very rapid expansion at the beginning, which scientists call inflation and which we will discuss later in this chapter, the expansion began to decelerate. Galaxies were then close together, and their mutual gravitational attraction slowed the expansion. After a few billion years, when galaxies were farther apart, the influence of gravity began to weaken. Dark energy then took over and caused the expansion to accelerate. (credit: modification of work by Ann Feild (STScI))

Deceleration works to make the age of the universe estimated by the simple relation $T_0 = 1/H$ seem older than it really is, whereas acceleration works to make it seem younger. By happy coincidence, our best estimates of how much deceleration and acceleration occurred lead to an answer for the age very close to $T_0 = 1/H$. The best current estimate is that the universe is 13.8 billion years old with an uncertainty of only about 100 million years.

Throughout this chapter, we have referred to the Hubble *constant*. We now know that the Hubble constant does change with time. It is, however, constant everywhere in the universe at any given time. When we say the Hubble constant is about 70 kilometers/second/million parsecs, we mean that this is the value of the Hubble constant at the current time.

Comparing Ages

We now have one estimate for the age of the universe from its expansion. Is this estimate consistent with other observations? For example, are the oldest stars or other astronomical objects younger than 13.8 billion years? After all, the universe has to be at least as old as the oldest objects in it.

In our Galaxy and others, the oldest stars are found in the globular clusters (Figure 29.5), which can be dated using the models of stellar evolution described in the chapter Stars from Adolescence to Old Age.

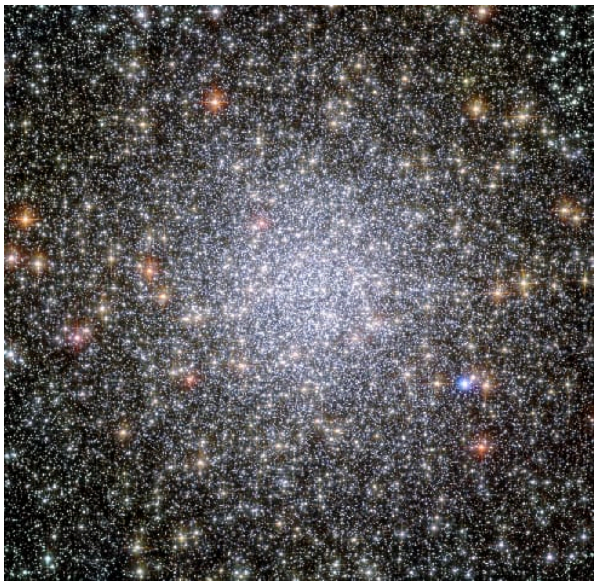


Figure 29.5 : Globular Cluster 47 Tucanae. This NASA/ESA Hubble Space Telescope image shows a globular cluster known as 47 Tucanae, since it is in the constellation of Tucana (The Toucan) in the southern sky. The second-brightest globular cluster in the night sky, it includes hundreds of thousands of stars. Globular clusters are among the oldest objects in our Galaxy and can be used to estimate its age. (credit: NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration)

The accuracy of the age estimates of the globular clusters has improved markedly in recent years for two reasons. First, models of interiors of globular cluster stars have been improved, mainly through better information about how atoms absorb radiation as they make their way from the center of a star out into space. Second, observations from satellites have improved the accuracy of our measurements of the distances to these clusters. The conclusion is that the oldest stars formed about 12–13 billion years ago.

This age estimate has recently been confirmed by the study of the spectrum of uranium in the stars. The isotope uranium-238 is radioactive and decays (changes into another element) over time. (Uranium-238 gets its designation because it has 92 protons and 146 neutrons.) We know (from how stars and supernovae make elements) how much uranium-238 is generally made compared to other elements. Suppose we measure the amount of uranium relative to nonradioactive elements in a very old star and in our own Sun, and compare the abundances. With those pieces of information, we can estimate how much longer the uranium has been decaying in the very old star because we know from our own Sun how much uranium decays in 4.5 billion years.

The line of uranium is very weak and hard to make out even in the Sun, but it has now been measured in one extremely old star using the European Very Large Telescope (Figure 29.6). Comparing the abundance with that in the solar system, whose age we know, astronomers estimate the star is 12.5 billion years old, with an uncertainty of about 3 billion years. While the uncertainty is large, this work is important confirmation of the ages estimated by studies of the globular cluster stars. Note that the uranium age estimate is completely independent; it does not depend on either the measurement of distances or on models of the interiors of stars.

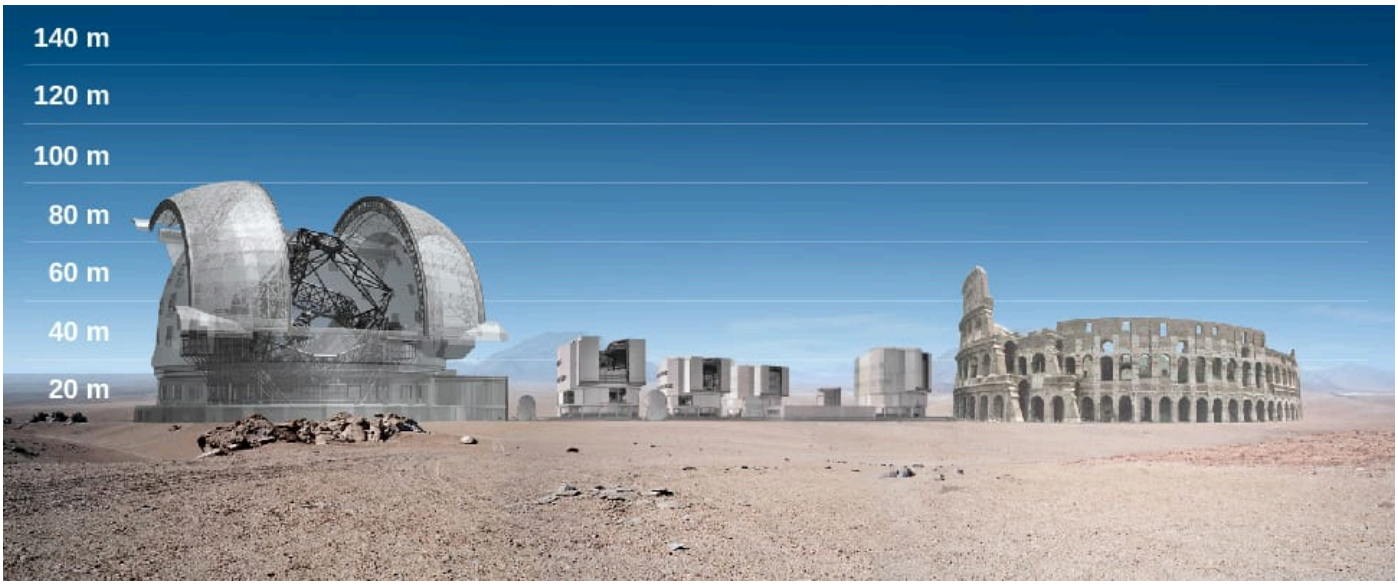


Figure 29.6 : European Extremely Large Telescope, European Very Large Telescope, and the Colosseum. The European Extremely Large Telescope (E-ELT) is currently under construction in Chile. This image compares the size of the E-ELT (left) with the four 8-meter telescopes of the European Very Large Telescope (center) and with the Colosseum in Rome (right). The mirror of the E-ELT will be 39 meters in diameter. Astronomers are building a new generation of giant telescopes in order to observe very distant galaxies and understand what they were like when they were newly formed and the universe was young. (credit: modification of work by ESO)

As we shall see later in this chapter, the globular cluster stars probably did not form until the expansion of the universe had been underway for at least a few hundred million years. Accordingly, their ages are consistent with the 13.8 billion-year age estimated from the expansion rate.

This page titled 18.2: The Age of the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 29.2: The Age of the Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

18.3: A Model of the Universe

Learning Objectives

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

- Explain how the rate of expansion of the universe affects its evolution
- Describe four possibilities for the evolution of the universe
- Explain what is expanding when we say that the universe is expanding
- Define critical density and the evidence that matter alone in the universe is much smaller than the critical density
- Describe what the observations say about the likely long-term future of the universe

Let's now use the results about the expansion of the universe to look at how these ideas might be applied to develop a model for the evolution of the universe as a whole. With this model, astronomers can make predictions about how the universe has evolved so far and what will happen to it in the future.

The Expanding Universe

Every model of the universe must include the expansion we observe. Another key element of the models is that the cosmological principle (which we discussed in The Evolution and Distribution of Galaxies) is valid: on the large scale, the universe at any given time is the same everywhere (homogeneous and isotropic). As a result, the expansion rate must be the same everywhere during any epoch of cosmic time. If so, we don't need to think about the entire universe when we think about the expansion, we can just look at any sufficiently large portion of it. (Some models for dark energy would allow the expansion rate to be different in different directions, and scientists are designing experiments to test this idea. However, until such evidence is found, we will assume that the cosmological principle applies throughout the universe.)

In Galaxies, we hinted that when we think of the expansion of the universe, it is more correct to think of space itself stretching rather than of galaxies moving through static space. Nevertheless, we have since been discussing the redshifts of galaxies as if they resulted from the motion of the galaxies themselves.

Now, however, it is time to finally put such simplistic notions behind us and take a more sophisticated look at the cosmic expansion. Recall from our discussion of Einstein's theory of general relativity (in the chapter on Black Holes and Curved Spacetime) that space—or, more precisely, spacetime—is not a mere backdrop to the action of the universe, as Newton thought. Rather, it is an active participant—affected by and in turn affecting the matter and energy in the universe.

Since the expansion of the universe is the stretching of all spacetime, all points in the universe are stretching together. Thus, the expansion began *everywhere at once*. Unfortunately for tourist agencies of the future, there is no location you can visit where the stretching of space began or where we can say that the Big Bang happened.

To describe just how space stretches, we say the cosmic expansion causes the universe to undergo a uniform change in *scale* over time. By scale we mean, for example, the distance between two clusters of galaxies. It is customary to represent the scale by the factor R; if R doubles, then the distance between the clusters has doubled. Since the universe is expanding at the same rate everywhere, the change in R tells us how much it has expanded (or contracted) at any given time. For a static universe, R would be constant as time passes. In an expanding universe, R increases with time.

If it is space that is stretching rather than galaxies moving through space, then why do the galaxies show redshifts in their spectra? When you were young and naïve—a few chapters ago—it was fine to discuss the redshifts of distant galaxies as if they resulted from their motion away from us. But now that you are an older and wiser student of cosmology, this view will simply not do.

A more accurate view of the redshifts of galaxies is that the light waves are stretched by the stretching of the space they travel through. Think about the light from a remote galaxy. As it moves away from its source, the light has to travel through space. If space is stretching during all the time the light is traveling, the light waves will be stretched as well. A redshift is a stretching of waves—the wavelength of each wave increases (Figure 29.7). Light from more distant galaxies travels for more time than light

from closer ones. This means that the light has stretched more than light from closer ones and thus shows a greater redshift.

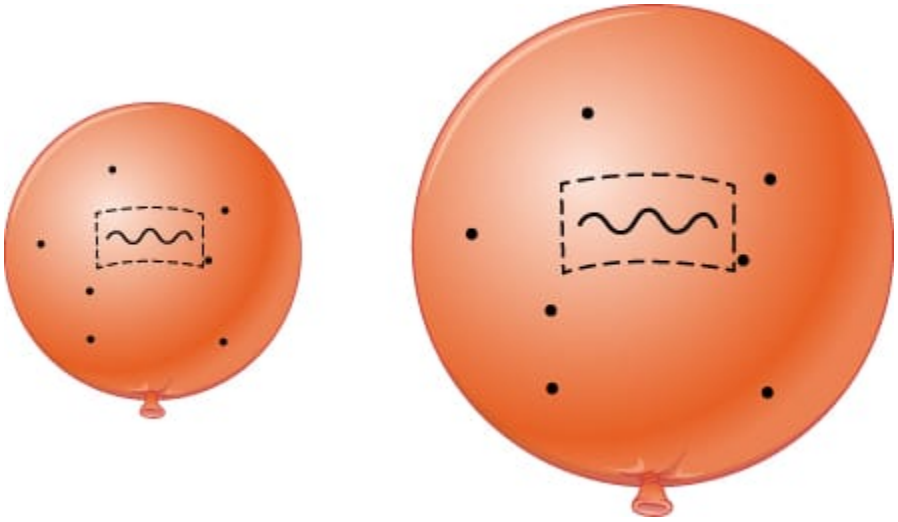


Figure 29.7 : Expansion and Redshift. As an elastic surface expands, a wave on its surface stretches. For light waves, the increase in wavelength would be seen as a redshift.

Thus, what the measured redshift of light from an object is telling us is how much the universe has expanded since the light left the object. If the universe has expanded by a factor of 2, then the wavelength of the light (and all electromagnetic waves from the same source) will have doubled.

Models of the Expansion

Before astronomers knew about dark energy or had a good measurement of how much matter exists in the universe, they made speculative models about how the universe might evolve over time. The four possible scenarios are shown in Figure 29.8. In this diagram, time moves forward from the bottom upward, and the scale of space increases by the horizontal circles becoming wider.

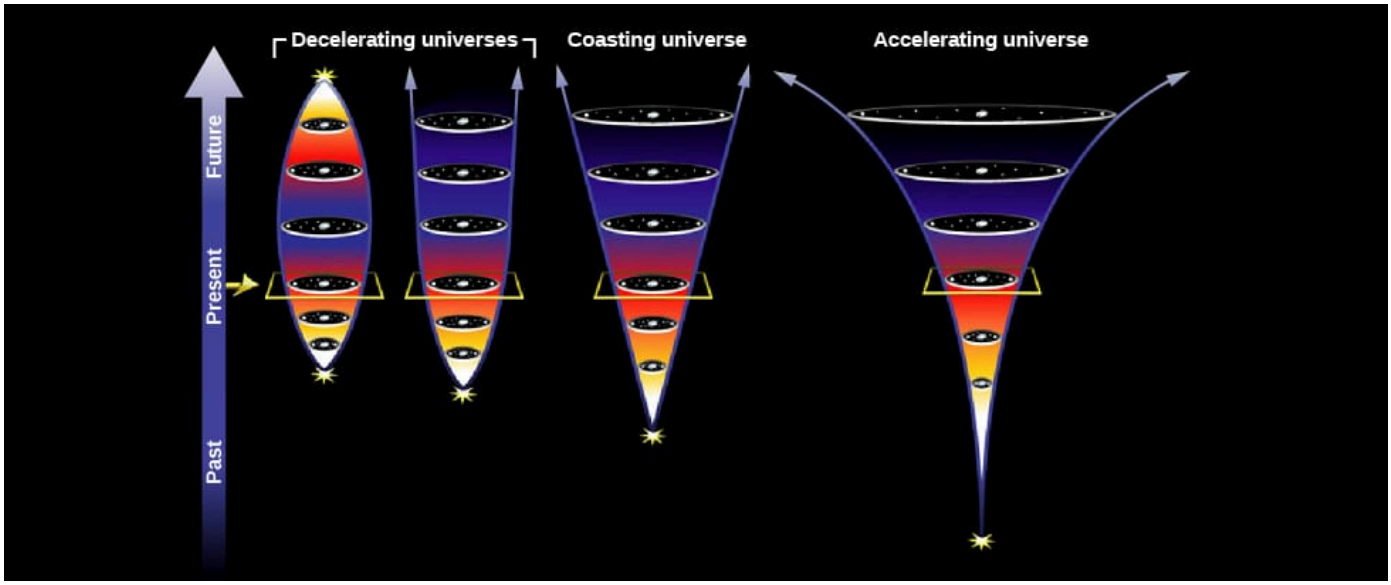


Figure 29.8 : Four Possible Models of the Universe. The yellow square marks the present in all four cases, and for all four, the Hubble constant is equal to the same value at the present time. Time is measured in the vertical direction. The first two universes on the left are ones in which the rate of expansion slows over time. The one on the left will eventually slow, come to a stop and reverse, ending up in a “big crunch,” while the one next to it will continue to expand forever, but ever-more slowly as time passes. The “coasting” universe is one that expands at a constant rate given by the Hubble constant throughout all of cosmic time. The accelerating universe on the right will continue to expand faster and faster forever. (credit: modification of work by NASA/ESA)

The simplest scenario of an expanding universe would be one in which R increases with time at a constant rate. But you already

know that life is not so simple. The universe contains a great deal of mass and its gravity decelerates the expansion—by a large amount if the universe contains a lot of matter, or by a negligible amount if the universe is nearly empty. Then there is the observed acceleration, which astronomers blame on a kind of dark energy.

Let's first explore the range of possibilities with models for different amounts of mass in the universe and for different contributions by dark energy. In some models—as we shall see—the universe expands forever. In others, it stops expanding and starts to contract. After looking at the extreme possibilities, we will look at recent observations that allow us to choose the most likely scenario.

We should perhaps pause for a minute to note how remarkable it is that we can do this at all. Our understanding of the principles that underlie how the universe works on the large scale and our observations of how the objects in the universe change with time allow us to model the evolution of the entire cosmos these days. It is one of the loftiest achievements of the human mind.

What astronomers look at in practice, to determine the kind of universe we live in, is the *average density* of the universe. This is the mass of matter (including the equivalent mass of energy)[1] that would be contained in each unit of volume (say, 1 cubic centimeter) if all the stars, galaxies, and other objects were taken apart, atom by atom, and if all those particles, along with the light and other energy, were distributed throughout all of space with absolute uniformity. If the average density is low, there is less mass and less gravity, and the universe will not decelerate very much. It can therefore expand forever. Higher average density, on the other hand, means there is more mass and more gravity and that the stretching of space might slow down enough that the expansion will eventually stop. An extremely high density might even cause the universe to collapse again.

For a given rate of expansion, there is a critical density—the mass per unit volume that will be just enough to slow the expansion to zero at some time infinitely far in the future. If the actual density is higher than this critical density, then the expansion will ultimately reverse and the universe will begin to contract. If the actual density is lower, then the universe will expand forever.

These various possibilities are illustrated in Figure 29.9. In this graph, one of the most comprehensive in all of science, we chart the development of the scale of space in the cosmos against the passage of time. Time increases to the right, and the scale of the universe, R, increases upward in the figure. Today, at the point marked "present" along the time axis, R is increasing in each model. We know that the galaxies are currently expanding away from each other, no matter which model is right. (The same situation holds for a baseball thrown high into the air. While it may eventually fall back down, near the beginning of the throw it moves upward most rapidly.)

The various lines moving across the graph correspond to different models of the universe. The straight dashed line corresponds to the empty universe with no deceleration; it intercepts the time axis at a time, T_0 (the Hubble time), in the past. This is not a realistic model but gives us a measure to compare other models to. The curves below the dashed line represent models with no dark energy and with varying amounts of deceleration, starting from the Big Bang at shorter times in the past. The curve above the dashed line shows what happens if the expansion is accelerating. Let's take a closer look at the future according to the different models.

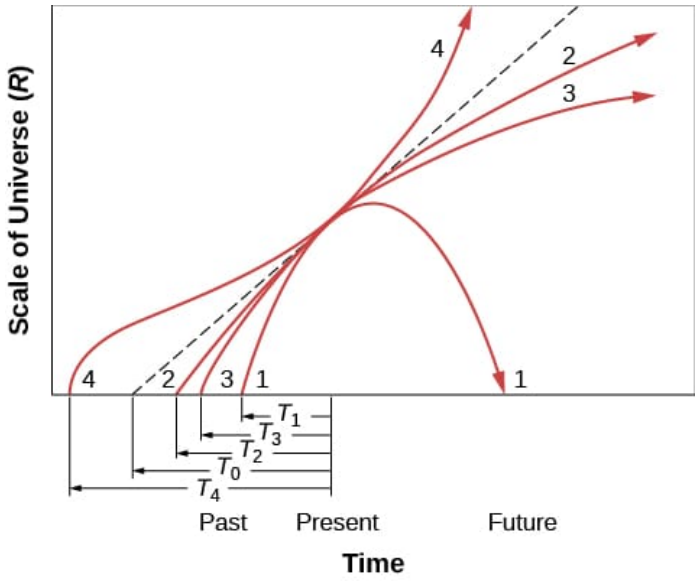


Figure 29.9 : Models of the Universe. This graph plots *R*, the scale of the universe, against time for various cosmological models. Curve 1 represents a universe where the density is greater than the critical value; this model predicts that the universe will eventually collapse. Curve 2 represents a universe with a density lower than critical; the universe will continue to expand but at an ever-slower rate. Curve 3 is a critical-density universe; in this universe, the expansion will gradually slow to a stop infinitely far in the future. Curve 4 represents a universe that is accelerating because of the effects of dark energy. The dashed line is for an empty universe, one in which the expansion is not slowed by gravity or accelerated by dark energy. Time is very compressed on this graph.

Let's start with curve 1 in Figure 29.9. In this case, the actual density of the universe is higher than the critical density and there is no dark energy. This universe will stop expanding at some time in the future and begin contracting. This model is called a closed universe and corresponds to the universe on the left in Figure 29.8. Eventually, the scale drops to zero, which means that space will have shrunk to an infinitely small size. The noted physicist John Wheeler called this the "big crunch," because matter, energy, space, and time would all be crushed out of existence. Note that the "big crunch" is the opposite of the Big Bang—it is an *implosion*. The universe is not expanding but rather collapsing in upon itself.

Some scientists speculated that another Big Bang might follow the crunch, giving rise to a new expansion phase, and then another contraction—perhaps oscillating between successive Big Bangs and big crunches indefinitely in the past and future. Such speculation was sometimes referred to as the oscillating theory of the universe. The challenge for theorists was how to describe the transition from collapse (when space and time themselves disappear into the big crunch) to expansion. With the discovery of dark energy, however, it does not appear that the universe will experience a big crunch, so we can put worrying about it on the back burner.

If the density of the universe is less than the critical density (curve 2 in Figure 29.9 and the universe second from the left in Figure 29.8), gravity is never important enough to stop the expansion, and so the universe expands forever. Such a universe is infinite and this model is called an open universe. Time and space begin with the Big Bang, but they have no end; the universe simply continues expanding, always a bit more slowly as time goes on. Groups of galaxies eventually get so far apart that it would be difficult for observers in any of them to see the others. (See the feature box on What Might the Universe Be Like in the Distant Future? for more about the distant future in the closed and open universe models.)

