Survey of Astronomy

Survey of Astronomy

ADAPTED BY JEAN CREIGHTON

UNIVERSITY OF WISCONSIN MILWAUKEE MILWAUKEE, WI



Survey of Astronomy by Adapted by Jean Creighton is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted.

How to Reuse & Attribute This Content



© Jan 8, 2019 OpenStax Astronomy. Textbook content produced by OpenStax Astronomy is licensed under a Creative Commons Attribution License 4.0 license.

Under this license, any user of this textbook or the textbook contents herein must provide proper attribution as follows:

The OpenStax name, OpenStax logo, OpenStax book covers, OpenStax CNX name, and OpenStax CNX logo are not subject to the creative commons license and may not be reproduced without the prior and express written consent of Rice University. For questions regarding this license, please contact support@openstax.org.

- If you use this textbook as a bibliographic reference, then you should cite it as follows: OpenStax Astronomy, Astronomy. OpenStax CNX. Jan 8, 2019 http://cnx.org/contents/ 2e737be8-ea65-48c3-aa0a-9f35b4c6a966@17.1.
- If you redistribute this textbook in a print format, then you must include on every physical page the following attribution: Download for free at http://cnx.org/contents/ 2e737be8-ea65-48c3-aa0a-9f35b4c6a966@17.1.

 If you redistribute part of this textbook, then you must retain in every digital format page view (including but not limited to EPUB, PDF, and HTML) and on every physical printed page the following attribution: Download for free at http://cnx.org/contents/

2e737be8-ea65-48c3-aa0a-9f35b4c6a966@17.1.

Contents

Acknowledgements	xi
Chapter 1 Science and the Universe: A Brief Tour Section 1.4: Numbers in Astronomy	1
Chapter 1 Science and the Universe: Section 1.5: Consequences of Light Travel Time	6
Chapter 1 Section 1.6: A Tour of the Universe	9
Chapter 1 Section 1.7: The Universe on the Large Scale	19
Chapter 2 Observing the Sky: The Birth of Astronomy Section 2.1: The Sky Above	24
Chapter 2 Section 2.2: Ancient Astronomy	45
Chapter 2 Section 2.3: Astrology and Astronomy	62
Chapter 2 Section 2.4: The Birth of Modern Astronomy	72
Chapter 3 Orbits and Gravity Section 3.1: The Laws of Planetary Motion	97
Chapter 3 Section 3.2: Newton's Great Synthesis	113
Chapter 3 Section 3.3: Newton's Universal Law of Gravitation	126
Chapter 4 Earth, Moon and Sky Section 4.2: The Seasons	144
Chapter 4 Section 4.5: Phases and Motions of the Moon	160
Chapter 4 Section 4.7: Eclipses of the Sun and Moon	170

Chapter 5 Radiation and Spectra Section 5.1: The Behavior of Light	193
Chapter 5 Section 5.2: The Electromagnetic Spectrum	209
Chapter 5 Section 5.3: Spectroscopy in Astronomy	225
Chapter 5 Section 5.4: The Structure of the Atom	236
Chapter 5 Section 5.5: The Formation of Spectral Lines	249
Chapter 5 Section 5.6: The Doppler Effect	259
Chapter 6 Astronomical Instruments Section 6.1: Telescopes	274
Chapter 6 Section 6.2: Telescopes Today	291
Chapter 6 Section 6.3: Visible-Light Detectors and	312
Instruments	
Chapter 6 Section 6.4: Radio Telescopes	319
Chapter 6 Section 6.5: Observations outside Earth's Atmosphere	333
Chapter 6 Section 6.6: The Future of Large Telescopes	343
Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System	356
Chapter 7 Section 7.2: Composition and Structure of Planets	379
Chapter 7 Section 7.3: Dating Planetary Surfaces	390
Chapter 7 Section 7.4: Origin of the Solar System	397
Chapter 17 Analyzing Starlight Section 17.1: The Brightness of Stars	411
Chapter 17 Section 17.2: Colors of Stars	421
Chapter 17 Section 17.3: The Spectra of Stars (and Brown Dwarfs)	427

Chapter 18 The Stars: A Celestial Census Section	455
Charden 10 Cardier 10 Charden in the line Manage	40.4
Chapter 18 Section 18.2: Measuring Stellar Masses	464
Chapter 18 Section 18.3: Diameters of Stars	481
Chapter 18 Section 18.4: The H-R Diagram	493
Chapter 19 Celestial Distances Section 19.1: Fundamental Units of Distance	522
Chapter 19 Section 19.2: Surveying the Stars	531
Chapter 19 Section 19.3: Variable Stars: One Key to Cosmic Distances	553
Chapter 19 Section 19.4: The H-R Diagram and Cosmic Distances	567
Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System	584
Chapter 21 Section 21.5: Exoplanets Everywhere: What We Are Learning	602
Chapter 23 The Death of Stars Section 23.1: The Death of Low-Mass Stars	624
Chapter 23 Section 23.2: Evolution of Massive Stars: An Explosive Finish	639
Chapter 23 Section 23.3: Supernova Observations	654
Chapter 23 Section 23.4: Pulsars and the Discovery of Neutron Stars	670
Chapter 23 Section 23.5: The Evolution of Binary Star Systems	684
Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts	693
Chapter 24 Black Holes and Curved Spacetime Section 24.1: Introduction General Relativity	723
Chapter 24 Section 24.2: Spacetime and Gravity	739

Chapter 24 Section 24.3: Tests of General	746
Relativity	
Chapter 24 Section 24.4: Time in General Relativity	752
Chapter 24 Section 24.5: Black Holes	758
Chapter 24 Section 24.6: Evidence for Black Holes	777
Chapter 24 Section 24.7: Gravitational Wave Astronomy	786
Chapter 25 The Milky Way Galaxy Section 25.1: The Architecture of the Galaxy	805
Chapter 26 Galaxies Section 26.1: The Discovery of Galaxies	827
Chapter 26 Section 26.2: Types of Galaxies	836
Chapter 26 Section 26.3: Properties of Galaxies	849
Chapter 26 Section 26.4: The Extragalactic Distance Scale	856
Chapter 26 Section 26.5: The Expanding Universe	864
Chapter 29 The Big Bang Section 29.3: The Beginning of the Universe	888
Chapter 29 Section 29.4: The Cosmic Microwave Background	903
Chapter 30 Life in the Universe Section 30.1: The Cosmic Context for Life	923
Chapter 30 Section 30.2: Astrobiology	932
Chapter 30 Section 30.3: Searching for Life beyond Earth	953
Chapter 30 Section 30.4: The Search for Extraterrestrial Intelligence	973
Appendix	1005

This remix of the *OpenStax* Astronomy text has been curated and edited for distribution in the University of Wisconsin Milwaukee Astronomy 185 course. Select chapters and sections have been remixed to support exploration of critical scientific questions and approaches K-12 teachers would incorporate in their classroom curricula.

As we publish this beta version of the text in January 2019, we have selected the following basic curriculum plan and the remixed version of this text includes the readings listed below.

Madeleine G. Pitsch Project Manager

Class 2 Readings

Section 2.1Titled:"Chapter 2 Observing the Sky: The Birth of Astronomy Section 2.1: The Sky Above"

URL: https://uwm.pressbooks.pub/astronomy/chapter/2-1/

Section 1.4 Titled: Chapter 1 Science and the Universe: A Brief Tour Section 1.4: Numbers in Astronomy"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-1/

Section 1.5 Titled: Chapter 1Science and the Universe: Section 1.5: Consequences of Light Travel Time

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-1-section-1-5/ Section 1.6 Titled: "Chapter 1 Section 1.6: A Tour of the Universe"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter1sectionatouroftheuniverse/

Section 1.7 Titled "Chapter 1 Section 1.7: The Universe on the Large Scale"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-1-section-1-7-the-universe-on-the-large-scale/

Class 3 Readings

Section 4.2 Titled: "Chapter 4 Earth, Moon and Sky Section 4.2: The Seasons"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-4/

Class 4 Readings: None

Class 5 Readings

Section 4.5 Titled: "Chapter 4 Earth, Moon and Sky Section 4.5: Phases and Motions of the Moon"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-4-earth-moon-and-sky-section-4-5-eclipses-of-the-sunand-moon/ Class 6 Readings

Section 4.7 Titled:"Chapter 4 Section 4.7: Eclipses of the Sun and Moon"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-4-section-4-7-eclipses-of-the-sun-and-moon/

Class 7 Readings

Section 2.2 Titled:"Chapter 2 Section 2.2: Ancient Astronomy"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-2-section-2-2-ancient-astronomy/

Section 2.4 Titled:"Chapter 2 Section 2.4: The Birth of Modern Astronomy"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-2-section-2-4-the-birth-of-modern-astronomy/

Section 3.1 Titled: "Chapter 3 Orbits and Gravity Section 3.1: The Laws of Planetary Motion"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-3/

Class 8 Readings: None

Class 9 Readings

Openstax Astronomy Textbook (URL: https://openstax.org/details/ books/astronomy)

Section 1.1 Titled: "The Nature of Astronomy"

URL: https://cnx.org/contents/LnN76Opl@17.1:fxG2yTd0@7/The-Nature-of-Astronomy

Section 1.2 Titled: "The Nature of Science" URL: https://cnx.org/contents/LnN76Opl@17.1:whULC0iB@8/The-Nature-of-Science

Class 10 Readings:

Section 3.2 Titled: "Chapter 3 Section 3.2: Newton's Great Synthesis"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-2-section-3-2-newtons-great-synthesis/

Section 3.3 Titled: "Chapter 3 Section 3.3: Newton's Universal Law of Gravitation"

URL: https://uwm.pressbooks.pub/astronomy/chapter/ chapter-3-section-3-3-newtons-universal-law-of-gravitation/

Chapter 1 Science and the Universe: A Brief Tour Section 1.4: Numbers in 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 $5 imes 10^8$. The small raised number after the 10, called an exponent, keeps track of the number of places we had to move the decimal point to the left to convert 500,000,000 to 5. If you are encountering this system for the first time or would like a refresher, we suggest you look at Appendix C and Example for more information. The second way we try to keep numbers simple is to use a consistent set of units-the metric International System of Units, or SI (from the French Système International d'Unités). The metric system is summarized in Appendix D (see Example).

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 3×10^5 kilometers per second (km/s), which makes a light-year 9.46×10^{12} kilometers. You might think that such a large unit would reach the nearest star easily, but the stars are far more remote than our imaginations might lead us to believe. Even the nearest star is 4.3 light-years away–more than 40 trillion kilometers. Other stars visible to the unaided eye are hundreds to thousands of light-years away (Figure 1).

Orion Nebula.



Figure 1. 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)

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 $7.92\ \times 10^{10}.$

Getting Familiar with a Light-Year

How many kilometers are there in a light-year?

Solution

Light travels $3\times 10^5~{\rm km}$ in 1 s. So, let's calculate how far it goes in a year:

- There are 60 (6×10^1) s in 1 min, and (6×10^1) min in 1 h.
- Multiply these together and you find that there are $3.6\times 10^3\,{\rm s/h.}$
- Thus, light covers 3×10^5 km/s $\times 3.6 \times 10^3$ s/h = 1.08×10^9 km/h.
- There are 24 or 2.4×10^1 h in a day, and 365.24 (3.65×10^2) 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.

Chapter 1 Science and the Universe: Section 1.5: Consequences of Light Travel Time

1.5 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) 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.

Telescope in Orbit.



Figure 1. 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)

Chapter 1 Section 1.6: A Tour of the Universe

1.6 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). 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.