At the critical density (curve 3), the universe can just barely expand forever. The critical-density universe has an age of exactly two-thirds T_0, where T_0 is the age of the empty universe. Universes that will someday begin to contract have ages less than two-thirds T_0.

In an empty universe (the dashed line Figure 29.9 and the coasting universe in Figure 29.8), neither gravity nor dark energy is important enough to affect the expansion rate, which is therefore constant throughout all time.

In a universe with dark energy, the rate of the expansion will increase with time, and the expansion will continue at an ever-faster rate. Curve 4 in Figure 29.9, which represents this universe, has a complicated shape. In the beginning, when the matter is all very close together, the rate of expansion is most influenced by gravity. Dark energy appears to act only over large scales and thus becomes more important as the universe grows larger and the matter begins to thin out. In this model, at first the universe slows down, but as space stretches, the acceleration plays a greater role and the expansion speeds up.

The Cosmic Tug of War

We might summarize our discussion so far by saying that a "tug of war" is going on in the universe between the forces that push everything apart and the gravitational attraction of matter, which pulls everything together. If we can determine who will win this tug of war, we will learn the ultimate fate of the universe.

The first thing we need to know is the density of the universe. Is it greater than, less than, or equal to the critical density? The critical density today depends on the value of the expansion rate today, H_0. If the Hubble constant is around 20 kilometers/second per million light-years, the critical density is about 10^{-26} kg/m^3. Let's see how this value compares with the actual density of the universe.

Example 29.1: Critical Density of the Universe

As we discussed, the critical density is that combination of matter and energy that brings the universe coasting to a stop at time infinity. Einstein's equations lead to the following expression for the critical density (ρ_{crit}):

$$\rho_{crit} = \frac{3H^2}{8\pi G}$$

where H is the Hubble constant and G is the universal constant of gravity (6.67×10^{-11} Nm^2/kg^2).

Solution

Let's substitute our values and see what we get. Take an H = 22 km/s per million light-years. We need to convert both km and light-years into meters for consistency. A million light-years = $10^6 \times 9.5 \times 10^{15}$ m = 9.5×10^{21} m. And 22 km/s = 2.2×10^4 m/s. That makes $H = 2.3 \times 10^{-18}$ /s and $H^2 = 5.36 \times 10^{-36}$ /s^2. So,

$$\rho_{crit} = \frac{3 \times 5.36 \times 10^{-36}}{8 \times 3.14 \times 6.67 \times 10^{-11}} = 9.6 \times 10^{-27}\,\mathrm{kg/m^3}$$

which we can round off to the 10^{-26} kg/m^3. (To make the units work out, you have to know that N, the unit of force, is the same as kg × m/s^2.)

Now we can compare densities we measure in the universe to this critical value. Note that density is mass per unit volume, but energy has an equivalent mass of $m = E/c^2$ (from Einstein's equation $E = mc^2$).

Exercise 18.3.1

a. A single grain of dust has a mass of about 1.1×10^{-13} kg. If the average mass-energy density of space is equal to the critical density on average, how much space would be required to produce a total mass-energy equal to a dust grain?
b. If the Hubble constant were twice what it actually is, how much would the critical density be?

Answer available at the end of this section.

We can start our survey of how dense the cosmos is by ignoring the dark energy and just estimating the density of all matter in the universe, including ordinary matter and dark matter. Here is where the cosmological principle really comes in handy. Since the universe is the same all over (at least on large scales), we only need to measure how much matter exists in a (large) representative sample of it. This is similar to the way a representative survey of a few thousand people can tell us whom the millions of residents of the US prefer for president.

There are several methods by which we can try to determine the average density of matter in space. One way is to count all the galaxies out to a given distance and use estimates of their masses, including dark matter, to calculate the average density. Such estimates indicate a density of about 1 to 2×10^{-27} kg/m^3 (10 to 20% of critical), which by itself is too small to stop the expansion.

A lot of the dark matter lies outside the boundaries of galaxies, so this inventory is not yet complete. But even if we add an estimate of the dark matter outside galaxies, our total won't rise beyond about 30% of the critical density. We'll pin these numbers down more precisely later in this chapter, where we will also include the effects of dark energy.

In any case, even if we ignore dark energy, the evidence is that the universe will continue to expand forever. The discovery of dark energy that is causing the rate of expansion to speed up only strengthens this conclusion. Things definitely do not look good for fans of the closed universe (big crunch) model.

Making Connections: What Might the Universe Be Like in the Distant Future?

> 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.
> —From the poem "Fire and Ice" by Robert Frost (1923)

Given the destructive power of impacting asteroids, expanding red giants, and nearby supernovae, our species may not be around in the remote future. Nevertheless, you might enjoy speculating about what it would be like to live in a much, much older universe.

The observed acceleration makes it likely that we will have continued expansion into the indefinite future. If the universe expands forever (R increases without limit), the clusters of galaxies will spread ever farther apart with time. As eons pass, the universe will get thinner, colder, and darker.

Within each galaxy, stars will continue to go through their lives, eventually becoming white dwarfs, neutron stars, and black holes. Low-mass stars might take a long time to finish their evolution, but in this model, we would literally have all the time in the world. Ultimately, even the white dwarfs will cool down to be black dwarfs, any neutron stars that reveal themselves as pulsars will slowly stop spinning, and black holes with accretion disks will one day complete their "meals." The remains of stars will all be dark and difficult to observe.

This means that the light that now reveals galaxies to us will eventually go out. Even if a small pocket of raw material were left in one unsung corner of a galaxy, ready to be turned into a fresh cluster of stars, we will only have to wait until the time that their evolution is also complete. And time is one thing this model of the universe has plenty of. There will surely come a time when all the stars are out, galaxies are as dark as space, and no source of heat remains to help living things survive. Then the lifeless galaxies will just continue to move apart in their lightless realm.

If this view of the future seems discouraging (from a human perspective), keep in mind that we fundamentally do not understand why the expansion rate is currently accelerating. Thus, our speculations about the future are just that: speculations. You might take heart in the knowledge that science is always a progress report. The most advanced ideas about the universe from a hundred years ago now strike us as rather primitive. It may well be that our best models of today will in a hundred or a thousand years also seem rather simplistic and that there are other factors determining the ultimate fate of the universe of which we are still completely unaware.

Ages of Distant Galaxies

In the chapter on Galaxies, we discussed how we can use Hubble's law to measure the distance to a galaxy. But that simple method only works with galaxies that are not too far away. Once we get to large distances, we are looking so far into the past that we must take into account changes in the rate of the expansion of the universe. Since we cannot measure these changes directly, we must assume one of the models of the universe to be able to convert large redshifts into distances.

This is why astronomers squirm when reporters and students ask them exactly how far away some newly discovered distant quasar or galaxy is. We really can't give an answer without first explaining the model of the universe we are assuming in calculating it (by which time a reporter or student is long gone or asleep). Specifically, we must use a model that includes the change in the expansion rate with time. The key ingredients of the model are the amounts of matter, including dark matter, and the equivalent mass (according to $E = mc^2$) of the dark energy along with the Hubble constant.

Elsewhere in this book, we have estimated the mass density of ordinary matter plus dark matter as roughly 0.3 times the critical density, and the mass equivalent of dark energy as roughly 0.7 times the critical density. We will refer to these values as the "standard model of the universe." The latest (slightly improved) estimates for these values and the evidence for them will be given later in this chapter. Calculations also require the current value of the Hubble constant. For Table 29.1, we have adopted a Hubble constant of 67.3 kilometers/second/million parsecs (rather than rounding it to 70 kilometers/second/million parsecs), which is consistent with the 13.8 billion-year age of the universe estimated by the latest observations.

Once we assume a model, we can use it to calculate the age of the universe at the time an object emitted the light we see. As an example, Table 29.1 lists the times that light was emitted by objects at different redshifts as fractions of the current age of the universe. The times are given for two very different models so you can get a feeling for the fact that the calculated ages are fairly similar. The first model assumes that the universe has a critical density of matter and no dark energy. The second model is the standard model described in the preceding paragraph. The first column in the table is the redshift, which is given by the equation $z = \Delta\lambda/\lambda_0$ and is a measure of how much the wavelength of light has been stretched by the expansion of the universe on its long journey to us.

Table 29.1: Ages of the Universe at Different Redshifts

Redshift	Percent of Current Age of Universe When the Light Was Emitted (mass = critical density)	Percent of Current Age of Universe When the Light Was Emitted (mass = 0.3 critical density; dark energy = 0.7 critical density)
0	100 (now)	100 (now)
0.5	54	63
1.0	35	43
2.0	19	24
3.0	13	16
4.0	9	11
5.0	7	9
8.0	4	5
11.9	2.1	2.7
Infinite	0	0

Notice that as we find objects with higher and higher redshifts, we are looking back to smaller and smaller fractions of the age of the universe. The highest observed redshifts as this book is being written are close to 12 (Figure 29.10). As Table 29.1 shows, we are seeing these galaxies as they were when the universe was only about 3% as old as it is now. They were already formed only about 700 million years after the Big Bang.

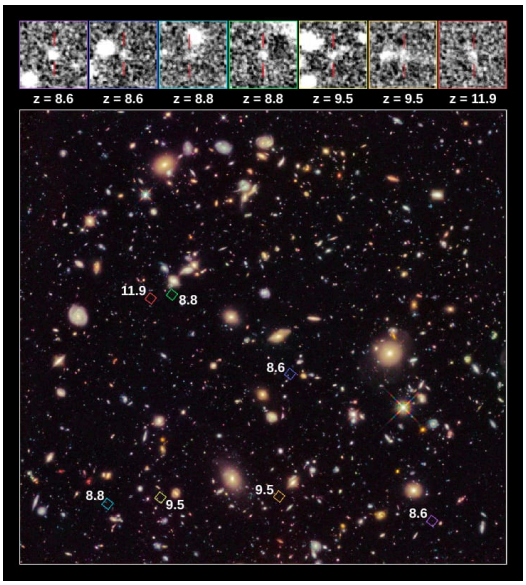


Figure 29.10 : Hubble Ultra-Deep Field. This image, called the Hubble Ultra Deep Field, shows faint galaxies, seen very far away and therefore very far back in time. The colored squares in the main image outline the locations of the galaxies. Enlarged views of each galaxy are shown in the black-and-white images. The red lines mark each galaxy's location. The "redshift" of each galaxy is indicated below each box, denoted by the symbol "z." The redshift measures how much a galaxy's ultraviolet and visible light has been stretched to infrared wavelengths by the universe's expansion. The larger the redshift, the more distant the galaxy, and therefore the further astronomers are seeing back in time. One of the seven galaxies may be a distance breaker, observed at a redshift of 11.9. If this redshift is confirmed by additional measurements, the galaxy is seen as it appeared only 380 million years after the Big Bang, when the universe was less than 3% of its present age. (credit: modification of work by NASA, ESA, R. Ellis (Caltech), and the UDF 2012 Team)

Footnotes

- 1By equivalent mass we mean that which would result if the energy were turned into mass using Einstein's formula, $E = mc^2$.

Answers

Exercise 18.3.1

a. In this case, the average mass-energy in a volume V of space is $E = \rho_{\text{crit}}V$. Thus, for space with critical density, we require that

$$V = \frac{E_{\text{grain}}}{\rho_{\text{crit}}} = \frac{1.1 \times 10^{-13}\,\text{kg}}{9.6 \times 10^{-26}\,\text{kg/m}^3} = 1.15 \times 10^{12}\,\text{m}^3 = (10{,}500\,\text{m})^3 \cong (10.5\,\text{km})^3$$

Thus, the sides of a cube of space with mass-energy density averaging that of the critical density would need to be slightly greater than 10 km to contain the total energy equal to a single grain of dust!

b. Since the critical density goes as the square of the Hubble constant, by doubling the Hubble parameter, the critical density would increase by a factor a four. So if the Hubble constant was 44 km/s per million light-years instead of 22 km/s per million light-years, the critical density would be $\rho_{\text{crit}} = 4 \times 9.6 \times 10^{-27}\,\text{kg/m}^3 = 3.8 \times 10^{-26}\,\text{kg/m}^3$.

This page titled 18.3: A Model of the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 29.3: A Model of the Universe by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

18.4: What Is the Universe Really Made Of?

Learning Objectives

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

- Specify what fraction of the density of the universe is contributed by stars and galaxies and how much ordinary matter (such as hydrogen, helium, and other elements we are familiar with here on Earth) makes up the overall density
- Describe how ideas about the contents of the universe have changed over the last 50 years
- Explain why it is so difficult to determine what dark matter really is
- Explain why dark matter helped galaxies form quickly in the early universe
- Summarize the evolution of the universe from the time the CMB was emitted to the present day

The model of the universe we described in the previous section is the simplest model that explains the observations. It assumes that general relativity is the correct theory of gravity throughout the universe. With this assumption, the model then accounts for the existence and structure of the CMB; the abundances of the light elements deuterium, helium, and lithium; and the acceleration of the expansion of the universe. All of the observations to date support the validity of the model, which is referred to as the standard (or concordance) model of cosmology.

Figure 29.21 and Table 29.2 summarize the current best estimates of the contents of the universe. Luminous matter in stars and galaxies and neutrinos contributes about 1% of the mass required to reach critical density. Another 4% is mainly in the form of hydrogen and helium in the space between stars and in intergalactic space. Dark matter accounts for about an additional 27% of the critical density. The mass equivalent of dark energy (according to $E = mc^2$) then supplies the remaining 68% of the critical density.

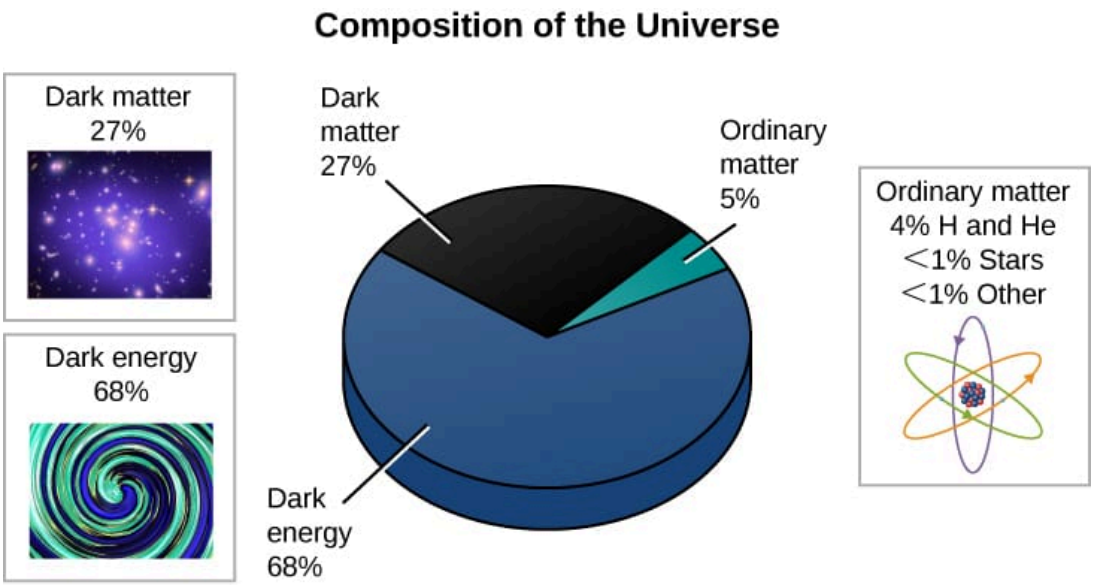


Figure 29.21 : Composition of the Universe. Only about 5% of all the mass and energy in the universe is matter with which we are familiar here on Earth. Most ordinary matter consists of hydrogen and helium located in interstellar and intergalactic space. Only about one-half of 1% of the critical density of the universe is found in stars. Dark matter and dark energy, which have not yet been detected in earthbound laboratories, account for 95% of the contents of the universe.

Table 29.2: What Different Kinds of Objects Contribute to the Density of the Universe

Object	Density as a Percent of Critical Density
Luminous matter (stars, etc.)	<1
Hydrogen and helium in interstellar and intergalactic space	4

Object	Density as a Percent of Critical Density
Dark matter	27
Equivalent mass density of the dark energy	68

This table should shock you. What we are saying is that 95% of the stuff of the universe is either dark matter or dark energy—neither of which has ever been detected in a laboratory here on Earth. This whole textbook, which has focused on objects that emit electromagnetic radiation, has generally been ignoring 95% of what is out there. Who says there aren't big mysteries yet to solve in science!

Figure 29.22 shows how our ideas of the composition of the universe have changed over just the past three decades. The fraction of the universe that we think is made of the same particles as astronomy students has been decreasing steadily.

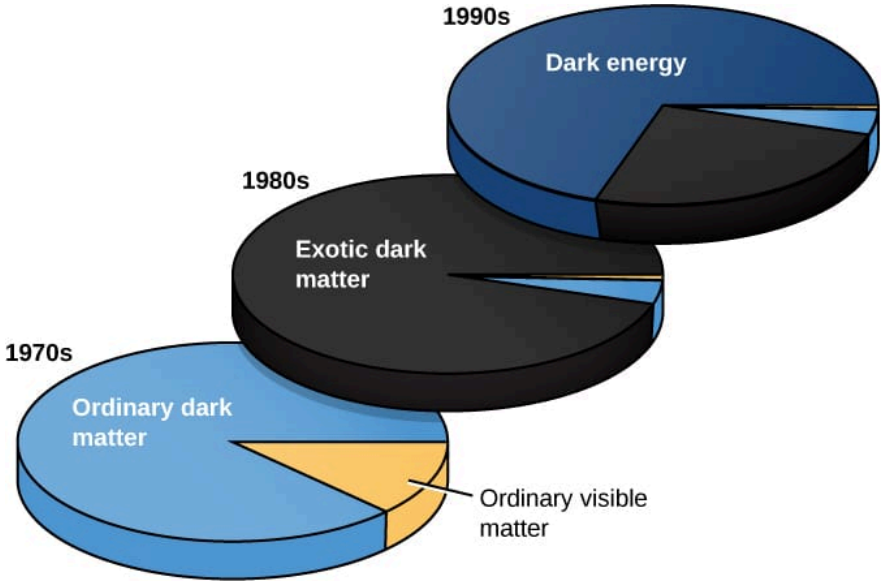


Figure 29.22 : Changing Estimates of the Content of the Universe. This diagram shows the changes in our understanding of the contents of the universe over the past three decades. In the 1970s, we suspected that most of the matter in the universe was invisible, but we thought that this matter might be ordinary matter (protons, neutrons, etc.) that was simply not producing electromagnetic radiation. By the 1980s, it was becoming likely that most of the dark matter was made of something we had not yet detected on Earth. By the late 1990s, a variety of experiments had shown that we live in a critical -density universe and that dark energy contributes about 70% of what is required to reach critical density. Note how the estimate of the relative importance of ordinary luminous matter (shown in yellow) has diminished over time.

What Is Dark Matter?

Many astronomers find the situation we have described very satisfying. Several independent experiments now agree on the type of universe we live in and on the inventory of what it contains. We seem to be very close to having a cosmological model that explains nearly everything. Others are not yet ready to jump on the bandwagon. They say, "show me the 96% of the universe we can't detect directly—for example, find me some dark matter!"

At first, astronomers thought that dark matter might be hidden in objects that appear dark because they emit no light (e.g., black holes) or that are too faint to be observed at large distances (e.g., planets or white dwarfs). However, these objects would be made of ordinary matter, and the deuterium abundance tells us that no more than 5% of the critical density consists of ordinary matter.

Another possible form that dark matter can take is some type of elementary particle that we have not yet detected here on Earth—a particle that has mass and exists in sufficient abundance to contribute 23% of the critical density. Some physics theories predict the existence of such particles. One class of these particles has been given the name WIMPs, which stands for weakly interacting massive particles. Since these particles do not participate in nuclear reactions leading to the production of deuterium, the deuterium abundance puts no limits on how many WIMPs might be in the universe. (A number of other exotic particles have also been suggested as prime constituents of dark matter, but we will confine our discussion to WIMPs as a useful example.)

If large numbers of WIMPs do exist, then some of them should be passing through our physics laboratories right now. The trick

is to catch them. Since by definition they interact only weakly (infrequently) with other matter, the chances that they will have a measurable effect are small. We don't know the mass of these particles, but various theories suggest that it might be a few to a few hundred times the mass of a proton. If WIMPs are 60 times the mass of a proton, there would be about 10 million of them passing through your outstretched hand every second—with absolutely no effect on you. If that seems too mind-boggling, bear in mind that neutrinos interact weakly with ordinary matter, and yet we were able to "catch" them eventually.

Despite the challenges, more than 30 experiments designed to detect WIMPS are in operation or in the planning stages. Predictions of how many times WIMPs might actually collide with the nucleus of an atom in the instrument designed to detect them are in the range of 1 event per year to 1 event per 1000 years per kilogram of detector. The detector must therefore be large. It must be shielded from radioactivity or other types of particles, such as neutrons, passing through it, and hence these detectors are placed in deep mines. The energy imparted to an atomic nucleus in the detector by collision with a WIMP will be small, and so the detector must be cooled to a very low temperature.

The WIMP detectors are made out of crystals of germanium, silicon, or xenon. The detectors are cooled to a few thousandths of a degree—very close to absolute zero. That means that the atoms in the detector are so cold that they are scarcely vibrating at all. If a dark matter particle collides with one of the atoms, it will cause the whole crystal to vibrate and the temperature therefore to increase ever so slightly. Some other interactions may generate a detectable flash of light.

A different kind of search for WIMPs is being conducted at the Large Hadron Collider (LHC) at CERN, Europe's particle physics lab near Geneva, Switzerland. In this experiment, protons collide with enough energy potentially to produce WIMPs. The LHC detectors cannot detect the WIMPs directly, but if WIMPs are produced, they will pass through the detectors, carrying energy away with them. Experimenters will then add up all the energy that they detect as a result of the collisions of protons to determine if any energy is missing.

So far, none of these experiments has detected WIMPs. Will the newer experiments pay off? Or will scientists have to search for some other explanation for dark matter? Only time will tell (Figure 29.23).



Figure 29.23 : Dark Matter. This cartoon from NASA takes a humorous look at how little we yet understand about dark matter. (credit: NASA)

Dark Matter and the Formation of Galaxies

As elusive as dark matter may be in the current-day universe, galaxies could not have formed quickly without it. Galaxies grew from density fluctuations in the early universe, and some had already formed only about 400–500 million years after the Big Bang. The observations with WMAP, Planck, and other experiments give us information on the size of those density fluctuations. It turns out that the density variations we observe are too small to have formed galaxies so soon after the Big Bang. In the hot, early universe, energetic photons collided with hydrogen and helium, and kept them moving so rapidly that gravity was still not strong enough to cause the atoms to come together to form galaxies. How can we reconcile this with the fact that galaxies *did* form and are all around us?

Our instruments that measure the CMB give us information about density fluctuations only for *ordinary matter*, which interacts with radiation. Dark matter, as its name indicates, does not interact with photons at all. Dark matter could have had much greater variations in density and been able to come together to form gravitational “traps” that could then have begun to attract ordinary matter immediately after the universe became transparent. As ordinary matter became increasingly concentrated, it could have turned into galaxies quickly thanks to these dark matter traps.

For an analogy, imagine a boulevard with traffic lights every half mile or so. Suppose you are part of a motorcade of cars accompanied by police who lead you past each light, even if it is red. So, too, when the early universe was opaque, radiation interacted with ordinary matter, imparting energy to it and carrying it along, sweeping past the concentrations of dark matter. Now suppose the police leave the motorcade, which then encounters some red lights. The lights act as traffic traps; approaching cars now have to stop, and so they bunch up. Likewise, after the early universe became transparent, ordinary matter interacted with radiation only occasionally and so could fall into the dark matter traps.

The Universe in a Nutshell

In the previous sections of this chapter, we traced the evolution of the universe progressively further back in time. Astronomical discovery has followed this path historically, as new instruments and new techniques have allowed us to probe ever closer to the beginning of time. The rate of expansion of the universe was determined from measurements of nearby galaxies. Determinations of the abundances of deuterium, helium, and lithium based on nearby stars and galaxies were used to put limits on how much ordinary matter is in the universe. The motions of stars in galaxies and of galaxies within clusters of galaxies could only be explained if there were large quantities of dark matter. Measurements of supernovae that exploded when the universe was about half as old as it is now indicated that the rate of expansion of the universe has sped up since those explosions occurred. Observations of extremely faint galaxies show that galaxies had begun to form when the universe was only 400–500 million years old. And observations of the CMB confirmed early theories that the universe was initially very hot.

But all this moving further and further backward in time might have left you a bit dizzy. So now let’s instead show how the universe evolves as time moves forward.

Figure 29.24 summarizes the entire history of the observable universe from the beginning in a single diagram. The universe was very hot when it began to expand. We have fossil remnants of the very early universe in the form of neutrons, protons, electrons, and neutrinos, and the atomic nuclei that formed when the universe was 3–4 minutes old: deuterium, helium, and a small amount of lithium. Dark matter also remains, but we do not yet know what form it is in.

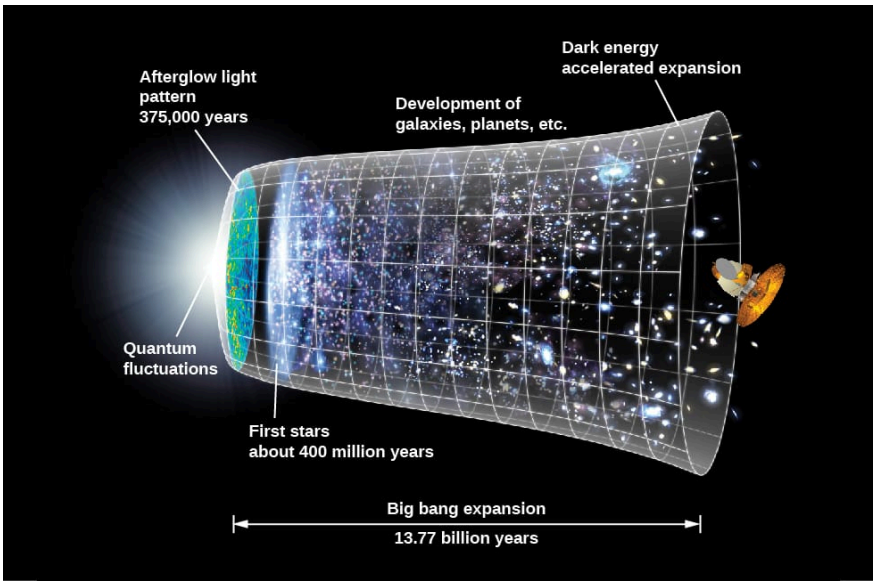


Figure 29.24 : History of the Universe. This image summarizes the changes that have occurred in the universe during the last 13.8 billion years. Protons, deuterium, helium, and some lithium were produced in the initial fireball. About 380,000 years after the Big Bang, the universe became transparent to electromagnetic radiation for the first time. COBE, WMAP, Planck, and other instruments have been used to study the radiation that was emitted at that time and that is still visible today (the CMB). The universe was then dark (except for this background radiation) until the first stars and galaxies began to form only a few hundred million years after the Big Bang. Existing space and ground-based telescopes have made substantial progress in studying the subsequent evolution of galaxies. (credit: modification of work by NASA/WMAP Science Team)

The universe gradually cooled; when it was about 380,000 years old, and at a temperature of about 3000 K, electrons combined with protons to form hydrogen atoms. At this point, as we saw, the universe became transparent to light, and astronomers have detected the CMB emitted at this time. The universe still contained no stars or galaxies, and so it entered what astronomers call “the dark ages” (since stars were not lighting up the darkness). During the next several hundred million years, small fluctuations in the density of the dark matter grew, forming gravitational traps that concentrated the ordinary matter, which began to form galaxies about 400–500 million years after the Big Bang.

By the time the universe was about a billion years old, it had entered its own renaissance: it was again blazing with radiation, but this time from newly formed stars, star clusters, and small galaxies. Over the next several billion years, small galaxies merged to form the giants we see today. Clusters and superclusters of galaxies began to grow, and the universe eventually began to resemble what we see nearby.

During the next 20 years, astronomers plan to build giant new telescopes both in space and on the ground to explore even further back in time. In 2021, the James Webb Space Telescope, a 6.5-meter telescope that is the successor to the Hubble Space Telescope, will be launched and assembled in space. The predictions are that with this powerful instrument (see Figure 29.1) we should be able to look back far enough to analyze in detail the formation of the first galaxies.

This page titled 18.4: What Is the Universe Really Made Of? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 29.6: What Is the Universe Really Made Of? by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

19: LIFE IN THE UNIVERSE

19.1: The Cosmic Context for Life
19.2: Astrobiology
19.3: Searching for Life beyond Earth
19.4: The Search for Extraterrestrial Intelligence

This page titled 19: Life in the Universe is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

19.1: The Cosmic Context for Life

Learning Objectives

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

- Describe the chemical and environmental conditions that make Earth hospitable to life
- Discuss the assumption underlying the Copernican principle and outline its implications for modern-day astronomers
- Understand the questions underlying the Fermi paradox

We saw that the universe was born in the Big Bang about 14 billion years ago. After the initial hot, dense fireball of creation cooled sufficiently for atoms to exist, all matter consisted of hydrogen and helium (with a very small amount of lithium). As the universe aged, processes within stars created the other elements, including those that make up Earth (such as iron, silicon, magnesium, and oxygen) and those required for life as we know it, such as carbon, oxygen, and nitrogen. These and other elements combined in space to produce a wide variety of compounds that form the basis of life on Earth. In particular, life on Earth is based on the presence of a key unit known as an organic molecule, a molecule that contains carbon. Especially important are the hydrocarbons, chemical compounds made up entirely of hydrogen and carbon, which serve as the basis for our biological chemistry, or *biochemistry*. While we do not understand the details of how life on Earth began, it is clear that to make creatures like us possible, events like the ones we described must have occurred, resulting in what is called the *chemical evolution* of the universe.

What Made Earth Hospitable to Life?

About 5 billion years ago, a cloud of gas and dust in this cosmic neighborhood began to collapse under its own weight. Out of this cloud formed the Sun and its planets, together with all the smaller bodies, such as comets, that also orbit the Sun (Figure 30.2). The third planet from the Sun, as it cooled, eventually allowed the formation of large quantities of liquid water on its surface.



Figure 30.2 : Comet Hyakutake. This image was captured in 1996 by NASA photographer Bill Ingalls. Comet impacts can deliver both water and a variety of interesting chemicals, including some organic chemicals, to Earth. (credit: NASA/Bill Ingalls)

The chemical variety and moderate conditions on Earth eventually led to the formation of molecules that could make copies of themselves (reproduce), which is essential for beginning life. Over the billions of years of Earth history, life evolved and became more complex. The course of evolution was punctuated by occasional planet-wide changes caused by collisions with some of the smaller bodies that did not make it into the Sun or one of its accompanying worlds. As we saw in the chapter on Earth as a Planet, mammals may owe their domination of Earth's surface to just such a collision 65 million years ago, which led to the extinction

of the dinosaurs (along with the majority of other living things). The details of such mass extinctions are currently the focus of a great deal of scientific interest.

Through many twisting turns, the course of evolution on Earth produced a creature with self-consciousness, able to ask questions about its own origins and place in the cosmos (Figure 30.3). Like most of Earth, this creature is composed of atoms that were forged in earlier generations of stars—in this case, assembled into both its body and brain. We might say that through the thoughts of human beings, the matter in the universe can become aware of itself.



Figure 30.3 : Young Human. Human beings have the intellect to wonder about their planet and what lies beyond it. Through them (and perhaps other intelligent life), the universe becomes aware of itself. (credit: Andrew Fraknoi)

Think about those atoms in your body for a minute. They are merely on loan to you from the lending library of atoms that make up our local corner of the universe. Atoms of many kinds circulate through your body and then leave it—with each breath you inhale and exhale and the food you eat and excrete. Even the atoms that take up more permanent residence in your tissues will not be part of you much longer than you are alive. Ultimately, you will return your atoms to the vast reservoir of Earth, where they will be incorporated into other structures and even other living things in the millennia to come.

This picture of *cosmic evolution*, of our descent from the stars, has been obtained through the efforts of scientists in many fields over many decades. Some of its details are still tentative and incomplete, but we feel reasonably confident in its broad outlines. It is remarkable how much we have been able to learn in the short time we have had the instruments to probe the physical nature of the universe.

The Copernican Principle

Our study of astronomy has taught us that we have always been wrong in the past whenever we have claimed that Earth is somehow unique. Galileo, using the newly invented technology of the telescope, showed us that Earth is not the center of the solar system, but merely one of a number of objects orbiting the Sun. Our study of the stars has demonstrated that the Sun itself is a rather undistinguished star, halfway through its long main-sequence stage like so many billions of others. There seems nothing special about our position in the Milky Way Galaxy either, and nothing surprising about our Galaxy's position in either its own group or its supercluster.

The discovery of planets around other stars confirms our idea that the formation of planets is a natural consequence of the formation of stars. We have identified thousands of exoplanets—planets orbiting around other stars, from huge ones orbiting close to their stars (informally called "hot Jupiters") down to planets smaller than Earth. A steady stream of exoplanet discoveries is leading to the conclusion that earthlike planets occur frequently—enough that there are likely many billions of "exo-Earths" in our own Milky Way Galaxy alone. From a planetary perspective, Earth-size planets are not unusual.

Philosophers of science sometimes call the idea that there is nothing special about our place in the universe the *Copernican principle*. Given all of the above, most scientists would be surprised if life were limited to our planet and had started nowhere

else. There are billions of stars in our Galaxy old enough for life to have developed on a planet around them, and there are billions of other galaxies as well. Astronomers and biologists have long conjectured that a series of events similar to those on the early Earth probably led to living organisms on many planets around other stars, and possibly even on other planets in our solar system, such as Mars.

The real scientific issue (which we do not currently know the answer to) is whether organic biochemistry is likely or unlikely in the universe at large. Are we a fortunate and exceedingly rare outcome of chemical evolution, or is organic biochemistry a regular part of the chemical evolution of the cosmos? We do not yet know the answer to this question, but data, even an exceedingly small amount (like finding “unrelated to us” living systems on a world like Europa), will help us arrive at it.

So Where Are They?

If the Copernican principle is applied to life, then biology may be rather common among planets. Taken to its logical limit, the Copernican principle also suggests that intelligent life like us might be common. Intelligence like ours has some very special properties, including an ability to make progress through the application of technology. Organic life around other (older) stars may have started a billion years earlier than we did on Earth, so they may have had a lot more time to develop advanced technology such as sending information, probes, or even life-forms between stars.

Faced with such a prospect, physicist Enrico Fermi asked a question several decades ago that is now called the *Fermi paradox*: where are they? If life and intelligence are common and have such tremendous capacity for growth, why is there not a network of galactic civilizations whose presence extends even into a “latecomer” planetary system like ours?

Several solutions have been suggested to the Fermi paradox. Perhaps life is common but intelligence (or at least technological civilization) is rare. Perhaps such a network will come about in the future but has not yet had the time to develop. Maybe there are invisible streams of data flowing past us all the time that we are not advanced enough or sensitive enough to detect. Maybe advanced species make it a practice not to interfere with immature, developing consciousness such as our own. Or perhaps civilizations that reach a certain level of technology then self-destruct, meaning there are no other civilizations now existing in our Galaxy. We do not yet know whether any advanced life is out there and, if it is, why we are not aware of it. Still, you might want to keep these issues in mind as you read the rest of this chapter.

Link to Learning

Is there a network of galactic civilizations beyond our solar system? If so, why can’t we see them? Explore the possibilities in the cartoon video “The Fermi Paradox—Where Are All the Aliens?”

This page titled 19.1: The Cosmic Context for Life is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 30.2: The Cosmic Context for Life by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

19.2: Astrobiology

Learning Objectives

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

- Describe the chemical building blocks required for life
- Describe the molecular systems and processes driving the origin and evolution of life
- Describe the characteristics of a habitable environment
- Describe some of the extreme conditions on Earth, and explain how certain organisms have adapted to these conditions

Scientists today take a multidisciplinary approach to studying the origin, evolution, distribution, and ultimate fate of life in the universe; this field of study is known as astrobiology. You may also sometimes hear this field referred to as *exobiology* or *bioastronomy.* Astrobiology brings together astronomers, planetary scientists, chemists, geologists, and biologists (among others) to work on the same problems from their various perspectives.

Among the issues that astrobiologists explore are the conditions in which life arose on Earth and the reasons for the extraordinary adaptability of life on our planet. They are also involved in identifying habitable worlds beyond Earth and in trying to understand in practical terms how to look for life on those worlds. Let's look at some of these issues in more detail.

The Building Blocks of Life

While no unambiguous evidence for life has yet been found anywhere beyond Earth, life's chemical building blocks have been detected in a wide range of extraterrestrial environments. Meteorites (which you learned about in Cosmic Samples and the Origin of the Solar System) have been found to contain two kinds of substances whose chemical structures mark them as having an extraterrestrial origin—amino acids and sugars. Amino acids are organic compounds that are the molecular building blocks of proteins. Proteins are key biological molecules that provide the structure and function of the body's tissues and organs and essentially carry out the "work" of the cell. When we examine the gas and dust around comets, we also find a number of organic molecules—compounds that on Earth are associated with the chemistry of life.

Expanding beyond our solar system, one of the most interesting results of modern radio astronomy has been the discovery of organic molecules in giant clouds of gas and dust between stars. More than 100 different molecules have been identified in these reservoirs of cosmic raw material, including formaldehyde, alcohol, and others we know as important stepping stones in the development of life on Earth. Using radio telescopes and radio spectrometers, astronomers can measure the abundances of various chemicals in these clouds. We find organic molecules most readily in regions where the interstellar dust is most abundant, and it turns out these are precisely the regions where star formation (and probably planet formation) happen most easily (Figure 30.4).



Figure 30.4 : Cloud of Gas and Dust. This cloud of gas and dust in the constellation of Scorpius is the sort of region where complex molecules are found. It is also the sort of cloud where new stars form from the reservoir of gas and dust in the cloud. Radiation from a group of hot stars (off the picture to the bottom left) called the Scorpius OB Association is “eating into” the cloud, sweeping it into an elongated shape and causing the reddish glow seen at its tip. (credit: Dr. Robert Gendler)

Clearly the early Earth itself produced some of the molecular building blocks of life. Since the early 1950s, scientists have tried to duplicate in their laboratories the chemical pathways that led to life on our planet. In a series of experiments known as the *Miller-Urey experiments*, pioneered by Stanley Miller and Harold Urey at the University of Chicago, biochemists have simulated conditions on early Earth and have been able to produce some of the fundamental building blocks of life, including those that form proteins and other large biological molecules known as nucleic acids (which we will discuss shortly).

Although these experiments produced encouraging results, there are some problems with them. The most interesting chemistry from a biological perspective takes place with hydrogen-rich or *reducing* gases, such as ammonia and methane. However, the early atmosphere of Earth was probably dominated by carbon dioxide (as Venus’ and Mars’ atmospheres still are today) and may not have contained an abundance of reducing gases comparable to that used in Miller-Urey type experiments. Hydrothermal vents—seafloor systems in which ocean water is superheated and circulated through crustal or mantle rocks before reemerging into the ocean—have also been suggested as potential contributors of organic compounds on the early Earth, and such sources would not require Earth to have an early reducing atmosphere.

Both earthly and extraterrestrial sources may have contributed to Earth’s early supply of organic molecules, although we have more direct evidence for the latter. It is even conceivable that life itself originated elsewhere and was seeded onto our planet—although this, of course, does not solve the problem of how that life originated to begin with.

Link to Learning

Hydrothermal vents are beginning to seem more likely as early contributors to the organic compounds found on Earth. Read about hydrothermal vents and watch videos and slideshows on these and other deep-sea wonders at the Woods Hole Oceanographic Institution website.

Try an interactive simulation of hydrothermal-vent circulation at the Dive and Discover website.

The Origin and Early Evolution of Life

The carbon compounds that form the chemical basis of life may be common in the universe, but it is still a giant step from these building blocks to a living cell. Even the simplest molecules of the genes (the basic functional units that carry the genetic, or hereditary, material in a cell) contain millions of molecular units, each arranged in a precise sequence. Furthermore, even the most primitive life required two special capabilities: a means of extracting energy from its environment, and a means of encoding and replicating information in order to make faithful copies of itself. Biologists today can see ways that either of these capabilities

might have formed in a natural environment, but we are still a long way from knowing how the two came together in the first life-forms.

We have no solid evidence for the pathway that led to the origin of life on our planet except for whatever early history may be retained in the biochemistry of modern life. Indeed, we have very little direct evidence of what Earth itself was like during its earliest history—our planet is so effective at resurfacing itself through plate tectonics (see the chapter on Earth as a Planet) that very few rocks remain from this early period. In the earlier chapter on Cratered Worlds, you learned that Earth was subjected to a heavy bombardment—a period of large impact events—some 3.8 to 4.1 billion years ago. Large impacts would have been energetic enough to heat-sterilize the surface layers of Earth, so that even if life had begun by this time, it might well have been wiped out.

When the large impacts ceased, the scene was set for a more peaceful environment on our planet. If the oceans of Earth contained accumulated organic material from any of the sources already mentioned, the ingredients were available to make living organisms. We do not understand in any detail the sequence of events that led from molecules to biology, but there is fossil evidence of microbial life in 3.5-billion-year-old rocks, and possible (debated) evidence for life as far back as 3.8 billion years.

Life as we know it employs two main molecular systems: the functional molecules known as proteins, which carry out the chemical work of the cell, and information-containing molecules of DNA (deoxyribonucleic acid) that store information about how to create the cell and its chemical and structural components. The origin of life is sometimes considered a “chicken and egg problem” because, in modern biology, neither of these systems works without the other. It is our proteins that assemble DNA strands in the precise order required to store information, but the proteins are created based on information stored in DNA. Which came first? Some origin of life researchers believe that prebiotic chemistry was based on molecules that could both store information and do the chemical work of the cell. It has been suggested that RNA (ribonucleic acid), a molecule that aids in the flow of genetic information from DNA to proteins, might have served such a purpose. The idea of an early “RNA world” has become increasingly accepted, but a great deal remains to be understood about the origin of life.

Perhaps the most important innovation in the history of biology, apart from the origin of life itself, was the discovery of the process of photosynthesis, the complex sequence of chemical reactions through which some living things can use sunlight to manufacture products that store energy (such as carbohydrates), releasing oxygen as one by-product. Previously, life had to make do with sources of chemical energy available on Earth or delivered from space. But the abundant energy available in sunlight could support a larger and more productive biosphere, as well as some biochemical reactions not previously possible for life. One of these was the production of oxygen (as a waste product) from carbon dioxide, and the increase in atmospheric levels of oxygen about 2.4 billion years ago means that oxygen-producing photosynthesis must have emerged and become globally important by this time. In fact, it is likely that oxygen-producing photosynthesis emerged considerably earlier.

Some forms of chemical evidence contained in ancient rocks, such as the solid, layered rock formations known as stromatolites, are thought to be the fossils of oxygen-producing photosynthetic bacteria in rocks that are almost 3.5 billion years old (Figure 30.5). It is generally thought that a simpler form of photosynthesis that does not produce oxygen (and is still used by some bacteria today) probably preceded oxygen-producing photosynthesis, and there is strong fossil evidence that one or the other type of photosynthesis was functioning on Earth at least as far back as 3.4 billion years ago.



Figure 30.5 : Stromatolites Preserve the Earliest Physical Representation of Life on Earth. In their reach for sunlight, the single-celled microbes formed mats that trapped sediments in the water above them. Such trapped sediments fell and formed layers on top of the mats. The microbes then climbed atop the sediment layers and trapped more sediment. What is found in the rock record are (a) the solidified, curved sedimentary layers that are signatures of biological activity. The earliest known stromatolite is 3.47 billion years old and is found in Western Australia. (b) This more recent example is in Lake Thetis, also in Western Australia. (credit a: modification of work by James St. John; credit b: modification of work by Ruth Ellison)

The free oxygen produced by photosynthesis began accumulating in our atmosphere about 2.4 billion years ago. The interaction of sunlight with oxygen can produce ozone (which has three atoms of oxygen per molecule, as compared to the two atoms per molecule in the oxygen we breathe), which accumulated in a layer high in Earth's atmosphere. As it does on Earth today, this ozone provided protection from the Sun's damaging ultraviolet radiation. This allowed life to colonize the landmasses of our planet instead of remaining only in the ocean.