Humanity's Home Base.



Figure 1. 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 2 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.

Earth and Moon, Drawn to Scale.

Figure 2. 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 $(3 \times 10^7 \text{ seconds})$ 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 3). 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.

Our Solar Family.



Figure 3. 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 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 4, with its spiral structure outlined by the blue light of hot adolescent stars.

Spiral Galaxy.



Figure 4. 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. <u>Figure 5</u> shows an image of the disk of the Galaxy as seen from our vantage point.

Milky Way Galaxy.



Figure 5. 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 6). 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.

Star Cluster.



Figure 6. 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."

Chapter 1 Section 1.7: The Universe on the Large Scale

1.7 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 75,000 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). 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.

Neighbor Galaxies.



Figure 1.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 2). 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.

Closest Spiral Galaxy.



Figure 2. 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 3).

Fornax Cluster of Galaxies.



Figure 3. 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 twentiethcentury 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.

Chapter 2 Observing the Sky: The Birth of Astronomy Section 2.1: The Sky Above

2. Thinking Ahead

Night Sky.



Figure 1. 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

24 | Chapter 2 Observing the Sky: The Birth of Astronomy Section 2.1: 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, worshiped, 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.

2.1 The Sky Above



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 which the heavens around 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 1), 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 1. 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). Some thought of it as an actual sphere of transparent crystalline material, with the stars embedded in it like tiny jewels.

Circles on the Celestial Sphere.



Figure 2. 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 3). 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.

Circling the South Celestial Pole.



Figure 3. 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.)

The Rotating Sky Lab created by the University of Nebraska–Lincoln provides an interactive demonstration that introduces the horizon coordinate system, the apparent rotation of the sky, and allows for exploration of the relationship between the horizon and celestial equatorial coordinate systems.

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").

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 4). 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.

Star Circles at Different Latitudes.



Figure 4. The turning of the sky looks different depending on your latitude on Earth. (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 5 and Table)-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.

Constellations on the Ecliptic.



Figure 5. 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.

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 6). 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."

The Celestial Tilt



Figure 6. 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.

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:

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

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 $\frac{75^{\circ}}{5h} = \frac{15^{\circ}}{1h}$. 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:

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

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

Check Your Learning

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:

The speed of the moon is $0.5^{\circ}/1$ h. To move a full 360°, the moon needs 720 h: $\frac{0.5^{\circ}}{1h} = \frac{360^{\circ}}{720h}$ Dividing 720 h by the conversion factor of 24 h/day reveals the lunar cycle is about 30 days. 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 5). (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 7). 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 7. (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.

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.

The direct evidence of our senses supports a geocentric perspective, with the celestial sphere pivoting on the celestial poles and rotating about a stationary Earth. We see only half of this sphere at one time, limited by the horizon; the point directly overhead is our zenith. The Sun's annual path on the celestial sphere is the ecliptic–a line that runs through the center of the zodiac, which

is the 18-degree-wide strip of the sky within which we always find the Moon and planets. The celestial sphere is organized into 88 constellations, or sectors.

Glossary

celestial equator

a great circle on the celestial sphere 90° from the celestial poles; where the celestial sphere intersects the plane of Earth's equator

celestial poles

points about which the celestial sphere appears to rotate; intersections of the celestial sphere with Earth's polar axis

celestial sphere

the apparent sphere of the sky; a sphere of large radius centered on the observer; directions of objects in the sky can be denoted by their position on the celestial sphere

circumpolar zone

those portions of the celestial sphere near the celestial poles that are either always above or always below the horizon

ecliptic

the apparent annual path of the Sun on the celestial sphere

geocentric

centered on Earth

horizon (astronomical)

a great circle on the celestial sphere 90° from the zenith; more popularly, the circle around us where the dome of the sky meets Earth

planet

today, any of the larger objects revolving about the Sun or any similar objects that orbit other stars; in ancient times, any object that moved regularly among the fixed stars

year

the period of revolution of Earth around the Sun

zenith

the point on the celestial sphere opposite the direction of gravity; point directly above the observer

zodiac

a belt around the sky about 18° wide centered on the ecliptic

Chapter 2 Section 2.2: Ancient Astronomy

2.2 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.

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 1). 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.

Earth's Round Shadow.



Figure 1. 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 Earthcentered view, which would dominate Western thinking for nearly two millennia.

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:

 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.
- 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 <u>Figure 2</u>. 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. 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 3). 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.



How Eratosthenes Measured the Size of Earth.

Figure 3. 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 4).



Figure 4. Just as the axis of a rapidly spinning top wobbles slowly in

54 | Chapter 2 Section 2.2: Ancient Astronomy

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 the Roman last great astronomer of 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 5 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.

Retrograde Motion of a Planet beyond Earth's Orbit.



Figure 5. 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.

This retrograde simulation of Mars illustrates the motion of Mars as seen from Earth as well as Earth's retrograde motion as seen from Mars. There is also an animation of the movement of the two planets relative to each other that creates the appearance of this motion. 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 5, 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 6). When the planet is at position x in Figure 6 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.



Ptolemy's Complicated Cosmological System.

Figure 6. 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.

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.

Ancient Greeks such as Aristotle recognized that Earth and the Moon are spheres, and understood the phases of the Moon, but because of their inability to detect stellar parallax, they rejected the idea that Earth moves. Eratosthenes measured the size of Earth with surprising precision. Hipparchus carried out many astronomical observations, making a star catalog, defining the system of stellar magnitudes, and discovering precession from the apparent shift in the position of the north celestial pole. Ptolemy of Alexandria summarized classic astronomy in his *Almagest*; he explained planetary motions, including retrograde motion, with remarkably good accuracy using a model centered on Earth. This geocentric model, based on combinations of uniform circular motion using epicycles, was accepted as authority for more than a thousand years.

Glossary

apparent magnitude

a measure of how bright a star looks in the sky; the larger the

number, the dimmer the star appears to us

cosmology

the study of the organization and evolution of the universe

epicycle

the circular orbit of a body in the Ptolemaic system, the center of which revolves about another circle (the deferent)

parallax

the apparent displacement of a nearby star that results from the motion of Earth around the Sun

precession (of Earth)

the slow, conical motion of Earth's axis of rotation caused principally by the gravitational pull of the Moon and Sun on Earth's equatorial bulge

retrograde motion

the apparent westward motion of a planet on the celestial sphere or with respect to the stars

Chapter 2 Section 2.3: Astrology and Astronomy

2.3 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-anddeath 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 1), 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.

Zodiac Signs.



Figure 1. 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, then 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.

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 schools 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.

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.

The ancient religion of astrology, with its main contribution to civilization a heightened interest in the heavens, began in Babylonia. It reached its peak in the Greco-Roman world, especially as recorded in the *Tetrabiblos* of Ptolemy. Natal astrology is based on the assumption that the positions of the planets at the time of our birth, as described by a horoscope, determine our future. However, modern tests clearly show that there is no evidence for this, even in a broad statistical sense, and there is no verifiable theory to explain what might cause such an astrological influence.

Glossary

astrology

the pseudoscience that deals with the supposed influences on human destiny of the configurations and locations in the sky of the Sun, Moon, and planets

horoscope

a chart used by astrologers that shows the positions along the zodiac and in the sky of the Sun, Moon, and planets at some given instant and as seen from a particular place on Earth-usually corresponding to the time and place of a person's birth

Chapter 2 Section 2.4: The Birth of Modern Astronomy

2.4 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 1).

Nicolaus Copernicus (1473–1543).



Figure 1. 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 Suncentered, 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).

Copernicus' System.



Figure 2. 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 3). 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.

Phases of Venus.



Figure 3. 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.

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 (<u>Figure 4</u>), 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.

Galileo Galilei (1564–1642).

Figure 4. 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.

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 5).

Telescope Used by Galileo.



Figure 5. 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 3.) 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.

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.

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.

Nicolaus Copernicus introduced the heliocentric cosmology to Renaissance Europe in his book De Revolutionibus. Although he retained the Aristotelian idea of uniform circular motion. Copernicus suggested that Earth is a planet and that the planets all circle about the Sun, dethroning Earth from its position at the center of the universe. Galileo was the father of both modern experimental physics and telescopic astronomy. He studied the acceleration of moving objects and, in 1610, began telescopic observations, discovering the nature of the Milky Way, the largescale features of the Moon, the phases of Venus, and four moons of Jupiter. Although he was accused of heresy for his support of heliocentric cosmology, Galileo is credited with observations and brilliant writings that convinced most of his scientific contemporaries of the reality of the Copernican theory.

For Further Exploration

Articles

Ancient Astronomy

Gingerich, O. "From Aristarchus to Copernicus." Sky & Telescope (November 1983): 410.

Gingerich, O. "Islamic Astronomy." Scientific American (April 1986): 74.

Astronomy and Astrology

Fraknoi, A. "Your Astrology Defense Kit." Sky & Telescope (August 1989): 146.

Copernicus and Galileo

Gingerich, O. "Galileo and the Phases of Venus." Sky & Telescope (December 1984): 520.

Gingerich, O. "How Galileo Changed the Rules of Science." Sky & Telescope (March 1993): 32.

Maran, S., and Marschall, L. "The Moon, the Telescope, and the Birth of the Modern World." Sky & Telescope (February 2009): 28.

Sobel, D. "The Heretic's Daughter: A Startling Correspondence Reveals a New Portrait of Galileo." *The New Yorker* (September 13, 1999): 52.

Websites

Ancient Astronomy

Aristarchos of Samos: http://adsabs.harvard.edu//full/seri/ JRASC/0075//0000029.000.html. By Dr. Alan Batten.

Claudius Ptolemy: http://www-history.mcs.st-and.ac.uk/ Biographies/Ptolemy.html. An interesting biography.

Hipparchus of Rhodes: http://www-history.mcs.standrews.ac.uk/Biographies/Hipparchus.html. An interesting biography.

Astronomy and Astrology

Astrology and Science: http://www.astrology-and-science.com/ hpage.htm. The best site for a serious examination of the issues with astrology and the research on whether it works.

Real Romance in the Stars: http://www.independent.co.uk/ voices/the-real-romance-in-the-stars-1527970.html. 1995 newspaper commentary attacking astrology.

Copernicus and Galileo

Galileo Galilei: http://www-history.mcs.st-andrews.ac.uk/ Biographies/Galileo.html. A good biography with additional links.

Galileo Project: http://galileo.rice.edu/. Rice University's repository of information on Galileo.

Nicolaus Copernicus: http://www-groups.dcs.stand.ac.uk/~history/Biographies/Copernicus.html. A biography including links to photos about his life.

```
Videos
```

Astronomy and Astrology

Astrology Debunked: https://www.youtube.com/ watch?v=y84HX2pMo5U. A compilation of scientists and magicians commenting skeptically on astrology (9:09).

Copernicus and Galileo

Galileo: http://www.biography.com/people/galileo-9305220. A brief biography (2:51).

Galileo's Battle for the Heavens: https://www.youtube.com/ watch?v=jvlr2iMWQyc. A NOVA episode on PBS (1:48:55)

Nicolaus Copernicus: http://www.biography.com/people/ nicolaus-copernicus-9256984. An overview of his life and work (2:41).

Collaborative Group Activities

- A. With your group, consider the question with which we began this chapter. How many ways can you think of to prove to a member of the "Flat Earth Society" that our planet is, indeed, round?
- B. Make a list of ways in which a belief in astrology (the notion that your life path or personality is controlled by the position of the Sun, Moon, and planets at the time of your birth) might be harmful to an individual or to society at large.
- C. Have members of the group compare their experiences with the night sky. Did you see the Milky Way? Can you identify any constellations? Make a list of reasons why you think so many

fewer people know the night sky today than at the time of the ancient Greeks. Discuss reasons for why a person, today, may want to be acquainted with the night sky.

- D. Constellations commemorate great heroes, dangers, or events in the legends of the people who name them. Suppose we had to start from scratch today, naming the patterns of stars in the sky. Whom or what would you choose to commemorate by naming a constellation after it, him, or her and why (begin with people from history; then if you have time, include living people as well)? Can the members of your group agree on any choices?
- E. Although astronomical mythology no longer holds a powerful sway over the modern imagination, we still find proof of the power of astronomical images in the number of products in the marketplace that have astronomical names. How many can your group come up with? (Think of things like Milky Way candy bars, Eclipse and Orbit gum, or Comet cleanser.)

Review Questions

- From where on Earth could you observe all of the stars during the course of a year? What fraction of the sky can be seen from the North Pole?
- Give four ways to demonstrate that Earth is spherical.
- Explain, according to both geocentric and heliocentric cosmologies, why we see retrograde motion of the planets.
- In what ways did the work of Copernicus and Galileo differ from the views of the ancient Greeks and of their contemporaries?
- What were four of Galileo's discoveries that were important to

astronomy?

- Explain the origin of the magnitude designation for determining the brightness of stars. Why does it seem to go backward, with smaller numbers indicating brighter stars?
- Ursa Minor contains the pole star, Polaris, and the asterism known as the Little Dipper. From most locations in the Northern Hemisphere, all of the stars in Ursa Minor are circumpolar. Does that mean these stars are also above the horizon during the day? Explain.
- How many degrees does the Sun move per day relative to the fixed stars? How many days does it take for the Sun to return to its original location relative to the fixed stars?
- How many degrees does the Moon move per day relative to the fixed stars? How many days does it take for the Moon to return to its original location relative to the fixed stars?
- Explain how the zodiacal constellations are different from the other constellations.
- The Sun was once thought to be a planet. Explain why.
- Is the ecliptic the same thing as the celestial equator? Explain.
- What is an asterism? Can you name an example?
- Why did Pythagoras believe that Earth should be spherical?
- How did Aristotle deduce that the Sun is farther away from Earth than the Moon?
- What are two ways in which Aristotle deduced that Earth is spherical?
- 92 | Chapter 2 Section 2.4: The Birth of Modern Astronomy

- How did Hipparchus discover the wobble of Earth's axis, known as *precession*?
- Why did Ptolemy have to introduce multiple circles of motion for the planets instead of a single, simple circle to represent the planet's motion around the Sun?
- Why did Copernicus want to develop a completely new system for predicting planetary positions? Provide two reasons.
- What two factors made it difficult, at first, for astronomers to choose between the Copernican heliocentric model and the Ptolemaic geocentric model?
- What phases would Venus show if the geocentric model were correct?

Thought Questions

- Describe a practical way to determine in which constellation the Sun is found at any time of the year.
- What is a constellation as astronomers define it today? What does it mean when an astronomer says, "I saw a comet in Orion last night"?
- Draw a picture that explains why Venus goes through phases the way the Moon does, according to the heliocentric cosmology. Does Jupiter also go through phases as seen from Earth? Why?
- Show with a simple diagram how the lower parts of a ship disappear first as it sails away from you on a spherical Earth. Use the same diagram to show why lookouts on old sailing

ships could see farther from the masthead than from the deck. Would there be any advantage to posting lookouts on the mast if Earth were flat? (Note that these nautical arguments for a spherical Earth were quite familiar to Columbus and other mariners of his time.)

- Parallaxes of stars were not observed by ancient astronomers. How can this fact be reconciled with the heliocentric hypothesis?
- Why do you think so many people still believe in astrology and spend money on it? What psychological needs does such a belief system satisfy?
- Consider three cosmological perspectives—the geocentric perspective, the heliocentric perspective, and the modern perspective—in which the Sun is a minor star on the outskirts of one galaxy among billions. Discuss some of the cultural and philosophical implications of each point of view.
- The north celestial pole appears at an altitude above the horizon that is equal to the observer's latitude. Identify Polaris, the North Star, which lies very close to the north celestial pole. Measure its altitude. (This can be done with a protractor. Alternatively, your fist, extended at arm's length, spans a distance approximately equal to 10°.) Compare this estimate with your latitude. (Note that this experiment cannot be performed easily in the Southern Hemisphere because Polaris itself is not visible in the south and no bright star is located near the south celestial pole.)
- What were two arguments or lines of evidence in support of the geocentric model?
- Although the Copernican system was largely correct to place the Sun at the center of all planetary motion, the model still
gave inaccurate predictions for planetary positions. Explain the flaw in the Copernican model that hindered its accuracy.

- During a retrograde loop of Mars, would you expect Mars to be brighter than usual in the sky, about average in brightness, or fainter than usual in the sky? Explain.
- The Great Pyramid of Giza was constructed nearly 5000 years ago. Within the pyramid, archaeologists discovered a shaft leading from the central chamber out of the pyramid, oriented for favorable viewing of the bright star Thuban at that time. Thinking about Earth's precession, explain why Thuban might have been an important star to the ancient Egyptians.
- Explain why more stars are circumpolar for observers at higher latitudes.
- What is the altitude of the north celestial pole in the sky from your latitude? If you do not know your latitude, look it up. If you are in the Southern Hemisphere, answer this question for the south celestial pole, since the north celestial pole is not visible from your location.
- If you were to drive to some city south of your current location, how would the altitude of the celestial pole in the sky change?
- Hipparchus could have warned us that the dates associated with each of the natal astrology sun signs would eventually be wrong. Explain why.
- Explain three lines of evidence that argue against the validity of astrology.
- What did Galileo discover about the planet Jupiter that cast doubt on exclusive geocentrism?

• What did Galileo discover about Venus that cast doubt on geocentrism?

Figuring for Yourself

- Suppose Eratosthenes had found that, in Alexandria, at noon on the first day of summer, the line to the Sun makes an angle 30° with the vertical. What, then, would he have found for Earth's circumference?
- Suppose Eratosthenes' results for Earth's circumference were quite accurate. If the diameter of Earth is 12,740 km, what is the length of his stadium in kilometers?
- Suppose you are on a strange planet and observe, at night, that the stars do not rise and set, but circle parallel to the horizon. Next, you walk in a constant direction for 8000 miles, and at your new location on the planet, you find that all stars rise straight up in the east and set straight down in the west, perpendicular to the horizon. How could you determine the circumference of the planet without any further observations? What is the circumference, in miles, of the planet?

Glossary

accelerate

to change velocity; to speed up, slow down, or change direction.

heliocentric

centered on the Sun

Chapter 3 Orbits and Gravity Section 3.1: The Laws of Planetary Motion

International Space Station.



Figure 1. This space habitat and laboratory orbits Earth once every 90 minutes. (credit: modification of work by NASA)

How would you find a new planet at the outskirts of our solar system that is too dim to be seen with the unaided eye and is so far away that it moves very slowly among the stars? This was the problem confronting astronomers during the nineteenth century as they tried to pin down a full inventory of our solar system.

If we could look down on the solar system from somewhere out in space, interpreting planetary motions would be much simpler. But the fact is, we must observe the positions of all the other planets from our own moving planet. Scientists of the Renaissance did not know the details of Earth's motions any better than the motions of the other planets. Their problem, as we saw in Observing the Sky: The Birth of Astronomy, was that they had to deduce the nature of all planetary motion using only their earthbound observations of the other planets' positions in the sky. To solve this complex problem more fully, better observations and better models of the planetary system were needed.

3.1 The 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 1). Brahe was the last and greatest of the pre-telescopic observers in Europe.

Tycho Brahe (1546–1601) and Johannes Kepler (1571–1630).





(b)

Figure 1. (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 1). 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 2.).

Conic Section



Figure 2. 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). 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.

Drawing an Ellipse.



(a)

(b)

Figure 3. (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 2*a* 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 4), 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.

Kepler's Second Law: The Law of Equal Areas.



Figure 4. 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.

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.

106 | Chapter 3 Orbits and Gravity Section 3.1: The Laws of Planetary Motion

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 (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.

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:

$$P = \sqrt{a^3}$$

 $P = \sqrt{50 \times 50 \times 50} = \sqrt{125,000} = 353.6$ years

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

Check Your Learning

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

ANSWER:

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

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.

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 $a^3 = 0.72 \times 0.72 \times 0.72 = 0.37$ (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.

Check Your Learning

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:

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.

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." To learn more about Johannes Kepler's life and his laws of planetary motion, as well as lots of information on the Kepler Mission, visit NASA's Kepler website and follow the links that interest you.

Key Concepts and Summary

Tycho Brahe's accurate observations of planetary positions provided the data used by Johannes Kepler to derive his three fundamental laws of planetary motion. Kepler's laws describe the behavior of planets in their orbits as follows: (1) planetary orbits are ellipses with the Sun at one focus; (2) in equal intervals, a planet's orbit sweeps out equal areas; and (3) the relationship between the orbital period (P) and the semimajor axis (*a*) of an orbit is given by $P^2 = a^3$ (when *a* is in units of AU and P is in units of Earth years).

Glossary

astronomical unit (AU)

the unit of length defined as the average distance between Earth and the Sun; this distance is about 1.5×10^8 kilometers

eccentricity

in an ellipse, the ratio of the distance between the foci to the major axis

ellipse

a closed curve for which the sum of the distances from any point on the ellipse to two points inside (called the foci) is always the same

focus

(plural: foci) one of two fixed points inside an ellipse from which the sum of the distances to any point on the ellipse is constant

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

major axis

the maximum diameter of an ellipse

orbit

the path of an object that is in revolution about another object or point

orbital period (P)

the time it takes an object to travel once around the Sun

orbital speed

the speed at which an object (usually a planet) orbits around the mass of another object; in the case of a planet, the speed at which each planet moves along its ellipse

semimajor axis

half of the major axis of a conic section, such as an ellipse

Chapter 3 Section 3.2: Newton's Great Synthesis

3.2 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 1). 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.

Isaac Newton (1643–1727), 1689 Portrait by Sir Godfrey Kneller.



Figure 1. 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 complexities Eventually, the involved. 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 2).

Demonstrating Newton's Third Law.



Figure 2. 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)

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 cm³, its density is 3 g/ cm³. Familiar materials span a considerable range in density, from artificial materials such as plastic insulating foam (less than 0.1 g/ cm³) to gold (19.3 g/cm³). Table 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/cm³) to a collapsed "star corpse" called a neutron star (10¹⁵ g/cm³).

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*.

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.). 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.

Conservation of Angular Momentum.



Figure 3. 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.

Key Concepts and Summary

In his *Principia*, Isaac Newton established the three laws that govern the motion of objects: (1) objects continue to be at rest or move with a constant velocity unless acted upon by an outside force; (2) an outside force causes an acceleration (and changes the momentum) for an object; and (3) for every action there is an equal and opposite reaction. Momentum is a measure of the motion of an object and depends on both its mass and its velocity. Angular momentum is a measure of the motion of a spinning or revolving object and depends on its mass, velocity, and distance from the point around which it revolves. The density of an object is its mass divided by its volume.

Footnotes

• 1 Generally we use standard metric (or SI) units in this book. The proper metric unit of density in that system is kg/m³. But to most people, g/cm³ provides a more meaningful unit because the density of water is exactly 1 g/cm³, and this is useful information for comparison. Density expressed in g/ cm³ is sometimes called specific density or specific weight.