The rise in oxygen levels was deadly to some microbes because, as a highly reactive chemical, it can irreversibly damage some of the biomolecules that early life had developed in the absence of oxygen. For other microbes, it was a boon: combining oxygen with organic matter or other reduced chemicals generates a lot of energy—you can see this when a log burns, for example—and many forms of life adopted this way of living. This new energy source made possible a great proliferation of organisms, which continued to evolve in an oxygen-rich environment.

The details of that evolution are properly the subject of biology courses, but the process of evolution by natural selection (survival of the fittest) provides a clear explanation for the development of Earth's remarkable variety of life-forms. It does not, however, directly solve the mystery of life's earliest beginnings. We hypothesize that life will arise whenever conditions are appropriate, but this hypothesis is just another form of the Copernican principle. We now have the potential to address this hypothesis with observations. If a second example of life is found in our solar system or a nearby star, it would imply that life emerges commonly enough that the universe is likely filled with biology. To make such observations, however, we must first decide where to focus our search.

Link to Learning

Just how did life arise in the first place? And could it have happened with a different type of chemistry? Watch the 15-minute video Making Matter Come Alive in which a chemistry expert explores some answers to these questions, from a 2011 TED Talk.

Habitable Environments

Among the staggering number of objects in our solar system, Galaxy, and universe, some may have conditions suitable for life, while others do not. Understanding what conditions and features make a habitable environment—an environment capable of hosting life—is important both for understanding how widespread habitable environments may be in the universe and for focusing

a search for life beyond Earth. Here, we discuss habitability from the perspective of the life we know. We will explore the basic requirements of life and, in the following section, consider the full range of environmental conditions on Earth where life is found. While we can't entirely rule out the possibility that other life-forms might have biochemistry based on alternatives to carbon and liquid water, such life "as we don't know it" is still completely speculative. In our discussion here, we are focusing on habitability for life that is chemically similar to that on Earth.

Life requires a solvent (a liquid in which chemicals can dissolve) that enables the construction of biomolecules and the interactions between them. For life as we know it, that solvent is water, which has a variety of properties that are critical to how our biochemistry works. Water is abundant in the universe, but life requires that water be in liquid form (rather than ice or gas) in order to properly fill its role in biochemistry. That is the case only within a certain range of temperatures and pressures—too high or too low in either variable, and water takes the form of a solid or a gas. Identifying environments where water is present within the appropriate range of temperature and pressure is thus an important first step in identifying habitable environments. Indeed, a "follow the water" strategy has been, and continues to be, a key driver in the exploration of planets both within and beyond our solar system.

Our biochemistry is based on molecules made of carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur. Carbon is at the core of organic chemistry. Its ability to form four bonds, both with itself and with the other elements of life, allows for the formation of a vast number of potential molecules on which to base biochemistry. The remaining elements contribute structure and chemical reactivity to our biomolecules, and form the basis of many of the interactions among them. These "biogenic elements," sometimes referred to with the acronym CHNOPS (carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur), are the raw materials from which life is assembled, and an accessible supply of them is a second requirement of habitability.

As we learned in previous chapters on nuclear fusion and the life story of the stars, carbon, nitrogen, oxygen, phosphorus, and sulfur are all formed by fusion within stars and then distributed out into their galaxy as those stars die. But how they are distributed among the planets that form within a new star system, in what form, and how chemical, physical, and geological processes on those planets cycle the elements into structures that are accessible to biology, can have significant impacts on the distribution of life. In Earth's oceans, for example, the abundance of phytoplankton (simple organisms that are the base of the ocean food chain) in surface waters can vary by a thousand-fold because the supply of nitrogen differs from place to place (Figure 30.6). Understanding what processes control the accessibility of elements at all scales is thus a critical part of identifying habitable environments.

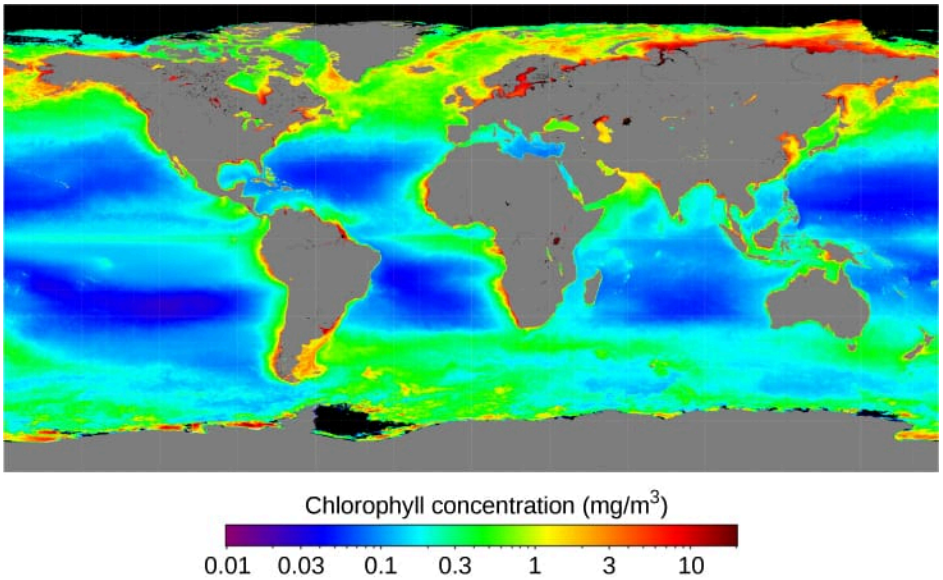


Figure 30.6 : Chlorophyll Abundance. The abundance of chlorophyll (an indicator of photosynthetic bacteria and algae) varies by almost a thousand-fold across the ocean basins. That variation is almost entirely due to the availability of nitrogen—one of the major "biogenic elements" in forms that can be used by life. (credit: modification of work by NASA, Gene C. Feldman)

With these first two requirements, we have the elemental raw materials of life and a solvent in which to assemble them into the complicated molecules that drive our biochemistry. But carrying out that assembly and maintaining the complicated biochemical machinery of life takes energy. You fulfill your own requirement for energy every time you eat food or take a breath, and you would not live for long if you failed to do either on a regular basis. Life on Earth makes use of two main types of energy: for you, these are the oxygen in the air you breathe and the organic molecules in your food. But life overall can use a much wider array of chemicals

and, while all animals require oxygen, many bacteria do not. One of the earliest known life processes, which still operates in some modern microorganisms, combines hydrogen and carbon dioxide to make methane, releasing energy in the process. There are microorganisms that “breathe” metals that would be toxic to us, and even some that breathe in sulfur and breathe out sulfuric acid. Plants and photosynthetic microorganisms have also evolved mechanisms to use the energy in light directly.

Water in the liquid phase, the biogenic elements, and energy are the fundamental requirements for habitability. But are there additional environmental constraints? We consider this in the next section.



Figure 30.7 : Grand Prismatic Spring in Yellowstone National Park. This hot spring, where water emerges from the bluish center at temperatures near the local boiling point (about 92 °C), supports a thriving array of microbial life. The green, yellow, and orange colors around the edges come from thick “mats” of photosynthetic bacteria. In fact, their coloration in part demonstrates their use of light energy—some wavelengths of incoming sunlight are selectively captured for energy; the rest are reflected back. Since it lacks the captured wavelengths, this light is now different in color than the sunlight that illuminates it. The blue part of the spring has temperatures too high to allow photosynthetic life (hence the lack of color except that supplied by water itself), but life is still present. Here, at nearly boiling temperatures, bacteria use the chemical energy supplied by the combination of hydrogen and other chemicals with oxygen. (credit: modification of work by Domenico Salvagnin)

Life in Extreme Conditions

At a chemical level, life consists of many types of molecules that interact with one another to carry out the processes of life. In addition to water, elemental raw materials, and energy, life also needs an environment in which those complicated molecules are stable (don’t break down before they can do their jobs) and their interactions are possible. Your own biochemistry works properly only within a very narrow range of about 10 °C in body temperature and two-tenths of a unit in blood pH (pH is a numerical measure of acidity, or the amount of free hydrogen ions). Beyond those limits, you are in serious danger.

Life overall must also have limits to the conditions in which it can properly work but, as we will see, they are much broader than human limits. The resources that fuel life are distributed across a very wide range of conditions. For example, there is abundant chemical energy to be had in hot springs that are essentially boiling acid (see Figure 30.7). This provides ample incentive for evolution to fill as much of that range with life as is biochemically possible. An organism (usually a microbe) that tolerates or even thrives under conditions that most of the life around us would consider hostile, such as very high or low temperature or acidity, is known as an extremophile (where the suffix *-phile* means “lover of”). Let’s have a look at some of the conditions that can challenge life and the organisms that have managed to carve out a niche at the far reaches of possibility.

Both high and low temperatures can cause a problem for life. As a large organism, you are able to maintain an almost constant body temperature whether it is colder or warmer in the environment around you. But this is not possible at the tiny size of microorganisms; whatever the temperature in the outside world is also the temperature of the microbe, and its biochemistry must be able to function at that temperature. High temperatures are the enemy of complexity—increasing thermal energy tends to break apart big molecules into smaller and smaller bits, and life needs to stabilize the molecules with stronger bonds and special proteins. But this approach has its limits.

Nevertheless, as noted earlier, high-temperature environments like hot springs and hydrothermal vents often offer abundant sources of chemical energy and therefore drive the evolution of organisms that can tolerate high temperatures (see Figure 30.8); such an organism is called a thermophile. Currently, the high temperature record holder is a methane-producing microorganism that can grow at 122 °C, where the pressure also is so high that water still does not boil. That's amazing when you think about it. We cook our food—meaning, we alter the chemistry and structure of its biomolecules—by boiling it at a temperature of 100 °C. In fact, food begins to cook at much lower temperatures than this. And yet, there are organisms whose biochemistry remains intact and operates just fine at temperatures 20 degrees higher.

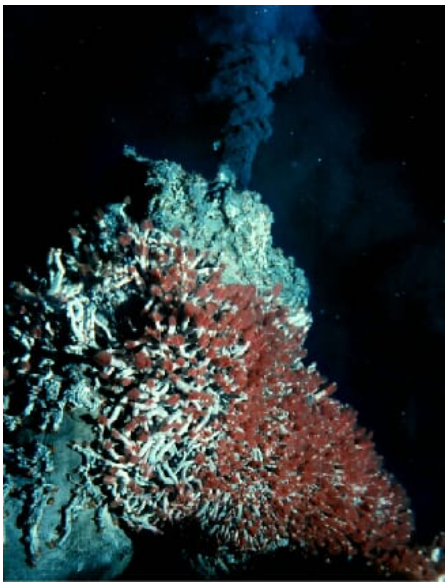


Figure 30.8 : Hydrothermal Vent on the Sea Floor. What appears to be black smoke is actually superheated water filled with minerals of metal sulfide. Hydrothermal vent fluid can represent a rich source of chemical energy, and therefore a driver for the evolution of microorganisms that can tolerate high temperatures. Bacteria feeding on this chemical energy form the base of a food chain that can support thriving communities of animals—in this case, a dense patch of red and white tubeworms growing around the base of the vent. (credit: modification of work by the University of Washington; NOAA/OAR/OER)

Cold can also be a problem, in part because it slows down metabolism to very low levels, but also because it can cause physical changes in biomolecules. Cell membranes—the molecular envelopes that surround cells and allow their exchange of chemicals with the world outside—are basically made of fatlike molecules. And just as fat congeals when it cools, membranes crystallize, changing how they function in the exchange of materials in and out of the cell. Some cold-adapted cells (called *psychrophiles*) have changed the chemical composition of their membranes in order to cope with this problem; but again, there are limits. Thus far, the coldest temperature at which any microbe has been shown to reproduce is about –25 °C.

Conditions that are very acidic or alkaline can also be problematic for life because many of our important molecules, like proteins and DNA, are broken down under such conditions. For example, household drain cleaner, which does its job by breaking down the chemical structure of things like hair clogs, is a very alkaline solution. The most acid-tolerant organisms (*acidophiles*) are capable of living at pH values near zero—about ten million times more acidic than your blood (Figure 30.9). At the other extreme, some *alkaliphiles* can grow at pH levels of about 13, which is comparable to the pH of household bleach and almost a million times more alkaline than your blood.



Figure 30.9 : Spain's Rio Tinto. With a pH close to 2, Rio Tinto is literally a river of acid. Acid-loving microorganisms (acidophiles) not only thrive in these waters, their metabolic activities help generate the acid in the first place. The rusty red color that gives the river its name comes from high levels of iron dissolved in the waters.

High levels of salts in the environment can also cause a problem for life because the salt blocks some cellular functions. Humans recognized this centuries ago and began to salt-cure food to keep it from spoiling—meaning, to keep it from being colonized by microorganisms. Yet some microbes have evolved to grow in water that is saturated in sodium chloride (table salt)—about ten times as salty as seawater (Figure 30.10).



Figure 30.10 : Salt Ponds. The waters of an evaporative salt works near San Francisco are colored pink by thriving communities of photosynthetic organisms. These waters are about ten times as salty as seawater—enough for sodium chloride to begin to crystallize out—yet some organisms can survive and thrive in these conditions. (credit: modification of work by NASA)

Very high pressures can literally squeeze life's biomolecules, causing them to adopt more compact forms that do not work very well. But we still find life—not just microbial, but even animal life—at the bottoms of our ocean trenches, where pressures are more than 1000 times atmospheric pressure. Many other adaptions to environmental "extremes" are also known. There is even an organism, *Deinococcus radiodurans,* that can tolerate ionizing radiation (such as that released by radioactive elements) a thousand times more intense than you would be able to withstand. It is also very good at surviving extreme desiccation (drying out) and a variety of metals that would be toxic to humans.

From many such examples, we can conclude that life is capable of tolerating a wide range of environmental extremes—so much

so that we have to work hard to identify places where life can't exist. A few such places are known—for example, the waters of hydrothermal vents at over 300 °C appear too hot to support any life—and finding these places helps define the possibility for life elsewhere. The study of extremophiles over the last few decades has expanded our sense of the range of conditions life can survive and, in doing so, has made many scientists more optimistic about the possibility that life might exist beyond Earth.

This page titled 19.2: Astrobiology is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 30.3: Astrobiology by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

19.3: Searching for Life beyond Earth

Learning Objectives

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

- Outline what we have learned from exploration of the environment on Mars
- Identify where in the solar system life is most likely sustainable and why
- Describe some key missions and their findings in our search for life beyond our solar system
- Explain the use of biomarkers in the search for evidence of life beyond our solar system

Astronomers and planetary scientists continue to search for life in the solar system and the universe at large. In this section, we discuss two kinds of searches. First is the direct exploration of planets within our own solar system, especially Mars and some of the icy moons of the outer solar system. Second is the even more difficult task of searching for evidence of life—a biomarker—on planets circling other stars. In the next section, we will examine SETI, the *search for extraterrestrial intelligence*. As you will see, the approaches taken in these three cases are very different, even though the goal of each is the same: to determine if life on Earth is unique in the universe.

Life on Mars

The possibility that Mars hosts, or has hosted, life has a rich history dating back to the "canals" that some people claimed to see on the martian surface toward the end of the nineteenth century and the beginning of the twentieth. With the dawn of the space age came the possibility to address this question up close through a progression of missions to Mars that began with the first successful flyby of a robotic spacecraft in 1964 and have led to the deployment of capable rovers, like *Curiosity* and *Perseverance*, with instruments to look for organic chemistry.

The earliest missions to Mars provided some hints that liquid water—one of life's primary requirements—may once have flowed on the surface, and later missions have strengthened this conclusion. The NASA Viking landers, whose purpose was to search directly for evidence of life on Mars, arrived on Mars in 1976. Viking's onboard instruments found no organic molecules (the stuff of which life is made), and no evidence of biological activity in the martian soils it analyzed.

This result is not particularly surprising because, despite the evidence of flowing liquid water in the past, liquid water on the surface of Mars is generally not stable today. Over much of Mars, temperatures and pressures at the surface are so low that pure water would either freeze or boil away (under very low pressures, water will boil at a much lower temperature than usual). To make matters worse, unlike Earth, Mars does not have a magnetic field and ozone layer to protect the surface from harmful solar ultraviolet radiation and energetic particles. However, Viking's analyses of the soil said nothing about whether life may have existed in Mars' distant past, when liquid water was more abundant. We do know that water in the form of ice exists in abundance on Mars, not so deep beneath its surface. Water vapor is also a constituent of the atmosphere of Mars.

Since the visit of Viking, our understanding of Mars has deepened spectacularly. Orbiting spacecraft have provided ever-more detailed images of the surface and detected the presence of minerals that could have formed only in the presence of liquid water. Two bold surface missions, the Mars Exploration Rovers *Spirit* and *Opportunity* (2004), followed by the much larger *Curiosity* Rover (2012), confirmed these remote-sensing data. All three rovers found abundant evidence for a past history of liquid water, revealed not only from the mineralogy of rocks they analyzed, but also from the unique layering of rock formations.

Curiosity has gone a step beyond evidence for water and confirmed the existence of habitable environments on ancient Mars. "Habitable" means not only that liquid water was present, but that life's requirements for energy and elemental raw materials could also have been met. The strongest evidence of an ancient habitable environment came from analyzing a very fine-grained rock called a mudstone—a rock type that is widespread on Earth but was unknown on Mars until *Curiosity* found it (see Figure 30.11). The mudstone can tell us a great deal about the wet environments in which they formed. The *Perseverance* rover is collecting samples of sedimentary rock in a former lakebed, to later return to Earth for laboratory analysis.



Figure 30.11 : Mudstone. Shown are the first holes drilled by NASA's *Curiosity* Mars rover into a mudstone, with "fresh" drill-pilings around the holes. Notice the difference in color between the red ancient martian surface and the gray newly exposed rock powder that came from the drill holes. Each drill hole is about 0.6 inch (1.6 cm) in diameter. (credit: modification of work by NASA/JPL-Caltech/MSSS)

Five decades of robotic exploration have allowed us to develop a picture of how Mars evolved through time. Early Mars had epochs of warmer and wetter conditions that would have been conducive to life at the surface. However, Mars eventually lost much of its early atmosphere and the surface water began to dry up. As that happened, the ever-shrinking reservoirs of liquid water on the martian surface became saltier and more acidic, until the surface finally had no significant liquid water and was bathed in harsh solar radiation. The surface thus became uninhabitable, but this might not be the case for the planet overall.

Reservoirs of ice and liquid water could still exist underground, where pressure and temperature conditions make it stable. There is recent evidence to suggest that liquid water (probably very salty water) can occasionally (and briefly) flow on the surface even today. Thus, Mars might even have habitable conditions in the present day, but of a much different sort than we normally think of on Earth.

Our study of Mars reveals a planet with a fascinating history—one that saw its ability to host surface life dwindle billions of years ago, but perhaps allowing life to adapt and survive in favorable environmental niches. Even if life did not survive, we expect that we might find evidence of life if it ever took hold on Mars. If it is there, it is hidden in the crust, and we are still learning how best to decipher that evidence.

Life in the Outer Solar System

The massive gas and ice giant planets of the outer solar system—Jupiter, Saturn, Uranus, and Neptune—are almost certainly not habitable for life as we know it, but some of their moons might be (see Figure 30.12). Although these worlds in the outer solar system contain abundant water, they receive so little warming sunlight in their distant orbits that it was long believed they would be "geologically dead" balls of hard-frozen ice and rock. But, as we saw in the chapter on Rings, Moons, and Pluto, missions to the outer solar system have found something much more interesting.

Jupiter's moon Europa revealed itself to the Voyager and Galileo missions as an active world whose icy surface apparently conceals an ocean with a depth of tens to perhaps a hundred kilometers. As the moon orbits Jupiter, the planet's massive gravity creates tides on Europa—just as our own Moon's gravity creates our ocean tides—and the friction of all that pushing and pulling generates enough heat to keep the water in liquid form (Figure 30.13). Similar tides act upon other moons if they orbit close to the planet. Scientists now think that six or more of the outer solar system's icy moons may harbor liquid water oceans for the same reason. Among these, Europa and Enceladus, a moon of Saturn, have thus far been of greatest interest to astrobiologists.