Glossary

angular momentum

the measure of the motion of a rotating object in terms of its speed and how widely the object's mass is distributed around its axis

density

the ratio of the mass of an object to its volume

momentum

the measure of the amount of motion of a body; the momentum of a body is the product of its mass and velocity; in the absence of an unbalanced force, momentum is conserved

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)

velocity

the speed and direction a body is moving–for example, 44 kilometers per second toward the north galactic pole

Chapter 3 Section 3.3: Newton's Universal Law of Gravitation

3.3 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 $(1/2)^2$, or 1/4 as large. Put the planet three times farther away, and the force is $(1/3)^2$, or 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:

$$F_{gravity} = G \frac{M_1 M_2}{R^2}$$

where Fgravity 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 m/s²).

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 m/s². 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.

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 $(1/2)^2 = 1/4$, so a person would weigh only one-fourth as much.

Check Your Learning

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:

With one-third its present mass, the gravitational force at the

surface would reduce by a factor of 1/3, so a person would weight only one-third as much.

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 1).¹
Astronauts in Free Fall.



Figure 1. While in space, astronauts are falling freely, so they experience "weightlessness." Clockwise from top left: Tracy Caldwell Dyson (NASA), Naoko Yamzaki (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 (M₁) and the planet (M₂), 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.

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.)

Check Your Learning

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:

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.

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.

Key Concepts and Summary

Gravity, the attractive force between all masses, is what keeps the planets in orbit. Newton's universal law of gravitation relates the gravitational force to mass and distance:

$$F_{gravity} = G \frac{M_1 M_2}{R^2}$$

The force of gravity is what gives us our sense of weight. Unlike mass, which is constant, weight can vary depending on the force of gravity (or acceleration) you feel. When Kepler's laws are reexamined in the light of Newton's gravitational law, it becomes clear that the masses of both objects are important for the third law, which becomes $a^3 = (M_1 + M_2) \times P^2$. Mutual gravitational effects permit us to calculate the masses of astronomical objects, from comets to galaxies.

Footnotes

• 1 In 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.

Glossary

gravity

the mutual attraction of material bodies or particles

For Further Exploration

Articles

Brahe and Kepler

Christianson, G. "The Celestial Palace of Tycho Brahe." Scientific American (February 1961): 118.

Gingerich, O. "Johannes Kepler and the Rudolphine Tables." Sky & Telescope (December 1971): 328. Brief article on Kepler's work.

Wilson, C. "How Did Kepler Discover His First Two Laws?" Scientific American (March 1972): 92.

Newton

Christianson, G. "Newton's Principia: A Retrospective." Sky & Telescope (July 1987): 18.

Cohen, I. "Newton's Discovery of Gravity." Scientific American (March 1981): 166.

Gingerich, O. "Newton, Halley, and the Comet." Sky & Telescope (March 1986): 230.

Sullivant, R. "When the Apple Falls." Astronomy (April 1998): 55. Brief overview.

The Discovery of Neptune

Sheehan, W., et al. "The Case of the Pilfered Planet: Did the British Steal Neptune?" *Scientific American* (December 2004): 92.

Websites

Brahe and Kepler

Johannes Kepler: His Life, His Laws, and Time: http://kepler.nasa.gov/Mission/JohannesKepler/. From NASA's Kepler mission.

Johannes Kepler: http://www.britannica.com/biography/ Johannes-Kepler. Encyclopedia Britannica article.

Johannes Kepler: http://www-history.mcs.st-andrews.ac.uk/ Biographies/Kepler.html. MacTutor article with additional links. Noble Dane: Images of Tycho Brahe: http://www.mhs.ox.ac.uk/ tycho/index.htm. A virtual museum exhibit from Oxford.

Newton

Sir Isaac Newton: http://www-groups.dcs.stand.ac.uk/~history//Biographies/Newton.html. MacTutor article with additional links.

Sir Isaac Newton: http://www.luminarium.org/sevenlit/ newton/newtonbio.htm. Newton Biography at the Luminarium.

The Discovery of Neptune

Adams,Airy,andtheDiscoveryofNeptune: http://www.mikeoates.org/lassell/adams-airy.htm.Adefense of Airy's role by historian Alan Chapman.

Mathematical Discovery of Planets: http://www-groups.dcs.stand.ac.uk/~history/HistTopics/Neptune_and_Pluto.html. MacTutor article.

Videos

Brahe and Kepler

"Harmony of the Worlds." This third episode of Carl Sagan's TV series Cosmos focuses on Kepler and his life and work.

Tycho Brahe, Johannes Kepler, and Planetary Motion: https://www.youtube.com/watch?v=x3ALuycrCwI. German-produced video, in English (14:27).

Newton

Beyond the Big Bang: Sir Isaac Newton's Law of Gravity: http://www.history.com/topics/enlightenment/videos/ beyond-the-big-bang-sir-isaac-newtons-law-of-gravity. From the History Channel (4:35).

Sir Isaac Newton versus Bill Nye: Epic Rap Battles of History: https://www.youtube.com/watch?v=8yis7GzlXNM. (2:47).

The Discovery of Neptune

Richard Feynman: On the Discovery of Neptune: https://www.youtube.com/watch?v=FgXQffVgZRs. A brief black-and-white Caltech lecture (4:33).

Collaborative Group Activities

- A. An eccentric, but very rich, alumnus of your college makes a bet with the dean that if you drop a baseball and a bowling ball from the tallest building on campus, the bowling ball would hit the ground first. Have your group discuss whether you would make a side bet that the alumnus is right. How would you decide who is right?
- B. Suppose someone in your astronomy class was unhappy about his or her weight. Where could a person go to weigh onefourth as much as he or she does now? Would changing the unhappy person's weight have any effect on his or her mass?
- C. When the Apollo astronauts landed on the Moon, some commentators commented that it ruined the mystery and "poetry" of the Moon forever (and that lovers could never gaze at the full moon in the same way again). Others felt that knowing more about the Moon could only enhance its interest

to us as we see it from Earth. How do the various members of your group feel? Why?

- D. [link] shows a swarm of satellites in orbit around Earth. What do you think all these satellites do? How many categories of functions for Earth satellites can your group come up with?
- E. The Making Connections feature box Astronomy and the Poets discusses how poets included the most recent astronomical knowledge in their poetry. Is this still happening today? Can your group members come up with any poems or songs that you know that deal with astronomy or outer space? If not, perhaps you could find some online, or by asking friends or roommates who are into poetry or music.

Review Questions

State Kepler's three laws in your own words.

Why did Kepler need Tycho Brahe's data to formulate his laws?

Which has more mass: an armful of feathers or an armful of lead? Which has more volume: a kilogram of feathers or a kilogram of lead? Which has higher density: a kilogram of feathers or a kilogram of lead?

Explain how Kepler was able to find a relationship (his third law) between the orbital periods and distances of the planets that did not depend on the masses of the planets or the Sun.

Write out Newton's three laws of motion in terms of what happens with the momentum of objects.

Which major planet has the largest . . .

- A. semimajor axis?
- B. average orbital speed around the Sun?

Chapter 3 Section 3.3: Newton's Universal Law of Gravitation | 139

- C. orbital period around the Sun?
- D. eccentricity?

Why do we say that Neptune was the first planet to be discovered through the use of mathematics?

Why was Brahe reluctant to provide Kepler with all his data at one time?

According to Kepler's second law, where in a planet's orbit would it be moving fastest? Where would it be moving slowest?

The gas pedal, the brakes, and the steering wheel all have the ability to accelerate a car-how?

Explain how a rocket can propel itself using Newton's third law.

A certain material has a mass of 565 g while occupying 50 cm^3 of space. What is this material? (Hint: Use [link].)

To calculate the momentum of an object, which properties of an object do you need to know?

To calculate the angular momentum of an object, which properties of an object do you need to know?

What was the great insight Newton had regarding Earth's gravity that allowed him to develop the universal law of gravitation?

Which of these properties of an object best quantifies its inertia: velocity, acceleration, volume, mass, or temperature?

Pluto's orbit is more eccentric than any of the major planets. What does that mean?

Why is Tycho Brahe often called "the greatest naked-eye astronomer" of all time?

Thought Questions

Is it possible to escape the force of gravity by going into orbit around Earth? How does the force of gravity in the International Space Station (orbiting an average of 400 km above Earth's surface) compare with that on the ground?

What is the momentum of an object whose velocity is zero? How does Newton's first law of motion include the case of an object at rest?

Evil space aliens drop you and your fellow astronomy student 1 km apart out in space, very far from any star or planet. Discuss the effects of gravity on each of you.

A body moves in a perfectly circular path at constant speed. Are there forces acting in such a system? How do you know?

As friction with our atmosphere causes a satellite to spiral inward, closer to Earth, its orbital speed increases. Why?

Use a history book, an encyclopedia, or the internet to find out what else was happening in England during Newton's lifetime and discuss what trends of the time might have contributed to his accomplishments and the rapid acceptance of his work.

Two asteroids begin to gravitationally attract one another. If one asteroid has twice the mass of the other, which one experiences the greater force? Which one experiences the greater acceleration?

How does the mass of an astronaut change when she travels from Earth to the Moon? How does her weight change?

If there is gravity where the International Space Station (ISS) is located above Earth, why doesn't the space station get pulled back down to Earth? Compare the density, weight, mass, and volume of a pound of gold to a pound of iron on the surface of Earth.

If identical spacecraft were orbiting Mars and Earth at identical radii (distances), which spacecraft would be moving faster? Why?

Figuring for Yourself

By what factor would a person's weight be increased if Earth had 10 times its present mass, but the same volume?

Suppose astronomers find an earthlike planet that is twice the size of Earth (that is, its radius is twice that of Earth's). What must be the mass of this planet such that the gravitational force (F_{gravity}) at the surface would be identical to Earth's?

What is the semimajor axis of a circle of diameter 24 cm? What is its eccentricity?

If 24 g of material fills a cube 2 cm on a side, what is the density of the material?

If 128 g of material is in the shape of a brick 2 cm wide, 4 cm high, and 8 cm long, what is the density of the material?

If the major axis of an ellipse is 16 cm, what is the semimajor axis? If the eccentricity is 0.8, would this ellipse be best described as mostly circular or very elongated?

What is the average distance from the Sun (in astronomical units) of an asteroid with an orbital period of 8 years?

What is the average distance from the Sun (in astronomical units) of a planet with an orbital period of 45.66 years?

In 1996, astronomers discovered an icy object beyond Pluto that was

142 | Chapter 3 Section 3.3: Newton's Universal Law of Gravitation

given the designation 1996 TL 66. It has a semimajor axis of 84 AU. What is its orbital period according to Kepler's third law?

Glossary

perturbation

a small disturbing effect on the motion or orbit of a body produced by a third body

Chapter 4 Earth, Moon and Sky Section 4.2: The Seasons

Southern Summer.



Figure 1. As captured with a fish-eye lens aboard the Atlantis Space Shuttle on December 9, 1993, Earth hangs above the Hubble Space Telescope as it is repaired. The reddish continent is Australia, its size and shape distorted by the special lens. Because the seasons in the Southern Hemisphere are opposite those in the Northern Hemisphere, it is summer in Australia on this December day. (credit: modification of work by NASA) If Earth's orbit is nearly a perfect circle (as we saw in earlier chapters), why is it hotter in summer and colder in winter in many

chapters), why is it hotter in summer and colder in winter in many places around the globe? And why are the seasons in Australia or Peru the opposite of those in the United States or Europe?

The story is told that Galileo, as he left the Hall of the Inquisition following his retraction of the doctrine that Earth rotates and revolves about the Sun, said under his breath, "But nevertheless it moves." Historians are not sure whether the story is true, but certainly Galileo knew that Earth was in motion, whatever church authorities said. It is the motions of Earth that produce the seasons and give us our measures of time and date. The Moon's motions around us provide the concept of the month and the cycle of lunar phases. In this chapter we examine some of the basic phenomena of our everyday world in their astronomical context.

4.2 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 1 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.

Seasons.



Figure 1. 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 2). 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.

The Sun's Rays in Summer and Winter.



Figure 2. (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 3). 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.

The Sun's Path in the Sky for Different Seasons.



Figure 3. 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 [link]). 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 3. 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.

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. 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 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, $90^{\circ} - 23^{\circ}$ (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*.

Earth on June 21.



Figure 4. 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. 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 5. 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.



Earth on December 21.

Figure 5. 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.

Seasonal Variations

As you can see in Figure 4, 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 $90^{\circ} - 23^{\circ} = 67^{\circ}$ 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 $90^\circ - 5^\circ = 85^\circ$ N of the equator.

Check Your Learning

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:

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 $74^{\circ} - 16^{\circ} = 58^{\circ}$. 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.

You can see an animation of the Sun's path during the seasons alongside a time-lapse view of light and shadow from a camera set up on the University of Nebraska campus.

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.

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.

Check Your Learning

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

ANSWER:

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°.

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.

Key Concepts and Summary

The familiar cycle of the seasons results from the 23.5° tilt of Earth's axis of rotation. At the summer solstice, the Sun is higher in the sky and its rays strike Earth more directly. The Sun is in the sky for more than half of the day and can heat Earth longer. At the winter solstice, the Sun is low in the sky and its rays come in at more of an angle; in addition, it is up for fewer than 12 hours, so those rays have less time to heat. At the vernal and autumnal equinoxes, the Sun is on

the celestial equator and we get about 12 hours of day and night. The seasons are different at different latitudes.

Footnotes

1 Note 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.

Chapter 4 Section 4.5: Phases and Motions of the Moon

4.5 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 1, 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 1, 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.



Phases of the Moon.

Figure 1. 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 onequarter 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 1), 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 <u>Figure 1</u>. 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.

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."

View this animation to see the phases of the Moon as it orbits Earth and as Earth orbits the Sun.

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 2). 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**.



The Moon without and with Rotation.

(b)

Figure 2. 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."

Key Concepts and Summary

The Moon's monthly cycle of phases results from the changing angle of its illumination by the Sun. The full moon is visible in the sky only during the night; other phases are visible during the day as well. Because its period of revolution is the same as its period of rotation, the Moon always keeps the same face toward Earth.

Glossary

phases of the Moon

the different appearance of light and dark on the Moon as seen from Earth during its monthly cycle, from new moon to full moon and back to new moon

sidereal month

the period of the Moon's revolution about Earth measured with respect to the stars

solar month

the time interval in which the phases repeat-say, from full to

full phase

synchronous rotation

when a body (for example, the Moon) rotates at the same rate that it revolves around another body

Chapter 4 Section 4.7: Eclipses of the Sun and Moon

4.7 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 1 illustrates the appearance of the Moon's shadow and what the Sun and Moon would look like from different points within the shadow.

Solar Eclipse.





Figure 1. (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 1 shows the distance of the observer varying at points A–D, 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 2. 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.

Geometry of a Total Solar Eclipse.



Figure 2. 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 will be total eclipses visible in the United States in 2017 and 2024, to which even 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 3). 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.)

The Sun's Corona.



Figure 3. 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 glasses¹ 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 <u>Figure 4</u>. 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

176 | Chapter 4 Section 4.7: Eclipses of the Sun and Moon

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.

Geometry of a Lunar Eclipse.



Figure 4. 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.

HOW TO OBSERVE SOLAR ECLIPSES

A total eclipse of the Sun is a spectacular sight and should not be

178 | Chapter 4 Section 4.7: Eclipses of the Sun and Moon

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 (<u>Figure 5</u> <u>a</u>).

2017 Total Solar and Lunar Eclipse



Figure 5. (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.)

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.

Key Concepts and Summary

The Sun and Moon have nearly the same angular size (about 1/2°). A solar eclipse occurs when the Moon moves between the Sun and Earth, casting its shadow on a part of Earth's surface. If the eclipse is total, the light from the bright disk of the Sun is completely blocked, and the solar atmosphere (the corona) comes into view. Solar eclipses take place rarely in any one location, but they are among the most spectacular sights in nature. A lunar eclipse takes place when the Moon moves into Earth's shadow; it is visible (weather permitting) from the entire night hemisphere of Earth.

For Further Exploration

Articles

Bakich, M. "Your Twenty-Year Solar Eclipse Planner." Astronomy (October 2008): 74. Describes the circumstances of upcoming total eclipses of the Sun.

Coco, M. "Not Just Another Pretty Phase." Astronomy (July 1994): 76. Moon phases explained.

Espenak, F., & Anderson, J. "Get Ready for America's Coast to Coast Experience." Sky & Telescope (February 2016): 22.

Gingerich, O. "Notes on the Gregorian Calendar Reform." Sky & Telescope (December 1982): 530.

Kluepfel, C. "How Accurate Is the Gregorian Calendar?" Sky & Telescope (November 1982): 417.

Krupp, E. "Calendar Worlds." Sky & Telescope (January 2001): 103. On how the days of the week got their names.

Krupp, E. "Behind the Curve." Sky & Telescope (September 2002): 68. On the reform of the calendar by Pope Gregory XIII.

MacRobert, A., & Sinnott, R. "Young Moon Hunting." Sky & Telescope (February 2005): 75. Hints for finding the Moon as soon after its new phase as possible.

Pasachoff, J. "Solar Eclipse Science: Still Going Strong." Sky & Telescope (February 2001): 40. On what we have learned and are still learning from eclipses.

Regas, D. "The Quest for Totality." Sky & Telescope (July 2012): 36. On eclipse chasing as a hobby.

Schaefer, B. "Lunar Eclipses That Changed the World." Sky & Telescope (December 1992): 639.

Schaefer, B. "Solar Eclipses That Changed the World." Sky & Telescope (May 1994): 36.

Websites

Ancient Observatories, Timeless Knowledge (Stanford Solar Center): http://solar-center.stanford.edu/AO/. An introduction to ancient sites where the movements of celestial objects were tracked over the years (with a special focus on tracking the Sun).

Astronomical Data Services: http://aa.usno.navy.mil/data/ index.php. This rich site from the U.S. Naval Observatory has information about Earth, the Moon, and the sky, with tables and online calculators.

Calendars through the Ages: http://www.webexhibits.org/ calendars/index.html. Like a good museum exhibit on the Web.

Calendar Zone: http://www.calendarzone.com/. Everything you wanted to ask or know about calendars and timekeeping, with links from around the world.

Eclipse 2017 Information and Safe Viewing Instructions: http://www.nsta.org/publications/press/extras/ files/solarscience/SolarScienceInsert.pdf.

Eclipse Maps: http://www.eclipse-maps.com/Eclipse-Maps/ Welcome.html. Michael Zeiler specializes in presenting helpful and interactive maps of where solar eclipses will be visible

Eclipse Predictions: http://astro.unl.edu/classaction/ animations/lunarcycles/eclipsetable.html. This visual calendar provides dates for upcoming solar and lunar eclipses through 2029.EclipseWise: http://www.eclipsewise.com/intro.html. An introductory site on future eclipses and eclipse observing by NASA's Fred Espenak.

HistoryoftheInternationalDateLine: http://www.staff.science.uu.nl/~gent0113/idl/idl.htm.FromR. H. van Gent at Utrecht University in the Netherlands.

Lunacy and the Full Moon: http://www.scientificamerican.com/ article/lunacy-and-the-full-moon/. This Scientific American article explores whether the Moon's phase is related to strange behavior.

Moon Phase Calculator: https://stardate.org/nightsky/moon. Keep track of the phases of the Moon with this calendar.

NASA Eclipse Website: http://eclipse.gsfc.nasa.gov/eclipse.html. This site, by NASA's eclipse expert Fred Espenak, contains a wealth of information on lunar and solar eclipses, past and future, as well as observing and photography links.

Phases of the Moon Gallery and Information: http://astropixels.com/moon/phases/

phasesgallery.html. Photographs and descriptions presented by NASA's Fred Espenak.

Time and Date Website: http://www.timeanddate.com/. Comprehensive resource about how we keep time on Earth; has time zone converters and many other historical and mathematical tools.

Walk through Time: The Evolution of Time Measurement through

the Ages (National Institute of Standards and Technology): http://www.nist.gov/pml/general/time/.

Videos

BillNye,theScienceGuy,ExplainstheSeasons: https://www.youtube.com/watch?v=KUU7IyfR34o.Forkids, but college students can enjoy the bad jokes, too (4:45).

Geography Lesson Idea: Time Zones: https://www.youtube.com/ watch?v=-j-SWKtWEcU. (3:11).

How to View a Solar Eclipse: http://www.exploratorium.edu/ eclipse/how-to-view-eclipse. (1:35).

Shadow of the Moon: https://www.youtube.com/ watch?v=XNcfKUJwnjM. This NASA video explains eclipses of the Sun, with discussion and animation, focusing on a 2015 eclipse, and shows what an eclipse looks like from space (1:54).

Strangest Time Zones in the World: https://www.youtube.com/ watch?v=uW6QqcmCfm8. (8:38).

Understanding Lunar Eclipses: https://www.youtube.com/ watch?v=lNi5UFpales. This NASA video explains why there isn't an eclipse every month, with good animation (1:58).

Collaborative Group Activities

- A. Have your group brainstorm about other ways (besides the Foucault pendulum) you could prove that it is our Earth that is turning once a day, and not the sky turning around us. (Hint: How does the spinning of Earth affect the oceans and the atmosphere?)
- B. What would the seasons on Earth be like if Earth's axis were not tilted? Discuss with your group how many things about life on Earth you think would be different.

- C. After college and graduate training, members of your U.S. student group are asked to set up a school in New Zealand. Describe some ways your yearly school schedule in the Southern Hemisphere would differ from what students are used to in the Northern Hemisphere.
- D. During the traditional U.S. Christmas vacation weeks, you are sent to the vicinity of the South Pole on a research expedition (depending on how well you did on your astronomy midterm, either as a research assistant or as a short-order cook!). Have your group discuss how the days and nights will be different there and how these differences might affect you during your stay.
- E. Discuss with your group all the stories you have heard about the full moon and crazy behavior. Why do members of your group think people associate crazy behavior with the full moon? What other legends besides vampire stories are connected with the phases of the Moon? (Hint: Think Professor Lupin in the Harry Potter stories, for example.)
- F. Your college town becomes the founding site for a strange new cult that worships the Moon. These true believers gather regularly around sunset and do a dance in which they must extend their arms in the direction of the Moon. Have your group discuss which way their arms will be pointing at sunset when the Moon is new, first quarter, full, and third quarter.
- G. Changes of the seasons play a large part in our yearly plans and concerns. The seasons have inspired music, stories, poetry, art, and much groaning from students during snowstorms. Search online to come up with some examples of the seasons being celebrated or overcome in fields other than science.
- H. Use the information in Appendix H and online to figure out when the next eclipse of the Sun or eclipse of the Moon will be visible from where your group is going to college or from where your group members live. What time of day will the eclipse be visible? Will it be a total or partial eclipse? What preparations can you make to have an enjoyable and safe

eclipse experience? How do these preparations differ between a solar and lunar eclipse?