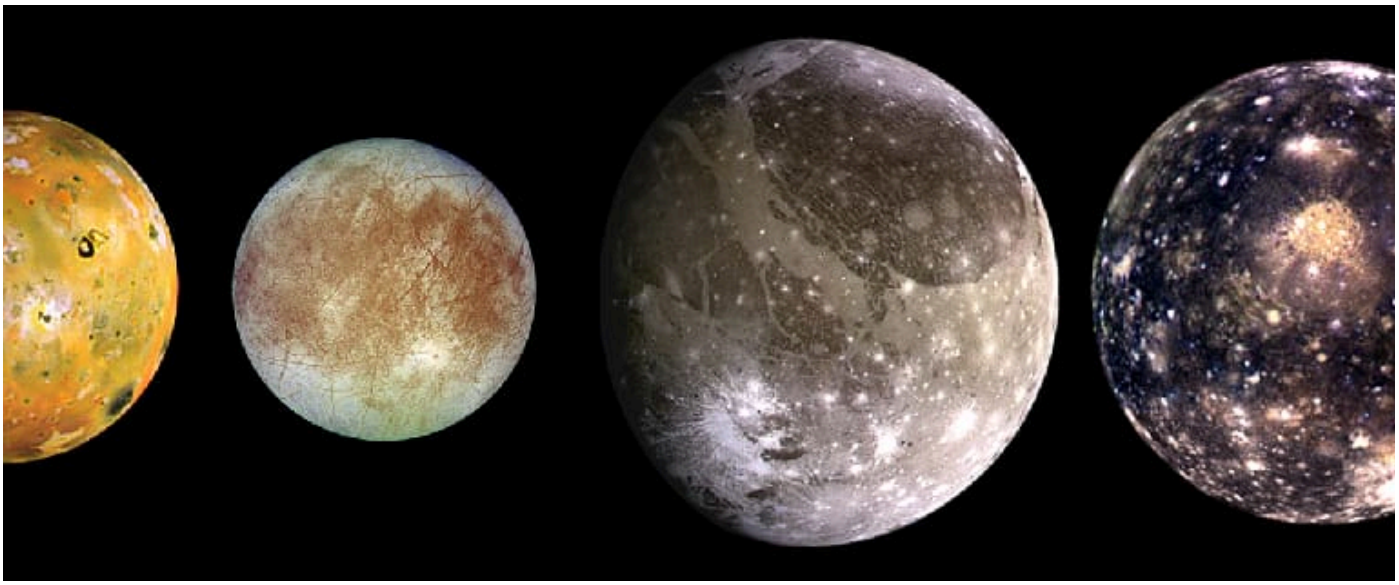


Figure 30.12 : Jupiter's Moons. The Galilean moons of Jupiter are shown to relative scale and arranged in order of their orbital distance from Jupiter. At far left, Io orbits closest to Jupiter and so experiences the strongest tidal heating by Jupiter's massive gravity. This effect is so strong that Io is thought to be the most volcanically active body in our solar system. At far right, Callisto shows a surface scarred by billions of years' worth of craters—an indication that the moon's surface is old and that Callisto may be far less active than its sibling moons. Between these hot and cold extremes, Europa, second from left, orbits at a distance where Jupiter's tidal heating may be "just right" to sustain a liquid water ocean beneath its icy crust. (credit: modification of work by NASA/JPL/DLR)

Europa has probably had an ocean for most or all of its history, but habitability requires more than just liquid water. Life also requires energy, and because sunlight does not penetrate below the kilometers-thick ice crust of Europa, this would have to be chemical energy. One of Europa's key attributes from an astrobiology perspective is that its ocean is most likely in direct contact with an underlying rocky mantle, and the interaction of water and rocks—especially at high temperatures, as within Earth's hydrothermal vent systems—yields a *reducing chemistry* (where molecules tend to give up electrons readily) that is like one half of a chemical battery. To complete the battery and provide energy that could be used by life requires that an *oxidizing chemistry* (where molecules tend to accept electrons readily) also be available. On Earth, when chemically reducing vent fluids meet oxygen-containing seawater, the energy that becomes available often supports thriving communities of microorganisms and animals on the sea floor, far from the light of the Sun.

The Galileo mission found that Europa's icy surface does contain an abundance of oxidizing chemicals. This means that availability of energy to support life depends very much on whether the chemistry of the surface and the ocean can mix, despite the kilometers of ice in between. That Europa's ice crust appears geologically "young" (only tens of millions of years old, on average) and that it is active makes it tantalizing to think that such mixing might indeed occur. Understanding whether and how much exchange occurs between the surface and ocean of Europa will be a key science objective of future missions to Europa, and a major step forward in understanding whether this moon could be a cradle of life.



Figure 30.13 : Jupiter's Moon Europa, as Imaged by NASA's Galileo Mission. The relative scarcity of craters on Europa suggests a surface that is "geologically young," and the network of colored ridges and cracks suggests constant activity and motion. Galileo's instruments also strongly suggested the presence of a massive ocean of salty liquid water beneath the icy crust. (credit: modification of work by NASA/JPL-Caltech/ SETI Institute)

In 2005, the Cassini mission performed a close flyby of a small (500-kilometer diameter) moon of Saturn, Enceladus (Figure 30.14), and made a remarkable discovery. Plumes of gas and icy material were venting from the moon's south polar region at a collective rate of about 250 kilograms of material per second. Several observations, including the discovery of salts associated with the icy material, suggest that their source is a liquid water ocean beneath tens of kilometers of ice. Although it remains to be shown definitively whether the ocean is local or global, transient or long-lived, it does appear to be in contact, and to have reacted, with a rocky interior. As on Europa, this is probably a necessary—though not sufficient—condition for habitability. What makes Enceladus so enticing to planetary scientists, though, are those plumes of material that seem to come directly from its ocean: samples of the interior are there for the taking by any spacecraft sent flying through. For a future mission, such samples could yield evidence not only of whether Enceladus is habitable but, indeed, of whether it is home to life.

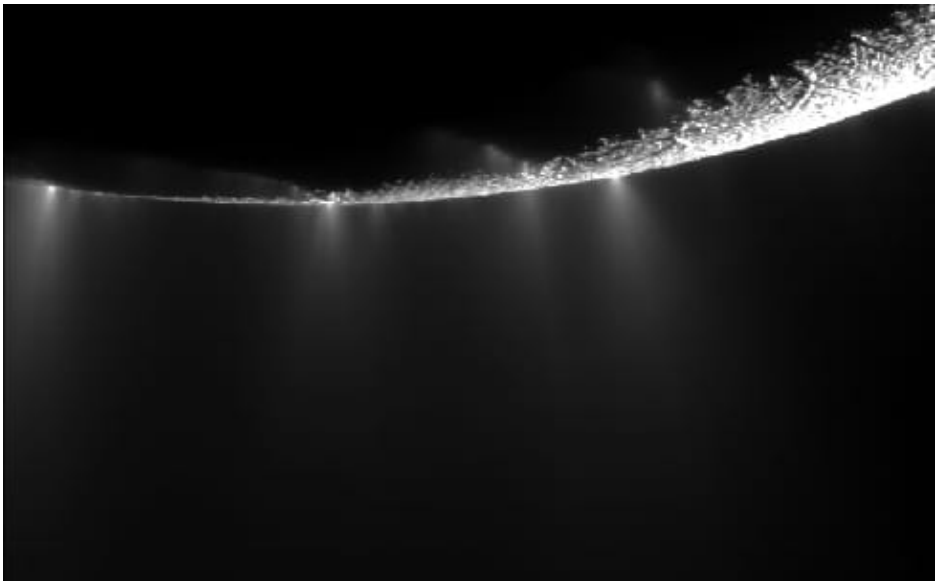


Figure 30.14 : Image of Saturn's Moon Enceladus from NASA's Cassini Mission. The south polar region was found to have multiple plumes of ice and gas that, combined, are venting about 250 kilograms of material per second into space. Such features suggest that Enceladus, like Europa, has a sub-ice ocean. (credit: NASA/JPL/ SSI)

Saturn's big moon Titan is very different from both Enceladus and Europa (see Figure 30.15). Although it may host a liquid water layer deep within its interior, it is the surface of Titan and its unusual chemistry that makes this moon such an interesting place.

Titan's thick atmosphere—the only one among moons in the solar system—is composed mostly of nitrogen but also of about 5% methane. In the upper atmosphere, the Sun's ultraviolet light breaks apart and recombines these molecules into more complex organic compounds that are collectively known as *tholins*. The tholins shroud Titan in an orange haze, and imagery from Cassini and from the Huygens probe that descended to Titan's surface show that heavier particles appear to accumulate on the surface, even forming "dunes" that are cut and sculpted by flows of liquid hydrocarbons (such as liquid methane). Some scientists see this organic chemical factory as a natural laboratory that may yield some clues about the solar system's early chemistry—perhaps even chemistry that could support the origin of life.

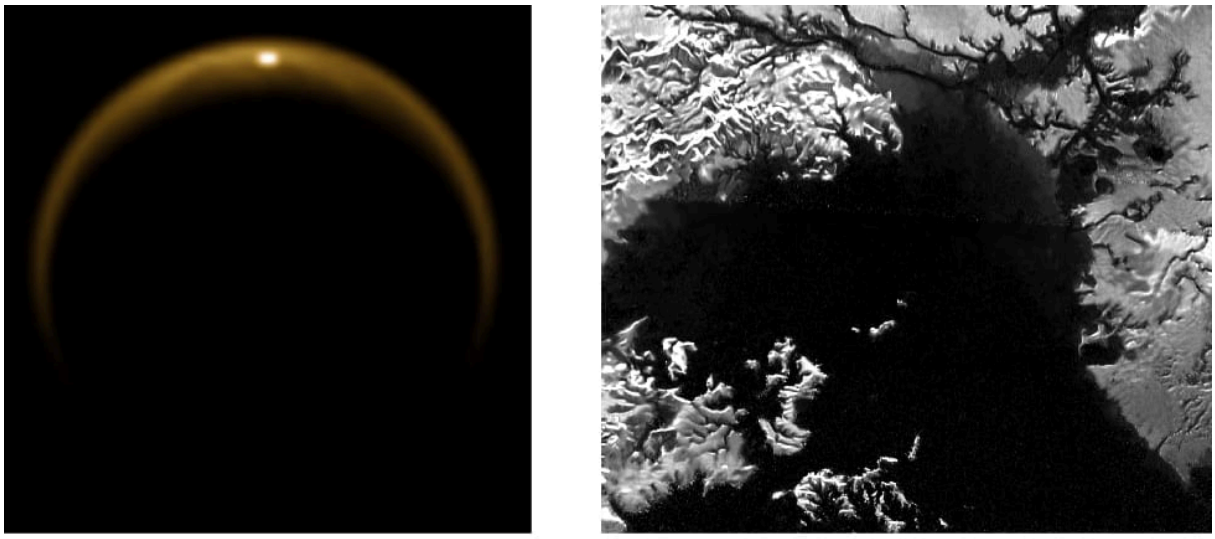


Figure 30.15 : Image of Saturn's Moon Titan from NASA's Cassini Mission. (a) The hazy orange glow comes from Titan's thick atmosphere (the only one known among the moons of the solar system). That atmosphere is mostly nitrogen but also contains methane and potentially a variety of complex organic compounds. The bright spot near the top of the image is sunlight reflected from a very flat surface—almost certainly a liquid. We see this effect, called "glint," when sunlight reflects off the surface of a lake or ocean. (b) Cassini radar imagery shows what look very much like landforms and lakes on the surface of Titan. But the surface lakes and oceans of Titan are not water; they are probably made of liquid hydrocarbons like methane and ethane. (credit a: modification of work by NASA/JPL/University of Arizona/DLR; credit b: modification of work by NASA/JPL-Caltech/ASI)

Link to Learning

In January 2005, the Huygens probe descended to the surface of Titan and relayed data, including imagery of the landing site, for about 90 minutes. You can watch a video about the descent of Huygens to Titan's surface.

Habitable Planets Orbiting Other Stars

One of the most exciting developments in astronomy during the last two decades is the ability to detect exoplanets—planets orbiting other stars. As we saw in the chapter on the formation of stars and planets, since the discovery of the first exoplanet in 1995, there have been thousands of confirmed detections, and many more candidates that are not yet confirmed. These include several dozen possibly habitable exoplanets. Such numbers finally allow us to make some predictions about exoplanets and their life-hosting potential. The majority of stars with mass similar to the Sun appear to host at least one planet, with multi-planet systems like our own not unusual. How many of these planets might be habitable, and how could we search for life there?

Link to Learning

The NASA Exoplanet Archive is an up-to-date searchable online source of data and tools on everything to do with exoplanets. Explore stellar and exoplanet parameters and characteristics, find the latest news on exoplanet discoveries, plot your own data interactively, and link to other related resources.

In evaluating the prospect for life in distant planetary systems, astrobiologists have developed the idea of a habitable zone—a region around a star where suitable conditions might exist for life. This concept focuses on life's requirement for liquid water, and the habitable zone is generally thought of as the range of distances from the central star in which water could be present in liquid form at a planet's surface. In our own solar system, for example, Venus has surface temperatures far above the boiling point of water and Mars has surface temperatures that are almost always below the freezing point of water. Earth, which orbits between the two, has a surface temperature that is "just right" to keep much of our surface water in liquid form.

Whether surface temperatures are suitable for maintaining liquid water depends on a planet's "radiation budget" —how much starlight energy it absorbs and retains—and whether or how processes like winds and ocean circulation distribute that energy around the planet. How much stellar energy a planet receives, in turn, depends on how much and what sort of light the star emits and how far the planet is from that star,[1] how much it reflects back to space, and how effectively the planet's atmosphere can retain heat through the greenhouse effect (see Earth as a Planet). All of these can vary substantially, and all matter a lot. For example, Venus receives about twice as much starlight per square meter as Earth but, because of its dense cloud cover, also reflects about twice as much of that light back to space as Earth does. Mars receives only about half as much starlight as Earth, but also reflects only about half as much. Thus, despite their differing orbital distances, the three planets actually absorb comparable amounts of sunlight energy. Why, then, are they so dramatically different?

As we learned in several chapters about the planets, some of the gases that make up planetary atmospheres are very effective at trapping infrared light—the very range of wavelengths at which planets radiate thermal energy back out to space—and this can raise the planet's surface temperature quite a bit more than would otherwise be the case. This is the same "greenhouse effect" that is of such concern for global warming on our planet. Earth's natural greenhouse effect, which comes mostly from water vapor and carbon dioxide in the atmosphere, raises our average surface temperature by about 33 °C over the value it would have if there were no greenhouse gases in the atmosphere. Mars has a very thin atmosphere and thus very little greenhouse warming (about 2 °C worth), while Venus has a massive carbon dioxide atmosphere that creates very strong greenhouse warming (about 510 °C worth). These worlds are much colder and much hotter, respectively, than Earth would be if moved into their orbits. Thus, we must consider the nature of any atmosphere as well as the distance from the star in evaluating the range of habitability.

Of course, as we have learned, stars also vary widely in the intensity and spectrum (the wavelengths of light) they emit. Some are much brighter and hotter (bluer), while others are significantly dimmer and cooler (redder), and the distance of the habitable zone varies accordingly. For example, the habitable zone around M-dwarf stars is 3 to 30 times closer in than for G-type (Sun-like) stars. There is a lot of interest in whether such systems could be habitable because—although they have some potential downsides for supporting life—M-dwarf stars are by far the most numerous and long-lived in our Galaxy.

The luminosity of stars like the Sun also increases over their main-sequence lifetime, and this means that the habitable zone migrates outward as a star system ages. Calculations indicate that the power output of the Sun, for example, has increased by at least 30% over the past 4 billion years. Thus, Venus was once within the habitable zone, while Earth received a level of solar energy insufficient to keep the modern Earth (with its present atmosphere) from freezing over. In spite of this, there is plenty of geological evidence that liquid water was present on Earth's surface billions of years ago. The phenomenon of increasing stellar output and an outwardly migrating habitable zone has led to another concept: the *continuously* habitable zone is defined by the range of orbits that would remain within the habitable zone during the entire lifetime of the star system. As you might imagine, the continuously habitable zone is quite a bit narrower than the habitable zone is at any one time in a star's history. The nearest star to the Sun, Proxima Centauri, is an M star that has a planet with a mass of at least 1.3 Earth masses, taking about 11 days to orbit. At the distance for such a quick orbit (0.05 AU), the planet may be in the habitable zone of its star, although whether conditions on such a planet near such a star are hospitable for life is a matter of great scientific debate.

Even when planets orbit within the habitable zone of their star, it is no guarantee that they are habitable. For example, Venus today has virtually no water, so even if it were suddenly moved to a "just right" orbit within the habitable zone, a critical requirement for life would still be lacking.

Scientists are working to understand all the factors that define the habitable zone and the habitability of planets orbiting within that zone because this will be our primary guide in targeting exoplanets on which to seek evidence of life. As technology for detecting exoplanets has advanced, so too has our potential to find Earth-size worlds within the habitable zones of their parent stars. Of the confirmed or candidate exoplanets known at the time of writing, nearly 300 are considered to be orbiting within the habitable zone and more than 10% of those are roughly Earth-size.

Link to Learning

Explore the habitable universe at the online Planetary Habitability Laboratory created by the University of Puerto Rico at Arecibo. See the potentially habitable exoplanets and other interesting places in the universe, watch video clips, and link to numerous related resources on astrobiology.

Biomarkers

Our observations suggest increasingly that Earth-size planets orbiting within the habitable zone may be common in the Galaxy—current estimates suggest that more than 40% of stars have at least one. But are any of them inhabited? With no ability to send probes there to sample, we will have to derive the answer from the light and other radiation that come to us from these faraway systems (Figure 30.16). What types of observations might constitute good evidence for life?

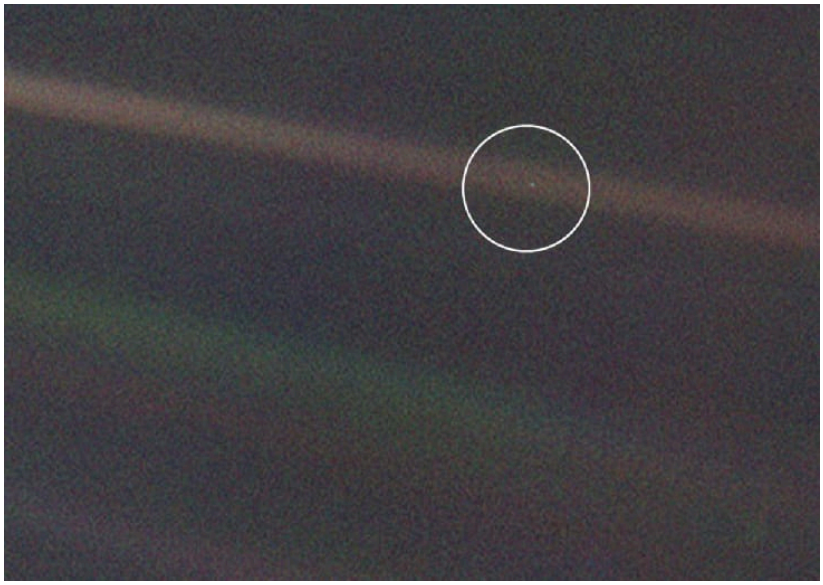


Figure 30.16 : Earth, as Seen by NASA's Voyager 1. In this image, taken from 4 billion miles away, Earth appears as a "pale blue dot" representing less than a pixel's worth of light. Would this light reveal Earth as a habitable and inhabited world? Our search for life on exoplanets will depend on an ability to extract information about life from the faint light of faraway worlds. (credit: modification of work by NASA/JPL-Caltech)

To be sure, we need to look for robust biospheres (atmospheres, surfaces, and/or oceans) capable of creating planet-scale change. Earth hosts such a biosphere: the composition of our atmosphere and the spectrum of light reflected from our planet differ considerably from what would be expected in the absence of life. Presently, Earth is the only body in our solar system for which this is true, despite the possibility that habitable conditions might prevail in the subsurface of Mars or inside the icy moons of the outer solar system. Even if life exists on these worlds, it is very unlikely that it could yield planet-scale changes that are both telescopically observable and clearly biological in origin.

What makes Earth "special" among the potentially habitable worlds in our solar system is that it has a photosynthetic biosphere. This requires the presence of liquid water at the planet's surface, where organisms have direct access to sunlight. The habitable zone concept focuses on this requirement for surface liquid water—even though we know that subsurface habitable conditions could prevail at more distant orbits—exactly because these worlds would have biospheres detectable at a distance.

Indeed, plants and photosynthetic microorganisms are so abundant at Earth's surface that they affect the color of the light that our planet reflects out into space—we appear greener in visible wavelengths and reflect more near-infrared light than we otherwise would. Moreover, photosynthesis has changed Earth's atmosphere at a large scale—more than 20% of our atmosphere comes from the photosynthetic waste product, oxygen. Such high levels would be very difficult to explain in the absence of life. Other gases, such as nitrous oxide and methane, when found simultaneously with oxygen, have also been suggested as possible indicators of life. When sufficiently abundant in an atmosphere, such gases could be detected by their effect on the spectrum of light that a planet emits or reflects. (As we saw in the chapter on exoplanets, astronomers today are beginning to have the

capability of detecting the spectrum of the atmospheres of some planets orbiting other stars.)

Astronomers have thus concluded that, at least initially, a search for life outside our solar system should focus on exoplanets that are as much like Earth as possible—roughly Earth-size planets orbiting in the habitable zone—and look for the presence of gases in the atmosphere or colors in the visible spectrum that are hard to explain except by the presence of biology. Simple, right? In reality, the search for exoplanet life poses many challenges.

As you might imagine, this task is more challenging for planetary systems that are farther away and, in practical terms, this will limit our search to the habitable worlds closest to our own. Should we become limited to a very small number of nearby targets, it will also become important to consider the habitability of planets orbiting the M-dwarfs we discussed above.

If we manage to separate out a clean signal from the planet and find some features in the light spectrum that might be indicative of life, we will need to work hard to think of any nonbiological process that might account for them. “Life is the hypothesis of last resort,” noted astronomer Carl Sagan—meaning that we must exhaust all other explanations for what we see before claiming to have found evidence of extraterrestrial biology. This requires some understanding of what processes might operate on worlds that we will know relatively little about; what we find on Earth can serve as a guide but also has potential to lead us astray (Figure 30.17).

Recall, for example, that it would be extremely difficult to account for the abundance of oxygen in Earth’s atmosphere except by the presence of biology. But it has been hypothesized that oxygen could build up to substantial levels on planets orbiting M-dwarf stars through the action of ultraviolet radiation on the atmosphere—with no need for biology. It will be critical to understand where such “false positives” might exist in carrying out our search.

We need to understand that we might not be able to detect biospheres even if they exist. Life has flourished on Earth for perhaps 3.5 billion years, but the atmospheric “biosignatures” that, today, would supply good evidence for life to distant astronomers have not been present for all of that time. Oxygen, for example, accumulated to detectable levels in our atmosphere only a little over 2 billion years ago. Could life on Earth have been detected before that time? Scientists are working actively to understand what additional features might have provided evidence of life on Earth during that early history, and thereby help our chances of finding life beyond.