I. On Mars, a day (often called a sol) is 24 hours and 40 minutes. Since Mars takes longer to go around the Sun, a year is 668.6 sols. Mars has two tiny moons, Phobos and Deimos. Phobos, the inner moon, rises in the west and sets in the east, taking 11 hours from moonrise to the next moonrise. Using your calculators and imaginations, have your group members come up with a calendar for Mars. (After you do your own, and only after, you can search online for the many suggestions that have been made for a martian calendar over the years.)

Review Questions

Discuss how latitude and longitude on Earth are similar to declination and right ascension in the sky.

What is the latitude of the North Pole? The South Pole? Why does longitude have no meaning at the North and South Poles?

Make a list of each main phase of the Moon, describing roughly when the Moon rises and sets for each phase. During which phase can you see the Moon in the middle of the morning? In the middle of the afternoon?

What are advantages and disadvantages of apparent solar time? How is the situation improved by introducing mean solar time and standard time?

What are the two ways that the tilt of Earth's axis causes the summers in the United States to be warmer than the winters?

Why is it difficult to construct a practical calendar based on the Moon's cycle of phases?

Explain why there are two high tides and two low tides each day. Strictly speaking, should the period during which there are two high tides be 24 hours? If not, what should the interval be?

What is the phase of the Moon during a total solar eclipse? During a total lunar eclipse?

On a globe or world map, find the nearest marked latitude line to your location. Is this an example of a great circle? Explain.

Explain three lines of evidence that indicate that the seasons in North America are not caused by the changing Earth-Sun distance as a result of Earth's elliptical orbit around the Sun.

What is the origin of the terms "a.m." and "p.m." in our timekeeping?

Explain the origin of the leap year. Why is it necessary?

Explain why the year 1800 was not a leap year, even though years divisible by four are normally considered to be leap years.

What fraction of the Moon's visible face is illuminated during first quarter phase? Why is this phase called first quarter?

Why don't lunar eclipses happen during every full moon?

Why does the Moon create tidal bulges on both sides of Earth instead of only on the side of Earth closest to the Moon?

Why do the heights of the tides change over the course of a month?

Explain how tidal forces are causing Earth to slow down.

Explain how tidal forces are causing the Moon to slowly recede from Earth.

Explain why the Gregorian calendar modified the nature of the leap year from its original definition in the Julian calendar.

The term *equinox* translates as "equal night." Explain why this translation makes sense from an astronomical point of view.

The term solstice translates as "Sun stop." Explain why this translation makes sense from an astronomical point of view.

Why is the warmest day of the year in the United States (or in the Northern Hemisphere temperate zone) usually in August rather than on the day of the summer solstice, in late June?

Thought Questions

When Earth's Northern Hemisphere is tilted toward the Sun during June, some would argue that the cause of our seasons is that the Northern Hemisphere is physically closer to the Sun than the Southern Hemisphere, and this is the primary reason the Northern Hemisphere is warmer. What argument or line of evidence could contradict this idea?

Where are you on Earth if you experience each of the following? (Refer to the discussion in Observing the Sky: The Birth of Astronomy as well as this chapter.)

- A. The stars rise and set perpendicular to the horizon.
- B. The stars circle the sky parallel to the horizon.
- C. The celestial equator passes through the zenith.
- D. In the course of a year, all stars are visible.
- E. The Sun rises on March 21 and does not set until September 21 (ideally).

In countries at far northern latitudes, the winter months tend to be so cloudy that astronomical observations are nearly impossible. Why can't good observations of the stars be made at those places during the summer months? What is the phase of the Moon if it . . .

- A. rises at 3:00 p.m.?
- B. is highest in the sky at sunrise?
- C. sets at 10:00 a.m.?

A car accident occurs around midnight on the night of a full moon. The driver at fault claims he was blinded momentarily by the Moon rising on the eastern horizon. Should the police believe him?

The secret recipe to the ever-popular veggie burgers in the college cafeteria is hidden in a drawer in the director's office. Two students decide to break in to get their hands on it, but they want to do it a few hours before dawn on a night when there is no Moon, so they are less likely to be caught. What phases of the Moon would suit their plans?

Your great-great-grandfather, who often exaggerated events in his own life, once told your relatives about a terrific adventure he had on February 29, 1900. Why would this story make you suspicious?

One year in the future, when money is no object, you enjoy your birthday so much that you want to have another one right away. You get into your supersonic jet. Where should you and the people celebrating with you travel? From what direction should you approach? Explain.

Suppose you lived in the crater Copernicus on the side of the Moon facing Earth.

- A. How often would the Sun rise?
- B. How often would Earth set?
- C. During what fraction of the time would you be able to see the stars?

In a lunar eclipse, does the Moon enter the shadow of Earth from the east or west side? Explain.

Describe what an observer at the crater Copernicus would see while the Moon is eclipsed on Earth. What would the same observer see during what would be a total solar eclipse as viewed from Earth?

The day on Mars is 1.026 Earth-days long. The martian year lasts 686.98 Earth-days. The two moons of Mars take 0.32 Earth-day (for Phobos) and 1.26 Earth-days (for Deimos) to circle the planet. You are given the task of coming up with a martian calendar for a new Mars colony. Would a solar or lunar calendar be better for tracking the seasons?

What is the right ascension and declination of the vernal equinox?

What is the right ascension and declination of the autumnal equinox?

What is the right ascension and declination of the Sun at noon on the summer solstice in the Northern Hemisphere?

During summer in the Northern Hemisphere, the North Pole is illuminated by the Sun 24 hours per day. During this time, the temperature often does not rise above the freezing point of water. Explain why.

On the day of the vernal equinox, the day length for all places on Earth is actually slightly longer than 12 hours. Explain why.

Regions north of the Arctic Circle are known as the "land of the midnight Sun." Explain what this means from an astronomical perspective.

In a part of Earth's orbit where Earth is moving faster than usual around the Sun, would the length of the sidereal day change? If so, how? Explain.

In a part of Earth's orbit where Earth is moving faster than usual around the Sun, would the length of the solar day change? If so, how? Explain.

190 | Chapter 4 Section 4.7: Eclipses of the Sun and Moon

If Sirius rises at 8:00 p.m. tonight, at what time will it rise tomorrow night, to the nearest minute? Explain.

What are three lines of evidence you could use to indicate that the phases of the Moon are not caused by the shadow of Earth falling on the Moon?

If the Moon rises at a given location at 6:00 p.m. today, about what time will it rise tomorrow night?

Explain why some solar eclipses are total and some are annular.

Why do lunar eclipses typically last much longer than solar eclipses?

Figuring for Yourself

Suppose Earth took exactly 300.0 days to go around the Sun, and everything else (the day, the month) was the same. What kind of calendar would we have? How would this affect the seasons?

Consider a calendar based entirely on the day and the month (the Moon's period from full phase to full phase). How many days are there in a month? Can you figure out a scheme analogous to leap year to make this calendar work?

If a star rises at 8:30 p.m. tonight, approximately what time will it rise two months from now?

What is the altitude of the Sun at noon on December 22, as seen from a place on the Tropic of Cancer?

Show that the Gregorian calendar will be in error by 1 day in about 3300 years.

Footnotes

• 1 Eclipse 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.

Glossary

lunar eclipse

an eclipse of the Moon, in which the Moon moves into the shadow of Earth; lunar eclipses can occur only at the time of full moon

solar eclipse

an eclipse of the Sun by the Moon, caused by the passage of the Moon in front of the Sun; solar eclipses can occur only at the time of the new moon

Chapter 5 Radiation and Spectra Section 5.1: The Behavior of Light

Thinking Ahead

Our Sun in Ultraviolet Light.



Figure 1. This photograph of the Sun was taken at several different wavelengths of ultraviolet, which our eyes cannot see, and then color coded so it reveals activity in our Sun's atmosphere that cannot be observed in visible light. This is why it is important to observe the Sun and other astronomical objects in wavelengths other than the visible band of the spectrum. This image was taken by a satellite from above Earth's atmosphere, which is necessary since Earth's atmosphere absorbs much of the ultraviolet light coming from space. (credit: modification of work by NASA) The nearest star is so far away that the fastest spacecraft humans have built would take almost 100,000 years to get there. Yet we very

much want to know what material this neighbor star is composed of and how it differs from our own Sun. How can we learn about the chemical makeup of stars that we cannot hope to visit or sample?

In astronomy, most of the objects that we study are completely beyond our reach. The temperature of the Sun is so high that a spacecraft would be fried long before it reached it, and the stars are much too far away to visit in our lifetimes with the technology now available. Even light, which travels at a speed of 300,000 kilometers per second (km/s), takes more than 4 years to reach us from the nearest star. If we want to learn about the Sun and stars, we must rely on techniques that allow us to analyze them from a distance.

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 1). 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.

James Clerk Maxwell (1831–1879).



Figure 1. 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 2). 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.

Making Waves.



Figure 2. 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 disturbances а family of possible electromagnetic 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 piece 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 nineteenthcentury 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 3×10^8 m/s), 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 3). 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.

Characterizing Waves.



Figure 3. 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 Yiolet. 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

Chapter 5 Radiation and Spectra Section 5.1: The Behavior of Light | 201

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:

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

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:

$$averagespeed = \frac{distance}{time}$$

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 $c = \lambda/t$. 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 = 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 × 10¹⁴ Hz?

Solution

Solving the wave equation for wavelength, we find:

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

Substituting our values gives:

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

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

Check Your Learning

"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:

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

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 (<u>Figure 4</u>). If

we stand twice as far from the source, our eyes will intercept twosquared (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.

Inverse Square Law for Light.



Figure 4. 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.

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 $(1/2)^2 = 1/4$ of the original intensity, or 0.6 W/m².

Check Your Learning

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:

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

Key Concepts and Summary

James Clerk Maxwell showed that whenever charged particles change their motion, as they do in every atom and molecule, they give off waves of energy. Light is one form of this electromagnetic radiation. The wavelength of light determines the color of visible radiation. Wavelength (λ) is related to frequency (f) and the speed of light (c) by the equation $c = \lambda f$. Electromagnetic radiation sometimes behaves like waves, but at other times, it behaves as if it were a particle–a little packet of energy, called a photon. The apparent brightness of a source of electromagnetic energy decreases with increasing distance from that source in proportion to the square of the distance–a relationship known as the inverse square law.

Glossary

electromagnetic radiation

radiation consisting of waves propagated through regularly varying electric and magnetic fields and traveling at the speed of light

frequency

the number of waves that cross a given point per unit time (in radiation)

inverse square law

(for light) the amount of energy (light) flowing through a given area in a given time decreases in proportion to the square of the distance from the source of energy or light

photon

a discrete unit (or "packet") of electromagnetic energy

wavelength

the distance from crest to crest or trough to trough in a wave

Chapter 5 Section 5.2: The Electromagnetic Spectrum

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 <u>Figure 1</u>, with some information about the waves in each part or band.

Radiation and Earth's Atmosphere.



Figure 1. 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

210 | Chapter 5 Section 5.2: The Electromagnetic Spectrum

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 2).

X-Ray Sky.



Figure 2. 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 accidentally 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 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.

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	$10^3 - 10^4 \text{ K}$	Stars
Infrared	10 ³ -10 ⁶	10–10 ³ К	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.

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 3. 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).



Radiation Laws Illustrated.