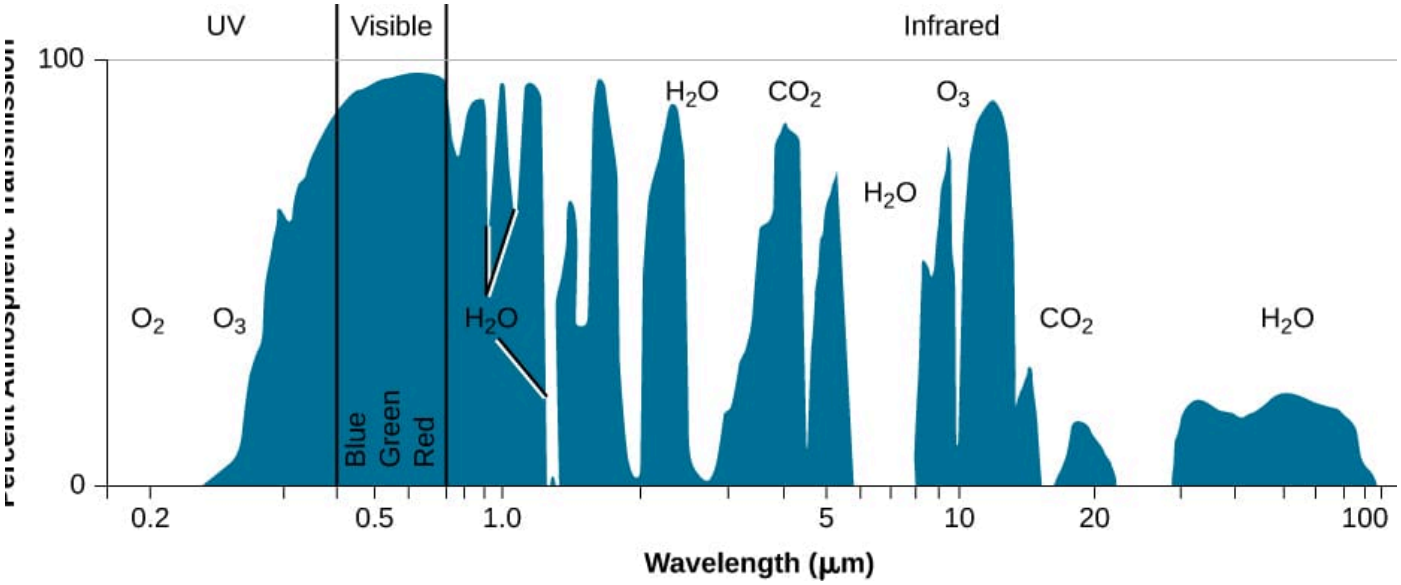


Figure 30.17 : Spectrum of Light Transmitted through Earth’s Atmosphere. This graph shows wavelengths ranging from ultraviolet (far left) to infrared. The many downward “spikes” come from absorption of particular wavelengths by molecules in Earth’s atmosphere. Some of these compounds, like water and the combination oxygen/ozone and methane, might reveal Earth as both habitable and inhabited. We will have to rely on this sort of information to seek life on exoplanets, but our spectra will be of much poorer quality than this one, in part because we will receive so little light from the planet. (credit: modification of work by NASA)

Footnotes

- 1The amount of starlight received per unit area of a planet’s surface (per square meter, for example) decreases with the

square of the distance from the star. Thus, when the orbital distance doubles, the illumination decreases by 4 times (2^2), and when the orbital distance increases tenfold, the illumination decreases by 100 times (10^2). Venus and Mars orbit the sun at about 72% and 152% of Earth's orbital distance, respectively, so Venus receives about $1/(0.72)^2 = 1.92$ (about twice) and Mars about $1/(1.52)^2 = 0.43$ (about half) as much light per square meter of planet surface as Earth does.

This page titled 19.3: Searching for Life beyond Earth is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 30.4: Searching for Life beyond Earth by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

19.4: The Search for Extraterrestrial Intelligence

Learning Objectives

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

- Explain why spaceships from extraterrestrial civilizations are unlikely to have visited us
- List efforts by humankind to communicate with other civilizations via messages on spacecraft
- Understand the various SETI programs scientists are undertaking

Given all the developments discussed in this chapter, it seems likely that life could have developed on many planets around other stars. Even if that life is microbial, we saw that we may soon have ways to search for chemical biosignatures. This search is of fundamental importance for understanding biology, but it does not answer the question, “Are we alone?” that we raised at the beginning of this chapter. When we ask this question, many people think of other intelligent creatures, perhaps beings that have developed technology similar to our own. If any intelligent, technical civilizations have arisen, as has happened on Earth in the most recent blink of cosmic time, how could we make contact with them?

This problem is similar to making contact with people who live in a remote part of Earth. If students in the United States want to converse with students in Australia, for example, they have two choices. Either one group gets on an airplane and travels to meet the other, or they communicate by sending a message remotely. Given how expensive airline tickets are, most students would probably select the message route.

In the same way, if we want to get in touch with intelligent life around other stars, we can travel, or we can try to exchange messages. Because of the great distances involved, interstellar space travel would be very slow and prohibitively expensive. The fastest spacecraft the human species has built so far would take almost 80,000 years to get to the nearest star. While we could certainly design a faster craft, the more quickly we require it to travel, the greater the energy cost involved. To reach neighboring stars in less than a human life span, we would have to travel close to the speed of light. In that case, however, the expense would become truly astronomical.

Interstellar Travel

Bernard Oliver, an engineer with an abiding interest in life elsewhere, made a revealing calculation about the costs of rapid interstellar space travel. Since we do not know what sort of technology we (or other civilizations) might someday develop, Oliver considered a trip to the nearest star (and back again) in a spaceship with a “perfect engine”—one that would convert its fuel into energy with 100% efficiency. Even with a perfect engine, the energy cost of a single round-trip journey at 70% the speed of light turns out to be equivalent to several hundred thousand years’ worth of total U.S. electrical energy consumption. The cost of such travel is literally out of this world.

This is one reason astronomers are so skeptical about claims that UFOs are spaceships from extraterrestrial civilizations. Given the distance and energy expense involved, it seems unlikely that the dozens of UFOs (and even UFO abductions) claimed each year could be visitors from other stars so fascinated by Earth civilization that they are willing to expend fantastically large amounts of energy or time to reach us. Nor does it seem credible that these visitors have made this long and expensive journey and then systematically avoided contacting our governments or political and intellectual leaders.

Not every UFO report has been explained (in many cases, the observations are sketchy or contradictory). But investigation almost always converts them to IFOs (identified flying objects) or NFOs (not-at-all flying objects). While some are hoaxes, others are natural phenomena, such as bright planets, ball lightning, fireballs (bright meteors), or even flocks of birds that landed in an oil slick to make their bellies reflective. Still others are human craft, such as private planes with some lights missing, or secret military aircraft. It is also interesting that the group of people who most avidly look at the night sky, the amateur astronomers, have never reported UFO sightings. Further, not a single UFO has ever left behind any physical evidence that can be tested in a laboratory and shown to be of nonterrestrial origin.

Another common aspect of belief that aliens are visiting Earth comes from people who have difficulty accepting human accomplishments. There are many books and TV shows, for example, that assert that humans could not have built the great

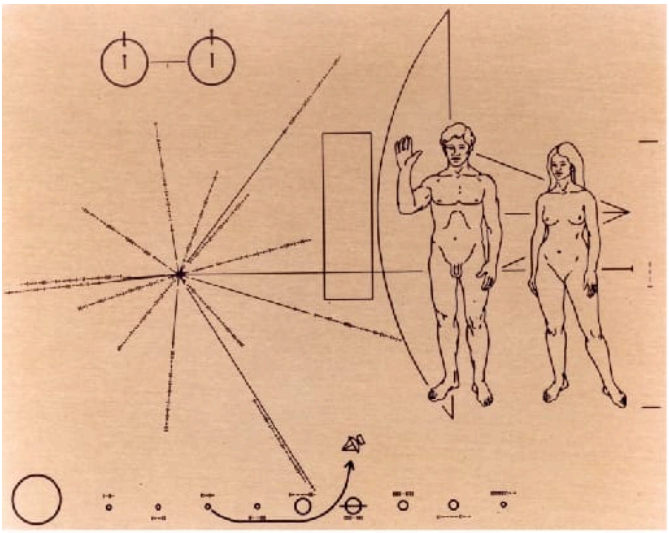
pyramids of Egypt, and therefore they must have been built by aliens. The huge statues (called Moai) on Easter Island are also sometimes claimed to have been built by aliens. Some people even think that the accomplishments of space exploration today are based on alien technology.

However, the evidence from archaeology and history is clear: ancient monuments were built by ancient *people,* whose brains and ingenuity were every bit as capable as ours are today, even if they didn't have electronic textbooks like you do.

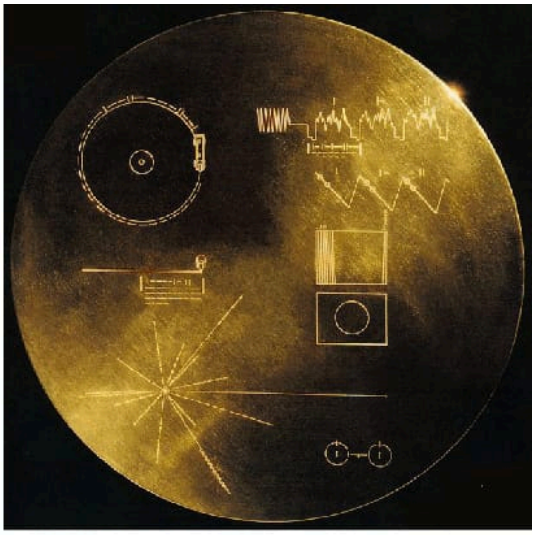
Messages on Spacecraft

While space travel by living creatures seems very difficult, robot probes can travel over long distances and over long periods of time. Five spacecraft—two Pioneers, two Voyagers, and New Horizons—are now leaving the solar system. At their coasting speeds, they will take hundreds of thousands or millions of years to get anywhere close to another star. On the other hand, they were the first products of human technology to go beyond our home system, so we wanted to put messages on board to show where they came from.

Each Pioneer carries a plaque with a pictorial message engraved on a gold-anodized aluminum plate (Figure 30.18). The Voyagers, launched in 1977, have audio and video records attached, which allowed the inclusion of over 100 photographs and a selection of music from around the world. Given the enormous space between stars in our section of the Galaxy, it is very unlikely that these messages will ever be received by anyone. They are more like a note in a bottle thrown into the sea by a shipwrecked sailor, with no realistic expectation of its being found soon but a slim hope that perhaps someday, somehow, someone will know of the sender's fate.



(a)



(b)

Figure 30.18 : Interstellar Messages. (a) This is the image engraved on the plaques aboard the Pioneer 10 and 11 spacecraft. The human figures are drawn in proportion to the spacecraft, which is shown behind them. The Sun and planets in the solar system can be seen at the bottom, with the trajectory that the spacecraft followed. The lines and markings in the left center show the positions and pulse periods for a number of pulsars, which might help locate the spacecraft's origins in space and time. (b) Encoded onto a gold-coated copper disk, the Voyager record contains 118 photographs, 90 minutes of music from around the world, greetings in almost 60 languages, and other audio material. It is a summary of the sights and sounds of Earth. (credit a, b: modification of work by NASA)

Making Connections: The Voyager Message

An Excerpt from the Voyager Record:

"We cast this message into the cosmos. It is likely to survive a billion years into our future, when our civilization is profoundly altered. . . . If [another] civilization intercepts Voyager and can understand these recorded contents, here is our message:

> This is a present from a small, distant world, a token of our sounds, our science, our images, our music, our thoughts, and our feelings. We are attempting to survive our time so we may live into yours. We hope, someday, having solved the problems we face, to join a community of galactic civilizations. This record represents our hope and our determination, and our goodwill in a vast and awesome universe."
>
> —Jimmy Carter, President of the United States of America, June 16, 1977

Communicating with the Stars

If direct visits among stars are unlikely, we must turn to the alternative for making contact: exchanging messages. Here the news is a lot better. We already use a messenger—light or, more generally, electromagnetic waves—that moves through space at the fastest speed in the universe. Traveling at 300,000 kilometers per second, light reaches the nearest star in only 4 years and does so at a tiny fraction of the cost of sending material objects. These advantages are so clear and obvious that we assume they will occur to any other species of intelligent beings that develop technology.

However, we have access to a wide spectrum of electromagnetic radiation, ranging from the longest-wavelength radio waves to the shortest-wavelength gamma rays. Which would be the best for interstellar communication? It would not be smart to select a wavelength that is easily absorbed by interstellar gas and dust, or one that is unlikely to penetrate the atmosphere of a planet like ours. Nor would we want to pick a wavelength that has lots of competition for attention in our neighborhood.

One final criterion makes the selection easier: we want the radiation to be inexpensive enough to produce in large quantities. When we consider all these requirements, radio waves turn out to be the best answer. Being the lowest-frequency (and lowest-energy) band of the spectrum, they are not very expensive to produce, and we already use them extensively for communications on Earth. They are not significantly absorbed by interstellar dust and gas. With some exceptions, they easily pass through Earth's atmosphere and through the atmospheres of the other planets we are acquainted with.

The Cosmic Haystack

Having made the decision that radio is the most likely means of communication among intelligent civilizations, we still have many questions and a daunting task ahead of us. Shall we *send* a message, or try to *receive* one? Obviously, if every civilization decides to receive only, then no one will be sending, and everyone will be disappointed. On the other hand, it may be appropriate for us to *begin* by listening, since we are likely to be among the most primitive civilizations in the Galaxy who are interested in exchanging messages.

We do not make this statement to insult the human species (which, with certain exceptions, we are rather fond of). Instead, we base it on the fact that humans have had the ability to receive (or send) a radio message across interstellar distances for only a few decades. Compared to the ages of the stars and the Galaxy, this is a mere instant. If there are civilizations out there that are ahead of us in development by even a short time (in the cosmic sense), they are likely to have a technology head start of many, many years.

In other words, we, who have just started, may well be the "youngest" species in the Galaxy with this capability (see the discussion in Example 30.1). Just as the youngest members of a community are often told to be quiet and listen to their elders for a while before they say something foolish, so may we want to begin our exercise in extraterrestrial communication by listening.

Even restricting our activities to listening, however, leaves us with an array of challenging questions. For example, if an extraterrestrial civilization's signal is too weak to be detected by our present-day radio telescopes, we will not detect them. In addition, it would be very expensive for an extraterrestrial civilization to broadcast on a huge number of channels. Most likely, they select one or a few channels for their particular message. Communicating on a narrow band of channels also helps distinguish an artificial message from the radio static that comes from natural cosmic processes. But the radio band contains an astronomically large number of possible channels. How can we know in advance which one they have selected, and how they have coded their message into the signal?

Table 30.1 summarizes these and other factors that scientists must grapple with when trying to tune in to radio messages from distant civilizations. Because their success depends on either guessing right about so many factors or searching through all the possibilities for each factor, some scientists have compared their quest to looking for a needle in a haystack. Thus, they like to say that the list of factors in Table 30.1 defines the *cosmic haystack problem*.

Table 30.1: The Cosmic Haystack Problem: Some Questions about an Extraterrestrial Message

Factors
From which direction (which star) is the message coming?
On what channels (or frequencies) is the message being broadcast?
How wide in frequency is the channel?
How strong is the signal (can our radio telescopes detect it)?
Is the signal continuous, or does it shut off at times (as, for example, a lighthouse beam does when it turns away from us)?
Does the signal drift (change) in frequency because of the changing relative motion of the source and the receiver?
How is the message encoded in the signal (how do we decipher it)?
Can we even recognize a message from a completely alien species? Might it take a form we don't at all expect?

Radio Searches

Although the cosmic haystack problem seems daunting, many other research problems in astronomy also require a large investment of time, equipment, and patient effort. And, of course, if we don't search, we're sure not to find anything.

The very first search was conducted by astronomer Frank Drake in 1960, using the 85-foot antenna at the National Radio Astronomy Observatory (Figure 30.19). Called Project Ozma, after the queen of the exotic Land of Oz in the children's stories of L. Frank Baum, his experiment involved looking at about 7200 channels and two nearby stars over a period of 200 hours. Although he found nothing, Drake demonstrated that we had the technology to do such a search, and set the stage for the more sophisticated projects that followed.



(a) (b)

Figure 30.19 : Project Ozma and the Allen Telescope Array. (a) This 25th anniversary photo shows some members of the Project Ozma team standing in front of the 85-foot radio telescope with which the 1960 search for extraterrestrial messages was performed. Frank Drake is in the back row, second from the right. (b) The Allen Telescope Array in California is made up of 42 small antennas linked together. This system allows simultaneous observations of multiple sources with millions of separate frequency channels. (credit a: modification of work by NRAO; credit b: modification of work by Colby Gutierrez-Kraybill)

Receivers are constantly improving, and the sensitivity of SETI programs—SETI stands for the search for extraterrestrial intelligence—is advancing rapidly. Equally important, modern electronics and software allow simultaneous searches on millions of frequencies (channels). If we can thus cover a broad frequency range, the cosmic haystack problem of guessing the right frequency largely goes away. One powerful telescope array (funded with an initial contribution from Microsoft founder Paul Allen) that is built for SETI searches is the Allen Telescope in Northern California. Other radio telescopes being used for such searches have included the giant Arecibo radio dish in Puerto Rico (which, sadly, collapsed due to storm damage in late 2020), the recently

completed, and even larger, FAST dish in China, and the Green Bank Telescope in West Virginia, which is the largest steerable radio telescope in the world.

In 2015, Silicon Valley investor and philanthropist Yuri Milner donated 100 million dollars to the University of California, Berkeley, to enhance the SETI endeavor for ten years. The project, called B*reakthrough: Listen*, is enlisting more telescopes around the world in the search and is developing sophisticated artificial intelligence programs to scour the incoming signals for intelligent messages. The aim is to search more stars and more channels than any SETI project so far.

What kind of signals do we hope to pick up? We on Earth are inadvertently sending out a flood of radio signals, dominated by military radar systems. This is a kind of leakage signal, similar to the wasted light energy that is beamed upward by poorly designed streetlights and advertising signs. Could we detect a similar leakage of radio signals from another civilization? The answer is just barely, but only for the nearest stars. For the most part, therefore, current radio SETI searches are looking for beacons, assuming that civilizations might be intentionally drawing attention to themselves or perhaps sending a message to another world or outpost that lies in our direction. Our prospects for success depend on how often civilizations arise, how long they last, and how patient they are about broadcasting their locations to the cosmos.

Voyagers in Astronomy: Jill Tarter: Trying to Make Contact

1997 was quite a year for Jill Cornell Tarter (Figure 30.20), one of the world's leading scientists in the SETI field. The SETI Institute announced that she would be the recipient of its first endowed chair (the equivalent of an endowed research professorship) named in honor of Bernard Oliver. The National Science Foundation approved a proposal by a group of scientists and educators she headed to develop an innovative hands-on high school curriculum based on the ideas of cosmic evolution (the topics of this chapter). And, at roughly the same time, she was being besieged with requests for media interviews as news reports identified her as the model for Ellie Arroway, the protagonist of *Contact*, Carl Sagan's best-selling novel about SETI. The book had been made into a high-budget science fiction film, starring Jodie Foster, who had talked with Tarter before taking the role.



Figure 30.20 Jill Tarter (credit: Christian Schidlowski)

Tarter is quick to point out, "Carl Sagan wrote a book about a woman who does what I do, not about me." Still, as the only woman in such a senior position in the small field of SETI, she was the center of a great deal of public attention. (However, colleagues and reporters pointed out that this was nothing compared to what would happen if her search for radio signals from other civilizations recorded a success.)

Being the only woman in a group is not a new situation to Tarter, who often found herself the only woman in her advanced science or math classes. Her father had encouraged her, both in her interest in science and her "tinkering." As an undergraduate at Cornell University, she majored in engineering physics. That training became key to putting together and maintaining the complex systems that automatically scan for signals from other civilizations.

Switching to astrophysics for her graduate studies, she wrote a PhD thesis that, among other topics, considered the formation of failed stars—those whose mass was not sufficient to ignite the nuclear reactions that power more massive stars like our own Sun. Tarter coined the term “brown dwarf” for these small, dim objects, and it has remained the name astronomers use ever since.

It was while she was still in graduate school that Stuart Bowyer, one of her professors at the University of California, Berkeley, asked her if she wanted to be involved in a small experiment to siphon off a bit of radiation from a radio telescope as astronomers used it year in and year out and see if there was any hint of an intelligently coded radio message buried in the radio noise. Her engineering and computer programming skills became essential to the project, and soon she was hooked on the search for life elsewhere.

Thus began an illustrious career working full time searching for what she calls the technosignatures (signs of technology) from extraterrestrial civilizations. Tarter has received many awards, including being elected fellow of the American Association for the Advancement of Science in 2002, the Adler Planetarium Women in Space Science Award in 2003, and a 2009 TED Prize.

Link to Learning

Watch the TED talk Jill Tarter gave on the fascination of the search for intelligence.

Example 30.1: The Drake Equation

At the first scientific meeting devoted to SETI, Frank Drake wrote an equation on the blackboard that took the difficult question of estimating the number of civilizations in the Galaxy and broke it down into a series of smaller, more manageable questions. Ever since then, both astronomers and students have used this Drake equation as a means of approaching the most challenging question: How likely is it that we are alone? Since this is at present an unanswerable question, astronomer Jill Tarter has called the Drake equation a “way of organizing our ignorance.” (See Figure 30.21.)



Figure 30.21 : Drake Equation. A plaque at the National Radio Astronomy Observatory commemorates the conference where the equation was first discussed. (credit: NRAO/NSF/AUI)

The form of the Drake equation is very simple. To estimate the number of communicating civilizations that currently exist in the Galaxy (we will define these terms more carefully in a moment), we multiply the rate of formation of such civilizations (number per year) by their average lifetime (in years). In symbols,

$$N = R_{\text{total}} \times L$$

To make this formula easier to use (and more interesting), however, Drake separated the rate of formation R_{total} into a series of probabilities:

$$R_{\text{total}} = R_{\text{star}} \times f_{\text{p}} \times f_{\text{e}} \times f_{\text{l}} \times f_{\text{i}} \times f_{\text{c}}$$

R_{star} is the rate of formation of stars like the Sun in our Galaxy, which is, very roughly, around 10 stars per year. Each of the other terms is a fraction or probability (less than or equal to 1.0), and the product of all these probabilities is itself the total probability that each star will have an intelligent, technological, communicating civilization that we might want to talk to. We have:

- f_{p} = the fraction of these stars with planets
- f_{e} = the fraction of the planetary systems that include habitable planets
- f_{l} = the fraction of habitable planets that actually support life
- f_{i} = the fraction of inhabited planets that develop advanced intelligence
- f_{c} = the fraction of these intelligent civilizations that develop science and the technology to build radio telescopes and transmitters

Each of these factors can be discussed and perhaps evaluated, but we must guess at many of the values. In particular, we don't know how to calculate the probability of something that happened once on Earth but has not been observed elsewhere—and these include the development of life, of intelligent life, and of technological life (the last three factors in the equation). One important advance in estimating the terms of the Drake equation comes from the recent discovery of exoplanets. When the Drake equation was first written, no one had any idea whether planets and planetary systems were common. Now we know they are—another example of the Copernican principle.