Figure 3. 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 3), 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 shorterwavelength 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 3. 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:

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

where the wavelength is in nanometers (one billionth of a meter) and the temperature is in K (the constant 3×10^{6} has units of nm \times K). This relationship is called **Wien's law**. For the Sun, the wavelength at which the maximum energy is emitted is 520 nanometers, which is near the middle of that portion of the electromagnetic spectrum called visible light. Characteristic temperatures of other astronomical objects, and the wavelengths at which they emit most of their power, are listed in Table.

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: $T = \frac{3 \times 10^6 nmK}{\lambda max} = \frac{3 \times 10^6 nmK}{1200 nm} = 2500k$

Check Your Learning

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

ANSWER:

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

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 and σ (Greek letter sigma) is a constant number (5.67 \times 10⁻⁸).

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 2⁴, 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.

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

 $L_A = 4\pi R_A{}^2\sigma T_A{}^4$ and $L_B = 4\pi R_B{}^2\sigma T_B{}^4$

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

 $\frac{L_B}{L_A} = \frac{4\pi R_B^2 \sigma T_B^4}{4\pi R_A^2 \sigma T_A^4} = \frac{R_B^2 T_B^4}{R_A^2 T_A^4}$ Because the two stars are the same size, R_A = R_B, leaving $\frac{T_B^4}{T_A^4} = \frac{(12,000K)^4}{(6,000K)^4} = 2^4 = 16$

Check Your Learning

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:

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

Key Concepts and Summary

The electromagnetic spectrum consists of gamma rays, X-rays, ultraviolet radiation, visible light, infrared, and radio radiation. Many of these wavelengths cannot penetrate the layers of Earth's atmosphere and must be observed from space, whereas others-such as visible light, FM radio and TV-can penetrate to Earth's surface. The emission of electromagnetic radiation is intimately connected to the temperature of the source. The higher the temperature of an idealized emitter of electromagnetic radiation, the shorter is the wavelength at which the maximum amount of radiation is emitted. The mathematical equation describing this relationship is known as Wien's law: $\lambda_{max} = (3 \times 10^6)/T$. The total power emitted per square meter increases with increasing temperature. The relationship between emitted energy flux and temperature is known as the Stefan-Boltzmann law: $F = \sigma T^4$.

Glossary

blackbody

an idealized object that absorbs all electromagnetic energy that falls onto it

electromagnetic spectrum

the whole array or family of electromagnetic waves, from radio to gamma rays

energy flux

the amount of energy passing through a unit area (for example, 1 square meter) per second; the units of flux are watts per square meter

gamma rays

photons (of electromagnetic radiation) of energy with

wavelengths no longer than 0.01 nanometer; the most energetic form of electromagnetic radiation

infrared

electromagnetic radiation of wavelength 10³–10⁶ nanometers; longer than the longest (red) wavelengths that can be perceived by the eye, but shorter than radio wavelengths

microwave

electromagnetic radiation of wavelengths from 1 millimeter to 1 meter; longer than infrared but shorter than radio waves

radio waves

all electromagnetic waves longer than microwaves, including radar waves and AM radio waves

Stefan-Boltzmann law

a formula from which the rate at which a blackbody radiates energy can be computed; the total rate of energy emission from a unit area of a blackbody is proportional to the fourth power of its absolute temperature: $F = \sigma T^4$

ultraviolet

electromagnetic radiation of wavelengths 10 to 400 nanometers; shorter than the shortest visible wavelengths

visible light

electromagnetic radiation with wavelengths of roughly 400–700 nanometers; visible to the human eye

Wien's law

formula that relates the temperature of a blackbody to the wavelength at which it emits the greatest intensity of radiation

X-rays

electromagnetic radiation with wavelengths between 0.01 nanometer and 20 nanometers; intermediate between those of ultraviolet radiation and gamma rays

Chapter 5 Section 5.3: Spectroscopy in 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 (<u>Figure 1</u>). Action of a Prism.



Figure 1. 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 1 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 (<u>Figure 2</u>). (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 2. 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 3).

Visible Spectrum of the Sun.



Figure 3. 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 signatures* by which their presence could be detected (Figure 4). 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.

Continuous Spectrum and Line Spectra from Different Elements.



Figure 4. 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 3 shows an absorption spectrum, whereas <u>Figure 4</u> 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.

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. 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.

Rainbow Refraction.



Figure 5. (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 <u>Figure 5</u>. 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.

Key Concepts and Summary

A spectrometer is a device that forms a spectrum, often utilizing the phenomenon of dispersion. The light from an astronomical source can consist of a continuous spectrum, an emission (bright line) spectrum, or an absorption (dark line) spectrum. Because each element leaves its spectral signature in the pattern of lines we observe, spectral analyses reveal the composition of the Sun and stars.

Glossary

absorption spectrum

a series or pattern of dark lines superimposed on a continuous spectrum

continuous spectrum

a spectrum of light composed of radiation of a continuous range of wavelengths or colors, rather than only certain discrete wavelengths

dispersion

separation of different wavelengths of white light through refraction of different amounts

emission spectrum

a series or pattern of bright lines superimposed on a continuous spectrum

spectrometer

an instrument for obtaining a spectrum; in astronomy, usually attached to a telescope to record the spectrum of a star, galaxy, or other astronomical object

Chapter 5 Section 5.4: The Structure of the Atom

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
- Explain the behavior of electrons within atoms and how electrons interact with light to move among energy levels

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 1). 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."

Rutherford's Experiment.



Figure 1. (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 2). 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 2. 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 3).

Helium Atom.



Figure 3. 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 (<u>Figure 4</u>), 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.

Isotopes of Hydrogen.



Figure 4. 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.

To explore the structure of atoms, go to the PhET Build and 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 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). 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.

Niels Bohr (1885–1962) and Max Planck (1858–1947).



Figure 5. (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.

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:

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

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

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

Check Your Learning

What is the energy of a yellow photon with a frequency of $5.5\times10^{14}\,\text{Hz}?$

ANSWER:

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

Key Concepts and Summary

Atoms consist of a nucleus containing one or more positively charged protons. All atoms except hydrogen can also contain one or more neutrons in the nucleus. Negatively charged electrons orbit the nucleus. The number of protons defines an element (hydrogen has one proton, helium has two, and so on) of the atom. Nuclei with the same number of protons but different numbers of neutrons are different isotopes of the same element. In the Bohr model of the atom, electromagnetic radiation. But when electrons go from lower levels to higher ones, they must absorb a photon of just the right energy, and when they go from higher levels to lower ones, they give off a photon of just the right energy. The energy of a photon is connected to the frequency of the electromagnetic wave it represents by Planck's formula, E = hf.

Glossary

energy level

a particular level, or amount, of energy possessed by an atom or ion above the energy it possesses in its least energetic state; also used to refer to the states of energy an electron can have in an atom

isotope

any of two or more forms of the same element whose atoms have the same number of protons but different numbers of neutrons

nucleus (of an atom)

the massive part of an atom, composed mostly of protons and neutrons, and about which the electrons revolve

Chapter 5 Section 5.5: The Formation of Spectral Lines

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 <u>Figure 1</u>. Bohr Model for Hydrogen.



Figure 1. 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 <u>Figure 2</u>. 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 2), 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.



Energy-Level Diagrams for Hydrogen.

Figure 2. (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. (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 3 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.

Three Kinds of Spectra.



Figure 3. 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.

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 nowionized 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.

Key Concepts and Summary

When electrons move from a higher energy level to a lower one, photons are emitted, and an emission line can be seen in the spectrum. Absorption lines are seen when electrons absorb photons and move to higher energy levels. Since each atom has its own characteristic set of energy levels, each is associated with a unique pattern of spectral lines. This allows astronomers to determine what elements are present in the stars and in the clouds of gas and dust among the stars. An atom in its lowest energy level is in the ground state. If an electron is in an orbit other than the least energetic one possible, the atom is said to be excited. If an atom has lost one or more electrons, it is called an ion and is said to be ionized. The spectra of different ions look different and can tell astronomers about the temperatures of the sources they are observing.

Glossary

excitation

the process of giving an atom or an ion an amount of energy greater than it has in its lowest energy (ground) state

ground state

the lowest energy state of an atom

ion

an atom that has become electrically charged by the addition or loss of one or more electrons

ionization

the process by which an atom gains or loses electrons

Chapter 5 Section 5.6: The Doppler Effect

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 1.



Doppler Effect.

Figure 1. (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.

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 $v = c \times \Delta\lambda/\lambda$.

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.

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:

 $v=c\times\frac{\Delta\lambda}{\lambda}=(3.0\times10^8m/s)(\frac{0.3nm}{656.3nm})=(3.0\times10^8m/s)(\frac{0.3\times10^9m}{656.3\times10^9m})=140,000m/s=140km/s$

Check Your Learning

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:

Because the light is shifted to a longer wavelength, the star is moving away from us:

 $v=c\times\frac{\Delta\lambda}{\lambda}=(3.0\times10^8m/s)(\frac{0.1nm}{500nm})=(3.0\times10^8m/s)(\frac{0.1\times10^{-9}m}{500\times10^{-9}m})=60,000m/s$

Its speed is 60,000 m/s.

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 lightyears away.

Key Concepts and Summary

If an atom is moving toward us when an electron changes orbits and produces a spectral line, we see that line shifted slightly toward the blue of its normal wavelength in a spectrum. If the atom is moving away, we see the line shifted toward the red. This shift is known as the Doppler effect and can be used to measure the radial velocities of distant objects.

For Further Exploration

Articles

Augensen, H. & Woodbury, J. "The Electromagnetic Spectrum." Astronomy (June 1982): 6.Darling, D. "Spectral Visions: The Long Wavelengths." Astronomy (August 1984): 16; "The Short Wavelengths." Astronomy (September 1984): 14.Gingerich, О. "Unlocking the Chemical Secrets of the Cosmos." Sky 87 Telescope (July 1981): 13.Stencil, R. et al. "Astronomical Spectroscopy." Astronomy (June 1978): 6.

Websites

Doppler Effect: http://www.physicsclassroom.com/class/waves/ Lesson-3/The-Doppler-Effect. A shaking bug and the Doppler Effect explained.Electromagnetic Spectrum: http://imagine.gsfc.nasa.gov/science/toolbox/ emspectrum1.html. An introduction to the electromagnetic spectrum from NASA's Imagine the Universe; note that you can click the "Advanced" button near the top and get a more detailed discussion.Rainbows: How They Form and How to See Them: http://www.livescience.com/30235-rainbows-formationexplainer.html. By meteorologist and amateur astronomer Joe Rao.

Videos

Doppler Effect: http://www.esa.int/spaceinvideos/Videos/2014/ 07/Doppler_effect_-_classroom_demonstration_video_VP05.

ESA video with Doppler ball demonstration and Doppler effect and Make satellites (4:48).How а Prism Works to Rainbow Colors: https://www.youtube.com/watch?v=JGqsi_LDUn0. Short video on how a prism bends light to make a rainbow of colors of the (2:44).Tour Electromagnetic Spectrum: https://www.youtube.com/

watch?v=HPcAWNIVI-8. NASA Mission Science video tour of the bands of the electromagnetic spectrum (eight short videos).

Introductions to Quantum Mechanics

Ford, Kenneth. The Quantum World. 2004. A well-written recent introduction by a physicist/educator.Gribbin, John. In Search of Schroedinger's Cat. 1984. Clear, very basic introduction to the fundamental ideas of quantum mechanics, by a British physicist and science writer.Rae, Alastair. Quantum Physics: A Beginner's Guide. 2005. Widely praised introduction by a British physicist.