Solution

Even if we don't know the answers, we can make some guesses and calculate the resulting number *N*. Let's start with the optimism implicit in the Copernican principle and set the last three terms equal to 1.0. If *R* is 10 stars/year and if we measure the average lifetime of a technological civilization in years, the units of years cancel. If we also assume that f_{p} is 0.1, and f_{e} is 1.0, the equation becomes

$$N = R_{\text{total}} \times L = L$$

Now we see the importance of the term L, the lifetime of a communicating civilization (measured in years). We have had this capability (to communicate at the distances of the stars) for only a few decades.

Exercise 19.4.1

Suppose we assume that this stage in our history lasts only one century.

Answer available at the end of this section.

Link to Learning

Read Frank Drake's own account of how he came up with his "equation." And here is a recent interview with Frank Drake by one of the authors of this textbook.

SETI outside the Radio Realm

For the reasons discussed above, most SETI programs search for signals at radio wavelengths. But in science, if there are other approaches to answering an unsolved question, we don't want to neglect them. So astronomers have been thinking about other ways we could pick up evidence for the existence of technologically advanced civilizations.

Recently, technology has allowed astronomers to expand the search into the domain of visible light. You might think that it would be hopeless to try to detect a flash of visible light from a planet given the brilliance of the star it orbits. This is why we usually

cannot measure the reflected light of planets around other stars. The feeble light of the planet is simply swamped by the “big light” in the neighborhood. So another civilization would need a mighty strong beacon to compete with their star.

However, in recent years, human engineers have learned how to make flashes of light brighter than the Sun. The trick is to “turn on” the light for a very brief time, so that the costs are manageable. But ultra-bright, ultra-short laser pulses (operating for periods of a billionth of a second) can pack a lot of energy and can be coded to carry a message. We also have the technology to detect such short pulses—not with human senses, but with special detectors that can be “tuned” to hunt automatically for such short bursts of light from nearby stars.

Why would any civilization try to outshine its own star in this way? It turns out that the cost of sending an ultra-short laser pulse in the direction of a few promising stars can be less than the cost of sweeping a continuous radio message across the whole sky. Or perhaps they, too, have a special fondness for light messages because one of their senses evolved using light. Several programs are now experimenting with “optical SETI” searches, which can be done with only a modest telescope. (The term *optical* here means using visible light.)

If we let our imaginations expand, we might think of other for *technosignature*s—signs of technology from other civilizations. For example, what if a truly advanced civilization should decide to (or need to) renovate its planetary system to maximize the area for life? It could do so by breaking apart some planets or moons and building a ring of solid material that surrounds or encloses the star and intercepts some or all of its light. This huge artificial ring or sphere might glow very brightly at infrared wavelengths, as the starlight it receives is eventually converted to heat and re-radiated into space. That infrared radiation could be detected by our instruments, and searches for such infrared sources are also underway (Figure 30.22).

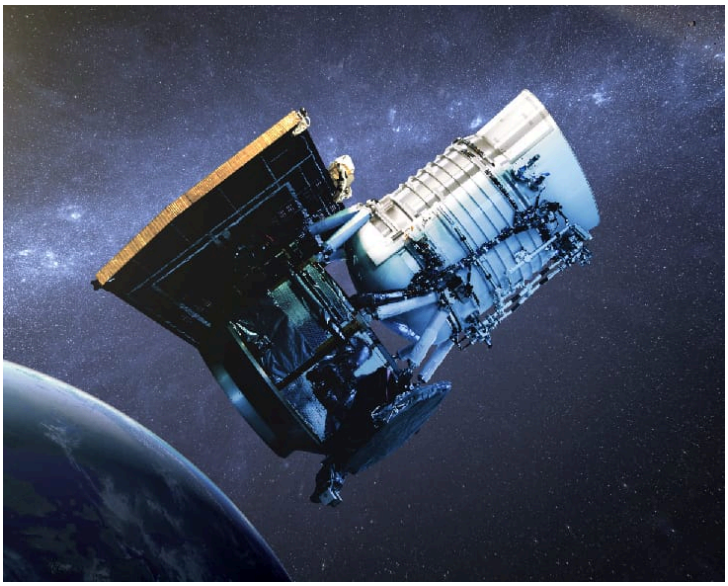


Figure 30.22 : Wide-Field Infrared Survey Explorer (WISE). Astronomers have used this infrared satellite to search for infrared signatures of enormous construction projects by very advanced civilizations, but their first survey did not reveal any. (credit: modification of work by NASA/JPL-Caltech)

Should We Transmit in Addition to Listening?

Our planet has some leakage of radio waves into space, from FM radio, television, military radars, and communication between Earth and our orbiting spacecraft. However, such leakage radiation is still quite weak, and therefore difficult to detect at the distances of the stars, at least with the radio technology we have. So at the present time our attempts to communicate with other civilizations that may be out there mostly involve trying to receive messages, but not sending any ourselves.

Some scientists, however, think that it is inconsistent to search for beacons from other civilizations without announcing our presence in a similar way. (We discussed earlier the problem that if every other civilization confined itself to listening, no one would ever get in touch.) So, should we be making regular attempts at sending easily decoded messages into space? Some scientists warn that our civilization is too immature and defenseless to announce ourselves at this early point in our development. The decision whether to transmit or not turns out to be an interesting reflection of how we feel about ourselves and our place in the universe.

Discussions of transmission also raise the question of who should speak for planet Earth. Today, anyone and everyone can broadcast radio signals, and many businesses, religious groups, and governments do. It would be a modest step for the same organizations to use or build large radio telescopes and begin intentional transmissions that are much stronger than the signals that leak from Earth today. And if we intercept a signal from an alien civilization, then the issue arises whether to reply.

Who should make the decision about whether, when, and how humanity announces itself to the cosmos? Is there freedom of speech when it comes to sending radio messages to other civilizations? Do all the nations of Earth have to agree before we send a signal strong enough that it has a serious chance of being received at the distances of the stars? How our species reaches a decision about these kinds of questions may well be a test of whether or not there is intelligent life on Earth.

Conclusion

Whether or not we ultimately turn out to be the only intelligent species in our part of the Galaxy, our exploration of the cosmos will surely continue. An important part of that exploration will still be the search for biomarkers from inhabited planets that have not produced technological creatures that send out radio signals. After all, creatures like butterflies and dolphins may never build radio antennas, but we are happy to share our planet with them and would be delighted to find their counterparts on other worlds.

Whether or not life exists elsewhere is just one of the unsolved problems in astronomy that we have discussed in this book. A humble acknowledgment of how much we have left to learn about the universe is one of the fundamental hallmarks of science. This should not, however, prevent us from feeling exhilarated about how much we have already managed to discover, and feeling curious about what else we might find out in the years to come.

Our progress report on the ideas of astronomy ends here, but we hope that your interest in the universe does not. We hope you will keep up with developments in astronomy through media and online, or by going to an occasional public lecture by a local scientist. Who, after all, can even guess all the amazing things that future research projects will reveal about both the universe and our connection with it?

Answers

Exercise 19.4.1

With our optimistic assumptions about the other factors, L = 100 years and N = 100 such civilizations in the entire Galaxy. In that case, there are so few other civilizations like ours that we are unlikely to detect any signals in a SETI search. But suppose the average lifetime is a million years; in that case, there are a million such civilizations in the Galaxy, and some of them may be within range for radio communication.
The most important conclusion from this calculation is that even if we are extremely optimistic about the probabilities, the only way we can expect success from SETI is if other civilizations are much older (and hence probably much more advanced) than ours.

This page titled 19.4: The Search for Extraterrestrial Intelligence is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 30.5: The Search for Extraterrestrial Intelligence by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

INDEX

A

absorption spectrum
5.3: Spectroscopy in Astronomy
accelerate
3.3: The Birth of Modern Astronomy
accretion
6.5: Formation of the Solar System
accretion disk
14.7: Evidence for Black Holes
16.7: Supermassive Black Holes- What Quasars really are
active galactic nuclei
16.6: Quasars
active galaxies
16.6: Quasars
active region
11.7: Solar Activity above the Photosphere
adaptive optics
5.8: Telescopes Today
agn
16.6: Quasars
amino acids
19.2: Astrobiology
Andromeda Galaxy
16.1: The Discovery of Galaxies
angular momentum
4.1: Newton's Great Synthesis
aperture
5.7: Telescopes
apparent brightness
12.1: The Brightness of Stars
apparent magnitude
3.1: Ancient Astronomy
Arctic Circle
2.4: The Seasons
Ascension
2.3: Earth and Sky
association
12.10: Star Clusters
asteroid
10.1: Asteroids
6.1: Overview of Our Planetary System
asteroid belt
10.1: Asteroids
astrobiology
19.2: Astrobiology
astrology
3.2: Astrology and Astronomy
astronomical
2.2: The Sky Above
astronomical unit
3.4: The Laws of Planetary Motion
astronomical unit (AU)
1.1: A Tour of the Universe
3.4: The Laws of Planetary Motion
AU
3.4: The Laws of Planetary Motion
Aurora
11.3: The Structure and Composition of the Sun

B

Balmer lines
12.4: The Spectra of Stars (and Brown Dwarfs)
baryon cycle
15.5: The Life Cycle of Cosmic Material
Big Bang
18.2: The Age of the Universe
binary stars
12.7: Measuring Stellar Masses
biomarker
19.3: Searching for Life beyond Earth
black hole
14.6: Black Holes
15.10: The Center of the Galaxy
blackbody
5.2: The Electromagnetic Spectrum
brown dwarf
12.4: The Spectra of Stars (and Brown Dwarfs)
12.7: Measuring Stellar Masses

C

Callisto
9.3: The Gailean Moons of Jupiter
CCD
5.9: Visible-Light Detectors and Instruments
celestial equator
2.2: The Sky Above
celestial poles
2.2: The Sky Above
celestial sphere
2.2: The Sky Above
Centaurs
10.5: The Origin and Fate of Comets and Related Objects
central bulge
15.1: The Architecture of the Galaxy
Chandrasekhar limit
14.1: The Death of Low-Mass Stars
Charon
10.6: Pluto and Charon
chromatic aberration
5.7: Telescopes
chromosphere
11.3: The Structure and Composition of the Sun
circumpolar zone
2.2: The Sky Above
closed universe
18.3: A Model of the Universe
CMB
17.2: The Cosmic Microwave Background
CME
11.7: Solar Activity above the Photosphere
color index
12.3: Colors of Stars
coma (comet)
10.4: The "Long-Haired" Comets
comet
10.4: The "Long-Haired" Comets
6.1: Overview of Our Planetary System
Conduction
11.4: The Solar Interior - Theory
conservation of energy
11.1: Sources of Sunshine- Thermal and Gravitational Energy
constellation
1.3: The Universe on the Large Scale
continuous spectrum
5.3: Spectroscopy in Astronomy
Convection
11.4: The Solar Interior - Theory
Copernicus
3.3: The Birth of Modern Astronomy
corona
11.3: The Structure and Composition of the Sun
coronal hole
11.3: The Structure and Composition of the Sun
coronal mass ejection
11.7: Solar Activity above the Photosphere
coronal mass ejection (CME)
11.7: Solar Activity above the Photosphere
cosmic microwave background
17.2: The Cosmic Microwave Background
cosmological constant
18.2: The Age of the Universe
Cosmology
18.2: The Age of the Universe
3.1: Ancient Astronomy
critical density
18.3: A Model of the Universe
cycle of the seasons
2.4: The Seasons

D

dark energy
18.2: The Age of the Universe
dark matter
15.7: The Mass of the Galaxy
18.4: What Is the Universe Really Made Of?
dark matter halo
15.1: The Architecture of the Galaxy
declination
2.3: Earth and Sky
degenerate gas
14.1: The Death of Low-Mass Stars
Degenerate Stars
14.1: The Death of Low-Mass Stars
density
4.1: Newton's Great Synthesis
deoxyribonucleic acid
19.2: Astrobiology
detector
5.7: Telescopes
deuterium
17.1: The Beginning of the Universe
differential force
2.6: Ocean Tides and the Moon
differential forces
2.6: Ocean Tides and the Moon

differential galactic rotation
15.6: Spiral Structure
differential rotation
11.6: The Solar Cycle
differentiation
6.2: Composition and Structure of Planets
Dispersion
5.3: Spectroscopy in Astronomy
DNA
19.2: Astrobiology
Doppler effect
5.6: The Doppler Effect
Drake equation
19.4: The Search for Extraterrestrial Intelligence
dwarf planet
10.6: Pluto and Charon

E

Eccentricity
3.4: The Laws of Planetary Motion
eclipsing binary
12.8: Diameters of Stars
Ecliptic
2.2: The Sky Above
eddening (interstellar)
15.4: Cosmic Dust
electromagnetic radiation
5.1: The Behavior of Light
electromagnetic spectrum
5.2: The Electromagnetic Spectrum
ellipse
3.4: The Laws of Planetary Motion
elliptical galaxy
16.4: Types of Galaxies
emission spectrum
5.3: Spectroscopy in Astronomy
energy flux
5.2: The Electromagnetic Spectrum
energy level
5.4: The Structure of the Atom
epicycle
3.1: Ancient Astronomy
Europa
9.3: The Gailean Moons of Jupiter
event horizon
14.6: Black Holes
Excitation
5.5: Formation of Spectral Lines
exoplanet
7.2: Planets beyond the Solar System- Search and Discovery
extremophile
19.2: Astrobiology
eyepiece
5.7: Telescopes

F

Fermi paradox
19.2: Astrobiology
fission
11.2: Mass, Energy, and the Theory of Relativity
flat universe
17.2: The Cosmic Microwave Background
focus
3.4: The Laws of Planetary Motion
5.7: Telescopes
frequency
5.1: The Behavior of Light
fusion
11.2: Mass, Energy, and the Theory of Relativity
17.1: The Beginning of the Universe

G

galactic nuclei
16.7: Supermassive Black Holes- What Quasars really are
gamma rays
5.2: The Electromagnetic Spectrum
Ganymede
9.3: The Gailean Moons of Jupiter
gene
19.2: Astrobiology
geocentric
2.2: The Sky Above
George Gamow
17.1: The Beginning of the Universe
Georges Lemaître
17.1: The Beginning of the Universe
giant
12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion
giant molecular clouds
13.1: Star Formation
giant planet
6.1: Overview of Our Planetary System
globular cluster
12.10: Star Clusters
grand unified theories
17.3: The Inflationary Universe
granulation
11.3: The Structure and Composition of the Sun
Gravitational Contraction
11.1: Sources of Sunshine- Thermal and Gravitational Energy
gravity
4.2: Newton's Universal Law of Gravitation
great circle
2.3: Earth and Sky
greenhouse effect
8.1: Life, Chemical Evolution, and Climate Change
greenhouse gas
8.1: Life, Chemical Evolution, and Climate Change
ground state
5.5: Formation of Spectral Lines

H

H II region
15.3: Interstellar Gas
H–R diagram
12.9: The H-R Diagram
habitable environment
19.2: Astrobiology
habitable zone
19.3: Searching for Life beyond Earth
halo
15.1: The Architecture of the Galaxy
heliocentric
3.3: The Birth of Modern Astronomy
helioseismology
11.5: The Solar Interior - Observations
helium flash
13.4: Further Evolution of Stars
horizon
2.2: The Sky Above
horoscope
3.2: Astrology and Astronomy
Hubble
16.1: The Discovery of Galaxies
Hubble Constant
16.3: The Expanding Universe
hubble telescope
1.2: Consequences of Light Travel Time
Hubble's law
16.3: The Expanding Universe
hydrostatic equilibrium
11.4: The Solar Interior - Theory
hypothesis
3.5: The Nature of Science

I

inflationary universe
17.3: The Inflationary Universe
infrared
5.2: The Electromagnetic Spectrum
interference
5.10: Radio Telescopes
interferometer
5.10: Radio Telescopes
interferometer array
5.10: Radio Telescopes
interstellar dust
15.2: The Interstellar Medium
interstellar extinction
15.4: Cosmic Dust
interstellar medium
15.2: The Interstellar Medium
inverse square law
5.1: The Behavior of Light
ion
5.5: Formation of Spectral Lines
ionization
5.5: Formation of Spectral Lines
iron meteorite
10.3: Meteorites - Stones from Heaven
irregular galaxy
16.4: Types of Galaxies
island universes
1.3: The Universe on the Large Scale
isotope
5.4: The Structure of the Atom

K

Kepler's first law
3.4: The Laws of Planetary Motion
Kepler's second law
3.4: The Laws of Planetary Motion
Kepler's third law
3.4: The Laws of Planetary Motion

Kuiper Belt
10.5: The Origin and Fate of Comets and Related Objects

L

lithium
17.1: The Beginning of the Universe
lo
9.3: The Gailean Moons of Jupiter
luminosity
12.1: The Brightness of Stars
lunar eclipse
2.7: Eclipses of the Sun and Moon

M

Magellanic Clouds
1.3: The Universe on the Large Scale
magnitude
12.1: The Brightness of Stars
main sequence
12.9: The H-R Diagram
major axis
3.4: The Laws of Planetary Motion
Margaret Huggins
12.4: The Spectra of Stars (and Brown Dwarfs)
mass extinction
10.7: Cosmic Influences on the Evolution of Earth
Maunder Minimum
11.6: The Solar Cycle
meridian
2.3: Earth and Sky
meteor
10.2: Meteors
6.1: Overview of Our Planetary System
meteor shower
10.2: Meteors
meteorite
10.3: Meteorites - Stones from Heaven
6.1: Overview of Our Planetary System
microwave
5.2: The Electromagnetic Spectrum
Milky Way Galaxy
15.1: The Architecture of the Galaxy
millisecond pulsar
14.5: The Evolution of Binary Star Systems
molecular cloud
15.3: Interstellar Gas
momentum
4.1: Newton's Great Synthesis

N

nebula
15.2: The Interstellar Medium
Nebulae
16.1: The Discovery of Galaxies
neutrino
11.2: Mass, Energy, and the Theory of Relativity
neutron star
14.2: Evolution of Massive Stars- An Explosive Finish
14.4: Pulsars and the Discovery of Neutron Stars
Newton's first law
4.1: Newton's Great Synthesis
Newton's second law
4.1: Newton's Great Synthesis
Newton's third law
4.1: Newton's Great Synthesis
north pole
2.4: The Seasons
Northern Hemisphere
2.4: The Seasons
nova
14.5: The Evolution of Binary Star Systems
nucleosynthesis
13.5: The Evolution of More Massive Stars
nucleus
10.4: The "Long-Haired" Comets
5.4: The Structure of the Atom
nucleus (of a comet)
10.4: The "Long-Haired" Comets

O

Oort Cloud
10.5: The Origin and Fate of Comets and Related Objects
open cluster
12.10: Star Clusters
open universe
18.3: A Model of the Universe
orbit
3.4: The Laws of Planetary Motion
orbital period
3.4: The Laws of Planetary Motion
orbital speed
3.4: The Laws of Planetary Motion
Organic Compound
19.2: Astrobiology
organic molecule
19.1: The Cosmic Context for Life

P

parallax
12.2: Surveying the Stars
3.1: Ancient Astronomy
parsec
12.2: Surveying the Stars
phases of the Moon
2.5: Phases and Motions of the Moon
photon
5.1: The Behavior of Light
photon decoupling time
17.2: The Cosmic Microwave Background
photosphere
11.3: The Structure and Composition of the Sun
photosynthesis
19.2: Astrobiology
8.1: Life, Chemical Evolution, and Climate Change
plage
11.7: Solar Activity above the Photosphere
planet
2.2: The Sky Above
planetary nebula
13.4: Further Evolution of Stars
planetesimals
6.4: Origin of the Solar System
plasma
11.3: The Structure and Composition of the Sun
Pluto
10.6: Pluto and Charon
population I star
15.8: Stellar Populations in the Galaxy
population II star
15.8: Stellar Populations in the Galaxy
Positron
11.2: Mass, Energy, and the Theory of Relativity
precession
3.1: Ancient Astronomy
precession of Earth
3.1: Ancient Astronomy
prime focus
5.7: Telescopes
prolate spheroid
2.6: Ocean Tides and the Moon
prominence
11.7: Solar Activity above the Photosphere
proper motion
12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion
protein
19.2: Astrobiology
protostar
13.1: Star Formation
pulsar
14.4: Pulsars and the Discovery of Neutron Stars

Q

quasar
16.6: Quasars

R

radar
5.10: Radio Telescopes
radial velocity
12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion
5.6: The Doppler Effect
radiation
11.4: The Solar Interior - Theory
radio waves
5.2: The Electromagnetic Spectrum
radioactivity
6.3: Dating Planetary Surfaces
reddening
15.4: Cosmic Dust
reddening (interstellar)
15.4: Cosmic Dust
redshift
16.3: The Expanding Universe
Reflecting Telescope
5.7: Telescopes
refracting telescope
5.7: Telescopes
refraction
2.4: The Seasons
resolution
5.8: Telescopes Today
retrograde motion
3.1: Ancient Astronomy
ribonucleic acid
19.2: Astrobiology

right ascension
2.3: Earth and Sky
RNA
19.2: Astrobiology
runaway greenhouse effect
8.2: The Massive Atmosphere of Venus

S

scientific notation
1.4: Numbers in Astronomy
seeing
5.8: Telescopes Today
semimajor axis
3.4: The Laws of Planetary Motion
SETI
19.4: The Search for Extraterrestrial Intelligence
sidereal month
2.5: Phases and Motions of the Moon
Singularity
14.6: Black Holes
SN 1987A
14.3: Supernova Observations
solar flare
11.7: Solar Activity above the Photosphere
solar month
2.5: Phases and Motions of the Moon
solar nebula
6.4: Origin of the Solar System
solar wind
11.3: The Structure and Composition of the Sun
Southern Hemisphere
2.4: The Seasons
space velocity
12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion
spectral class
12.4: The Spectra of Stars (and Brown Dwarfs)
spectrometer
5.3: Spectroscopy in Astronomy
spectroscopic binary
12.7: Measuring Stellar Masses
spiral arm
15.6: Spiral Structure
spiral galaxy
16.4: Types of Galaxies
star clusters
16.1: The Discovery of Galaxies
stellar wind
13.1: Star Formation
stony meteorite
10.3: Meteorites - Stones from Heaven
stromatolites
19.2: Astrobiology
sunspot
11.6: The Solar Cycle
sunspot cycle
11.6: The Solar Cycle
supermassive black hole
15.10: The Center of the Galaxy
supernova
14.3: Supernova Observations
supernovae
14.3: Supernova Observations
synchronous rotation
2.5: Phases and Motions of the Moon
synchrotron radiation
9.2: The Giant Planets

T

tail
10.4: The "Long-Haired" Comets
telescope
5.7: Telescopes
terrestrial planet
6.1: Overview of Our Planetary System
thermophile
19.2: Astrobiology
tidal force
9.3: The Gailean Moons of Jupiter
tidal heating
9.3: The Gailean Moons of Jupiter
tides
2.6: Ocean Tides and the Moon
Titan
9.4: Titan and Triton
transit
7.2: Planets beyond the Solar System- Search and Discovery
transit depth
7.2: Planets beyond the Solar System- Search and Discovery
transition region
11.3: The Structure and Composition of the Sun
triton
9.4: Titan and Triton
Tropic of Cancer
2.4: The Seasons
Tropic of Capricorn
2.4: The Seasons
type Ia supernova
16.2: The Extragalactic Distance Scale
type II supernova
14.2: Evolution of Massive Stars- An Explosive Finish

U

ultraviolet
5.2: The Electromagnetic Spectrum

V

velocity
4.1: Newton's Great Synthesis
visible light
5.2: The Electromagnetic Spectrum
visual binary
12.7: Measuring Stellar Masses

W

wavelength
5.1: The Behavior of Light
white dwarf
12.9: The H-R Diagram
White Dwarf Explosions
14.5: The Evolution of Binary Star Systems
white dwarfs
14.1: The Death of Low-Mass Stars
Wien's law
5.2: The Electromagnetic Spectrum
William Huggins
12.4: The Spectra of Stars (and Brown Dwarfs)

Z

Zodiac
2.2: The Sky Above

This page titled Index is shared under a license and was authored, remixed, and/or curated by Alessandro Baldi.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a license and was authored, remixed, and/or curated by Alessandro Baldi.

DETAILED LICENSING

Overview

Title: The Astronomer's Guide to the Universe

Webpages: 144

All licenses found:

- CC BY 4.0: 92.4% (133 pages)
- Undeclared: 6.9% (10 pages)
- Public Domain: 0.7% (1 page)

By Page

- The Astronomer's Guide to the Universe - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Public Domain*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: The Modern Cosmos - *CC BY 4.0*
    - 1.1: A Tour of the Universe - *CC BY 4.0*
    - 1.2: Consequences of Light Travel Time - *CC BY 4.0*
    - 1.3: The Universe on the Large Scale - *CC BY 4.0*
    - 1.4: Numbers in Astronomy - *CC BY 4.0*
  - 2: The Universe through Your Eyes - *CC BY 4.0*
    - 2.1: Thinking Ahead - *CC BY 4.0*
    - 2.2: The Sky Above - *CC BY 4.0*
    - 2.3: Earth and Sky - *CC BY 4.0*
    - 2.4: The Seasons - *CC BY 4.0*
    - 2.5: Phases and Motions of the Moon - *CC BY 4.0*
    - 2.6: Ocean Tides and the Moon - *CC BY 4.0*
    - 2.7: Eclipses of the Sun and Moon - *CC BY 4.0*
  - 3: The Science of Astronomy - *CC BY 4.0*
    - 3.1: Ancient Astronomy - *CC BY 4.0*
    - 3.2: Astrology and Astronomy - *CC BY 4.0*
    - 3.3: The Birth of Modern Astronomy - *CC BY 4.0*
    - 3.4: The Laws of Planetary Motion - *CC BY 4.0*
    - 3.5: The Nature of Science - *CC BY 4.0*
  - 4: Making Sense of the Universe - *CC BY 4.0*
    - 4.1: Newton's Great Synthesis - *CC BY 4.0*
    - 4.2: Newton's Universal Law of Gravitation - *CC BY 4.0*
  - 5: Light- Messenger of the Cosmos - *CC BY 4.0*
    - 5.1: The Behavior of Light - *CC BY 4.0*
    - 5.2: The Electromagnetic Spectrum - *CC BY 4.0*
    - 5.3: Spectroscopy in Astronomy - *CC BY 4.0*
    - 5.4: The Structure of the Atom - *CC BY 4.0*
    - 5.5: Formation of Spectral Lines - *CC BY 4.0*
    - 5.6: The Doppler Effect - *CC BY 4.0*
    - 5.7: Telescopes - *CC BY 4.0*
    - 5.8: Telescopes Today - *CC BY 4.0*
    - 5.9: Visible-Light Detectors and Instruments - *CC BY 4.0*
    - 5.10: Radio Telescopes - *CC BY 4.0*
    - 5.11: Observations outside Earth's Atmosphere - *CC BY 4.0*
    - 5.12: The Future of Large Telescopes - *CC BY 4.0*
  - 6: Introduction to the Solar System - *CC BY 4.0*
    - 6.1: Overview of Our Planetary System - *CC BY 4.0*
    - 6.2: Composition and Structure of Planets - *CC BY 4.0*
    - 6.3: Dating Planetary Surfaces - *CC BY 4.0*
    - 6.4: Origin of the Solar System - *CC BY 4.0*
    - 6.5: Formation of the Solar System - *CC BY 4.0*
  - 7: Worlds Around Other Stars - *CC BY 4.0*
    - 7.1: Evidence That Planets Form around Other Stars - *CC BY 4.0*
    - 7.2: Planets beyond the Solar System- Search and Discovery - *CC BY 4.0*
    - 7.3: Exoplanets Everywhere - What we are Learning - *CC BY 4.0*
    - 7.4: New Perspectives on Planet Formation - *CC BY 4.0*
  - 8: Terrestrial Planets - *CC BY 4.0*
    - 8.1: Life, Chemical Evolution, and Climate Change - *CC BY 4.0*
    - 8.2: The Massive Atmosphere of Venus - *CC BY 4.0*
    - 8.3: Water and Life on Mars - *CC BY 4.0*
    - 8.4: Divergent Planetary Evolution - *CC BY 4.0*
  - 9: Jovian Planets - *CC BY 4.0*
    - 9.1: Exploring the Outer Planets - *CC BY 4.0*
    - 9.2: The Giant Planets - *CC BY 4.0*
    - 9.3: The Gailean Moons of Jupiter - *CC BY 4.0*
    - 9.4: Titan and Triton - *CC BY 4.0*
  - 10: Asteroids, Comets, and Dwarf Planets - *CC BY*

4.0
- 10.1: Asteroids - *CC BY 4.0*
- 10.2: Meteors - *CC BY 4.0*
- 10.3: Meteorites - Stones from Heaven - *CC BY 4.0*
- 10.4: The "Long-Haired" Comets - *CC BY 4.0*
- 10.5: The Origin and Fate of Comets and Related Objects - *CC BY 4.0*
- 10.6: Pluto and Charon - *CC BY 4.0*
- 10.7: Cosmic Influences on the Evolution of Earth - *CC BY 4.0*

◦ 11: The Sun, Our Star - *CC BY 4.0*
- 11.1: Sources of Sunshine- Thermal and Gravitational Energy - *CC BY 4.0*
- 11.2: Mass, Energy, and the Theory of Relativity - *CC BY 4.0*
- 11.3: The Structure and Composition of the Sun - *CC BY 4.0*
- 11.4: The Solar Interior - Theory - *CC BY 4.0*
- 11.5: The Solar Interior - Observations - *CC BY 4.0*
- 11.6: The Solar Cycle - *CC BY 4.0*
- 11.7: Solar Activity above the Photosphere - *CC BY 4.0*
- 11.8: Space Weather - *CC BY 4.0*

◦ 12: The Stellar Census - *CC BY 4.0*
- 12.1: The Brightness of Stars - *CC BY 4.0*
- 12.2: Surveying the Stars - *CC BY 4.0*
- 12.3: Colors of Stars - *CC BY 4.0*
- 12.4: The Spectra of Stars (and Brown Dwarfs) - *CC BY 4.0*
- 12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion - *CC BY 4.0*
- 12.6: A Stellar Census - *CC BY 4.0*
- 12.7: Measuring Stellar Masses - *CC BY 4.0*
- 12.8: Diameters of Stars - *CC BY 4.0*
- 12.9: The H-R Diagram - *CC BY 4.0*
- 12.10: Star Clusters - *CC BY 4.0*
- 12.11: Checking Out the Theory - *CC BY 4.0*

◦ 13: Stellar Evolution - *CC BY 4.0*
- 13.1: Star Formation - *CC BY 4.0*
- 13.2: The H-R and the Study of Stellar Evolution - *CC BY 4.0*
- 13.3: Evolution from the Main Sequence to Red Giants - *CC BY 4.0*
- 13.4: Further Evolution of Stars - *CC BY 4.0*
- 13.5: The Evolution of More Massive Stars - *CC BY 4.0*

◦ 14: White Dwarfs, Neutron Stars, and Black Holes - *CC BY 4.0*
- 14.1: The Death of Low-Mass Stars - *CC BY 4.0*
- 14.2: Evolution of Massive Stars- An Explosive Finish - *CC BY 4.0*
- 14.3: Supernova Observations - *CC BY 4.0*
- 14.4: Pulsars and the Discovery of Neutron Stars - *CC BY 4.0*
- 14.5: The Evolution of Binary Star Systems - *CC BY 4.0*
- 14.6: Black Holes - *CC BY 4.0*
- 14.7: Evidence for Black Holes - *CC BY 4.0*

◦ 15: The Milky Way, Our Island Universe - *CC BY 4.0*
- 15.1: The Architecture of the Galaxy - *CC BY 4.0*
- 15.2: The Interstellar Medium - *CC BY 4.0*
- 15.3: Interstellar Gas - *CC BY 4.0*
- 15.4: Cosmic Dust - *CC BY 4.0*
- 15.5: The Life Cycle of Cosmic Material - *CC BY 4.0*
- 15.6: Spiral Structure - *CC BY 4.0*
- 15.7: The Mass of the Galaxy - *CC BY 4.0*
- 15.8: Stellar Populations in the Galaxy - *CC BY 4.0*
- 15.9: The Formation of the Galaxy - *CC BY 4.0*
- 15.10: The Center of the Galaxy - *CC BY 4.0*

◦ 16: The Expanding Universe of Galaxies - *CC BY 4.0*
- 16.1: The Discovery of Galaxies - *CC BY 4.0*
- 16.2: The Extragalactic Distance Scale - *CC BY 4.0*
- 16.3: The Expanding Universe - *CC BY 4.0*
- 16.4: Types of Galaxies - *CC BY 4.0*
- 16.5: Properties of Galaxies - *CC BY 4.0*
- 16.6: Quasars - *CC BY 4.0*
- 16.7: Supermassive Black Holes- What Quasars really are - *CC BY 4.0*
- 16.8: Quasars as Probes of Evolution in the Universe - *CC BY 4.0*

◦ 17: The Big Bang and the Origin of the Cosmos - *CC BY 4.0*
- 17.1: The Beginning of the Universe - *CC BY 4.0*
- 17.2: The Cosmic Microwave Background - *CC BY 4.0*
- 17.3: The Inflationary Universe - *CC BY 4.0*

◦ 18: Dark Matter and Dark Energy - *CC BY 4.0*
- 18.1: The Challenge of Dark Matter - *CC BY 4.0*
- 18.2: The Age of the Universe - *CC BY 4.0*
- 18.3: A Model of the Universe - *CC BY 4.0*
- 18.4: What Is the Universe Really Made Of? - *CC BY 4.0*

◦ 19: Life in the Universe - *CC BY 4.0*
- 19.1: The Cosmic Context for Life - *CC BY 4.0*
- 19.2: Astrobiology - *CC BY 4.0*
- 19.3: Searching for Life beyond Earth - *CC BY 4.0*
- 19.4: The Search for Extraterrestrial Intelligence - *CC BY 4.0*

◦ Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

This page titled The Astronomer's Guide to the Universe is shared under a license and was authored, remixed, and/or curated by Alessandro Baldi.

DETAILED LICENSING

Overview

Title: The Astronomer's Guide to the Universe

Webpages: 144

All licenses found:

- CC BY 4.0: 92.4% (133 pages)
- Undeclared: 6.9% (10 pages)
- Public Domain: 0.7% (1 page)

By Page

- The Astronomer's Guide to the Universe - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Public Domain*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: The Modern Cosmos - *CC BY 4.0*
    - 1.1: A Tour of the Universe - *CC BY 4.0*
    - 1.2: Consequences of Light Travel Time - *CC BY 4.0*
    - 1.3: The Universe on the Large Scale - *CC BY 4.0*
    - 1.4: Numbers in Astronomy - *CC BY 4.0*
  - 2: The Universe through Your Eyes - *CC BY 4.0*
    - 2.1: Thinking Ahead - *CC BY 4.0*
    - 2.2: The Sky Above - *CC BY 4.0*
    - 2.3: Earth and Sky - *CC BY 4.0*
    - 2.4: The Seasons - *CC BY 4.0*
    - 2.5: Phases and Motions of the Moon - *CC BY 4.0*
    - 2.6: Ocean Tides and the Moon - *CC BY 4.0*
    - 2.7: Eclipses of the Sun and Moon - *CC BY 4.0*
  - 3: The Science of Astronomy - *CC BY 4.0*
    - 3.1: Ancient Astronomy - *CC BY 4.0*
    - 3.2: Astrology and Astronomy - *CC BY 4.0*
    - 3.3: The Birth of Modern Astronomy - *CC BY 4.0*
    - 3.4: The Laws of Planetary Motion - *CC BY 4.0*
    - 3.5: The Nature of Science - *CC BY 4.0*
  - 4: Making Sense of the Universe - *CC BY 4.0*
    - 4.1: Newton's Great Synthesis - *CC BY 4.0*
    - 4.2: Newton's Universal Law of Gravitation - *CC BY 4.0*
  - 5: Light- Messenger of the Cosmos - *CC BY 4.0*
    - 5.1: The Behavior of Light - *CC BY 4.0*
    - 5.2: The Electromagnetic Spectrum - *CC BY 4.0*
    - 5.3: Spectroscopy in Astronomy - *CC BY 4.0*
    - 5.4: The Structure of the Atom - *CC BY 4.0*
    - 5.5: Formation of Spectral Lines - *CC BY 4.0*
    - 5.6: The Doppler Effect - *CC BY 4.0*
    - 5.7: Telescopes - *CC BY 4.0*
    - 5.8: Telescopes Today - *CC BY 4.0*
    - 5.9: Visible-Light Detectors and Instruments - *CC BY 4.0*
    - 5.10: Radio Telescopes - *CC BY 4.0*
    - 5.11: Observations outside Earth's Atmosphere - *CC BY 4.0*
    - 5.12: The Future of Large Telescopes - *CC BY 4.0*
  - 6: Introduction to the Solar System - *CC BY 4.0*
    - 6.1: Overview of Our Planetary System - *CC BY 4.0*
    - 6.2: Composition and Structure of Planets - *CC BY 4.0*
    - 6.3: Dating Planetary Surfaces - *CC BY 4.0*
    - 6.4: Origin of the Solar System - *CC BY 4.0*
    - 6.5: Formation of the Solar System - *CC BY 4.0*
  - 7: Worlds Around Other Stars - *CC BY 4.0*
    - 7.1: Evidence That Planets Form around Other Stars - *CC BY 4.0*
    - 7.2: Planets beyond the Solar System- Search and Discovery - *CC BY 4.0*
    - 7.3: Exoplanets Everywhere - What we are Learning - *CC BY 4.0*
    - 7.4: New Perspectives on Planet Formation - *CC BY 4.0*
  - 8: Terrestrial Planets - *CC BY 4.0*
    - 8.1: Life, Chemical Evolution, and Climate Change - *CC BY 4.0*
    - 8.2: The Massive Atmosphere of Venus - *CC BY 4.0*
    - 8.3: Water and Life on Mars - *CC BY 4.0*
    - 8.4: Divergent Planetary Evolution - *CC BY 4.0*
  - 9: Jovian Planets - *CC BY 4.0*
    - 9.1: Exploring the Outer Planets - *CC BY 4.0*
    - 9.2: The Giant Planets - *CC BY 4.0*
    - 9.3: The Gailean Moons of Jupiter - *CC BY 4.0*
    - 9.4: Titan and Triton - *CC BY 4.0*
  - 10: Asteroids, Comets, and Dwarf Planets - *CC BY*

4.0
- 10.1: Asteroids - *CC BY 4.0*
- 10.2: Meteors - *CC BY 4.0*
- 10.3: Meteorites - Stones from Heaven - *CC BY 4.0*
- 10.4: The "Long-Haired" Comets - *CC BY 4.0*
- 10.5: The Origin and Fate of Comets and Related Objects - *CC BY 4.0*
- 10.6: Pluto and Charon - *CC BY 4.0*
- 10.7: Cosmic Influences on the Evolution of Earth - *CC BY 4.0*

◦ 11: The Sun, Our Star - *CC BY 4.0*
- 11.1: Sources of Sunshine- Thermal and Gravitational Energy - *CC BY 4.0*
- 11.2: Mass, Energy, and the Theory of Relativity - *CC BY 4.0*
- 11.3: The Structure and Composition of the Sun - *CC BY 4.0*
- 11.4: The Solar Interior - Theory - *CC BY 4.0*
- 11.5: The Solar Interior - Observations - *CC BY 4.0*
- 11.6: The Solar Cycle - *CC BY 4.0*
- 11.7: Solar Activity above the Photosphere - *CC BY 4.0*
- 11.8: Space Weather - *CC BY 4.0*

◦ 12: The Stellar Census - *CC BY 4.0*
- 12.1: The Brightness of Stars - *CC BY 4.0*
- 12.2: Surveying the Stars - *CC BY 4.0*
- 12.3: Colors of Stars - *CC BY 4.0*
- 12.4: The Spectra of Stars (and Brown Dwarfs) - *CC BY 4.0*
- 12.5: Using Spectra to Measure Stellar Radius, Composition, and Motion - *CC BY 4.0*
- 12.6: A Stellar Census - *CC BY 4.0*
- 12.7: Measuring Stellar Masses - *CC BY 4.0*
- 12.8: Diameters of Stars - *CC BY 4.0*
- 12.9: The H-R Diagram - *CC BY 4.0*
- 12.10: Star Clusters - *CC BY 4.0*
- 12.11: Checking Out the Theory - *CC BY 4.0*

◦ 13: Stellar Evolution - *CC BY 4.0*
- 13.1: Star Formation - *CC BY 4.0*
- 13.2: The H-R and the Study of Stellar Evolution - *CC BY 4.0*
- 13.3: Evolution from the Main Sequence to Red Giants - *CC BY 4.0*
- 13.4: Further Evolution of Stars - *CC BY 4.0*
- 13.5: The Evolution of More Massive Stars - *CC BY 4.0*

◦ 14: White Dwarfs, Neutron Stars, and Black Holes - *CC BY 4.0*
- 14.1: The Death of Low-Mass Stars - *CC BY 4.0*
- 14.2: Evolution of Massive Stars- An Explosive Finish - *CC BY 4.0*
- 14.3: Supernova Observations - *CC BY 4.0*
- 14.4: Pulsars and the Discovery of Neutron Stars - *CC BY 4.0*
- 14.5: The Evolution of Binary Star Systems - *CC BY 4.0*
- 14.6: Black Holes - *CC BY 4.0*
- 14.7: Evidence for Black Holes - *CC BY 4.0*

◦ 15: The Milky Way, Our Island Universe - *CC BY 4.0*
- 15.1: The Architecture of the Galaxy - *CC BY 4.0*
- 15.2: The Interstellar Medium - *CC BY 4.0*
- 15.3: Interstellar Gas - *CC BY 4.0*
- 15.4: Cosmic Dust - *CC BY 4.0*
- 15.5: The Life Cycle of Cosmic Material - *CC BY 4.0*
- 15.6: Spiral Structure - *CC BY 4.0*
- 15.7: The Mass of the Galaxy - *CC BY 4.0*
- 15.8: Stellar Populations in the Galaxy - *CC BY 4.0*
- 15.9: The Formation of the Galaxy - *CC BY 4.0*
- 15.10: The Center of the Galaxy - *CC BY 4.0*

◦ 16: The Expanding Universe of Galaxies - *CC BY 4.0*
- 16.1: The Discovery of Galaxies - *CC BY 4.0*
- 16.2: The Extragalactic Distance Scale - *CC BY 4.0*
- 16.3: The Expanding Universe - *CC BY 4.0*
- 16.4: Types of Galaxies - *CC BY 4.0*
- 16.5: Properties of Galaxies - *CC BY 4.0*
- 16.6: Quasars - *CC BY 4.0*
- 16.7: Supermassive Black Holes- What Quasars really are - *CC BY 4.0*
- 16.8: Quasars as Probes of Evolution in the Universe - *CC BY 4.0*

◦ 17: The Big Bang and the Origin of the Cosmos - *CC BY 4.0*
- 17.1: The Beginning of the Universe - *CC BY 4.0*
- 17.2: The Cosmic Microwave Background - *CC BY 4.0*
- 17.3: The Inflationary Universe - *CC BY 4.0*

◦ 18: Dark Matter and Dark Energy - *CC BY 4.0*
- 18.1: The Challenge of Dark Matter - *CC BY 4.0*
- 18.2: The Age of the Universe - *CC BY 4.0*
- 18.3: A Model of the Universe - *CC BY 4.0*
- 18.4: What Is the Universe Really Made Of? - *CC BY 4.0*

◦ 19: Life in the Universe - *CC BY 4.0*
- 19.1: The Cosmic Context for Life - *CC BY 4.0*
- 19.2: Astrobiology - *CC BY 4.0*
- 19.3: Searching for Life beyond Earth - *CC BY 4.0*
- 19.4: The Search for Extraterrestrial Intelligence - *CC BY 4.0*

◦ Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

This page titled The Astronomer's Guide to the Universe is shared under a license and was authored, remixed, and/or curated by Alessandro Baldi.