Collaborative Group Activities

- A. Have your group make a list of all the electromagnetic wave technology you use during a typical day.
- B. How many applications of the Doppler effect can your group think of in everyday life? For example, why would the highway patrol find it useful?
- C. Have members of your group go home and "read" the face of your radio set and then compare notes. If you do not have a radio, research "broadcast radio frequencies" to find answers to the following questions. What do all the words and symbols mean? What frequencies can your radio tune to? What is the frequency of your favorite radio station? What is its wavelength?
- D. If your instructor were to give you a spectrometer, what kind of spectra does your group think you would see from each of the following: (1) a household lightbulb, (2) the Sun, (3) the "neon lights of Broadway," (4) an ordinary household flashlight, and (5) a streetlight on a busy shopping street?
- E. Suppose astronomers want to send a message to an alien civilization that is living on a planet with an atmosphere very similar to that of Earth's. This message must travel through space, make it through the other planet's atmosphere, and be noticeable to the residents of that planet. Have your group

discuss what band of the electromagnetic spectrum might be best for this message and why. (Some people, including noted physicist Stephen Hawking, have warned scientists not to send such messages and reveal the presence of our civilization to a possible hostile cosmos. Do you agree with this concern?)

Review Questions

What distinguishes one type of electromagnetic radiation from another? What are the main categories (or bands) of the electromagnetic spectrum?

What is a wave? Use the terms wavelength and frequency in your definition.

Is your textbook the kind of idealized object (described in section on radiation laws) that absorbs all the radiation falling on it? Explain. How about the black sweater worn by one of your classmates?

Where in an atom would you expect to find electrons? Protons? Neutrons?

Explain how emission lines and absorption lines are formed. In what sorts of cosmic objects would you expect to see each? Explain how the Doppler effect works for sound waves and give some familiar examples.

What kind of motion for a star does not produce a Doppler effect? Explain.

Describe how Bohr's model used the work of Maxwell.

Explain why light is referred to as electromagnetic radiation. Explain the difference between radiation as it is used in most everyday language and radiation as it is used in an astronomical context.

What are the differences between light waves and sound waves? Which type of wave has a longer wavelength: AM radio waves (with frequencies in the kilohertz range) or FM radio waves (with frequencies in the megahertz range)? Explain.

Explain why astronomers long ago believed that space must be filled with some kind of substance (the "aether") instead of the vacuum we know it is today.

Explain what the ionosphere is and how it interacts with some radio waves.

Which is more dangerous to living things, gamma rays or X-rays? Explain.

Explain why we have to observe stars and other astronomical objects from above Earth's atmosphere in order to fully learn about their properties.

Explain why hotter objects tend to radiate more energetic photons compared to cooler objects.

Explain how we can deduce the temperature of a star by determining its color.

Explain what dispersion is and how astronomers use this phenomenon to study a star's light.

Explain why glass prisms disperse light.

Explain what Joseph Fraunhofer discovered about stellar spectra. Explain how we use spectral absorption and emission lines to determine the composition of a gas.

Explain the results of Rutherford's gold foil experiment and how they changed our model of the atom.

Is it possible for two different atoms of carbon to have different numbers of neutrons in their nuclei? Explain.

What are the three isotopes of hydrogen, and how do they differ? Explain how electrons use light energy to move among energy levels within an atom.

Explain why astronomers use the term "blueshifted" for objects moving toward us and "redshifted" for objects moving away from us.

If spectral line wavelengths are changing for objects based on the radial velocities of those objects, how can we deduce which type of atom is responsible for a particular absorption or emission line?

Thought Questions

Make a list of some of the many practical consequences of Maxwell's theory of electromagnetic waves (television is one example).

With what type of electromagnetic radiation would you observe:

- A. A star with a temperature of 5800 K?
- B. A gas heated to a temperature of one million K?
- C. A person on a dark night?

Why is it dangerous to be exposed to X-rays but not (or at least much less) dangerous to be exposed to radio waves? Go outside on a clear night, wait 15 minutes for your eyes to adjust to the dark, and look carefully at the brightest stars. Some should look slightly red and others slightly blue. The primary factor that determines the color of a star is its temperature. Which is hotter: a blue star or a red one? Explain

Water faucets are often labeled with a red dot for hot water and a blue dot for cold. Given Wien's law, does this labeling make sense? Suppose you are standing at the exact center of a park surrounded by a circular road. An ambulance drives completely around this road, with siren blaring. How does the pitch of the siren change as it circles around you?

How could you measure Earth's orbital speed by photographing the spectrum of a star at various times throughout the year? (Hint: Suppose the star lies in the plane of Earth's orbit.)

Astronomers want to make maps of the sky showing sources of Xrays or gamma rays. Explain why those X-rays and gamma rays must be observed from above Earth's atmosphere.

The greenhouse effect can be explained easily if you understand the laws of blackbody radiation. A greenhouse gas blocks the transmission of infrared light. Given that the incoming light to Earth is sunlight with a characteristic temperature of 5800 K (which peaks in the visible part of the spectrum) and the outgoing light from Earth has a characteristic temperature of about 300 K (which peaks in the infrared part of the spectrum), explain how greenhouse gases cause Earth to warm up. As part of your answer, discuss that greenhouse gases block both incoming and outgoing infrared light. Explain why these two effects don't simply cancel each other, leading to no net temperature change. An idealized radiating object does not reflect or scatter any radiation but instead absorbs all of the electromagnetic energy that falls on it. Can you explain why astronomers call such an object a blackbody? Keep in mind that even stars, which shine brightly in a variety of colors, are considered blackbodies. Explain why. Why are ionized gases typically only found in very hightemperature environments?

Explain why each element has a unique spectrum of absorption or emission lines.

Figuring for Yourself

What is the wavelength of the carrier wave of a campus radio station, broadcasting at a frequency of 97.2 MHz (million cycles per second or million hertz)?

What is the frequency of a red laser beam, with a wavelength of 670 nm, which your astronomy instructor might use to point to slides during a lecture on galaxies?

You go to a dance club to forget how hard your astronomy midterm was. What is the frequency of a wave of ultraviolet light coming from a blacklight in the club, if its wavelength is 150 nm? What is the energy of the photon with the frequency you calculated in Exercise?

If the emitted infrared radiation from Pluto, has a wavelength of maximum intensity at 75,000 nm, what is the temperature of Pluto assuming it follows Wien's law?

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

Glossary

Doppler effect

the apparent change in wavelength or frequency of the radiation from a source due to its relative motion away from or toward the observer

radial velocity

motion toward or away from the observer; the component of relative velocity that lies in the line of sight

Chapter 6 Astronomical Instruments Section 6.1: Telescopes

Thinking Ahead

Hubble Space Telescope (HST).



Figure 1. This artist's impression shows the Hubble above Earth, with the rectangular solar panels that provide it with power seen to the left and right.

If you look at the sky when you are far away from city lights, there seem to be an overwhelming number of stars up there. In reality, only about 9000 stars are visible to the unaided eye (from both hemispheres of our planet). The light from most stars is so weak that by the time it reaches Earth, it cannot be detected by the human eye. How can we learn about the vast majority of objects in the universe that our unaided eyes simply cannot see?

In this chapter, we describe the tools astronomers use to extend

274 | Chapter 6 Astronomical Instruments Section 6.1: Telescopes
their vision into space. We have learned almost everything we know about the universe from studying electromagnetic radiation, as discussed in the chapter on Radiation and Spectra. In the twentieth century, our exploration of space made it possible to detect electromagnetic radiation at all wavelengths, from gamma rays to radio waves. The different wavelengths carry different kinds of information, and the appearance of any given object often depends on the wavelength at which the observations are made.

6.1 Telescopes



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 1). 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.

Orion Region at Different Wavelengths.



Figure 1. 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 2). 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.

Two Pre-Telescopic Observatories.



Figure 2. (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 lightgathering 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 $\pi d^2/4$, where *d* is the diameter of the circle.)

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:

$$A = \frac{\pi d^2}{4}$$

the area of a 1-m telescope is:

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

and the area of a 4-m telescope is

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

Check Your Learning

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

ANSWER:

 $\frac{12.6m^2}{0.79m^2} = 16$

Therefore, with 16 times the area, a 4-m telescope collects 16 times the light of a 1-m telescope.

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 3). 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.

Formation of an Image by a Simple Lens.



Figure 3. 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 3, 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 4), 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.



Refracting and Reflecting Telescopes.

Figure 4. Light enters a refracting telescope through a lens at the upper end, which focuses the light near the bottom of the

Chapter 6 Astronomical Instruments Section 6.1: Telescopes | 283

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 4). 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 5). 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.

Focus Arrangements for Reflecting Telescopes.



Prime focus

Newtonian focus

Cassegrain focus

Figure 5. 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 4). Most large professional telescopes have a Casse grain 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.

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.

286 | Chapter 6 Astronomical Instruments Section 6.1: Telescopes

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.

A telescope collects the faint light from astronomical sources and brings it to a focus, where an instrument can sort the light according to wavelength. Light is then directed to a detector, where a permanent record is made. The light-gathering power of a telescope is determined by the diameter of its aperture, or opening-that is, by the area of its largest or primary lens or mirror. The primary optical element in a telescope is either a convex lens (in a refracting telescope) or a concave mirror (in a reflector) that brings the light to a focus. Most large telescopes are reflectors; it is easier to manufacture and support large mirrors because the light does not have to pass through glass.

Glossary

aperture

diameter of the primary lens or mirror of a telescope

chromatic aberration

distortion that causes an image to appear fuzzy when each wavelength coming into a transparent material focuses at a different spot

detector

device sensitive to electromagnetic radiation that makes a record of astronomical observations

eyepiece

magnifying lens used to view the image produced by the objective lens or primary mirror of a telescope

focus

(of telescope) point where the rays of light converged by a mirror or lens meet

prime focus

point in a telescope where the objective lens or primary mirror focuses the light

reflecting telescope

telescope in which the principal light collector is a concave mirror

refracting telescope

telescope in which the principal light collector is a lens or system of lenses

telescope

instrument for collecting visible-light or other electromagnetic radiation

Chapter 6 Section 6.2: Telescopes Today

6.2 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 (<u>Figure 1</u>).

Large Telescope Mirror.



Figure 1. 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, 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.

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)	Mauna Kea, HI	First light 2025 (estimated)	www.tmt.org
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)	Mauna Kea, 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(LSST)	The Cerro Pachón, Chile	First light 2021	www.lsst.org
8.3	Subaru Telescope	Mauna Kea, 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	Mauna Kea, HI (North) and Cerro Pachón, Chile (South)	First light 1999 (North), First light 2000 (South)	www.gemini.edu

Large Single-Dish Visible-Light and Infrared Telescopes

Aperture (m)	Telescope Name	Location	Status	Website
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

Large Single-Dish Visible-Light and Infrared Telescopes

The differences between the Palomar telescope and the modern Gemini North telescope (to take an example) are easily seen in <u>Figure 2</u>. 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.

Modern Reflecting Telescopes.



Figure 2. (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 Mauna Kea, 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 3). 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.

Thirty-Six Eyes Are Better Than One.



Figure 3. The mirror of the 10-meter Keck telescope is composed of 36 hexagonal sections. (credit: NASA)

Learn more about the Keck Observatory on Mauna Kea 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.

GEORGE ELLERY HALE: MASTER TELESCOPE BUILDER

George Ellery Hale (Figure 4) 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.

George Ellery Hale (1868–1938).



Figure 4. 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 5).

World's Largest Refractor.



Figure 5. 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), the desert peaks of Arizona, the Canary Islands in the Atlantic Ocean, and Mauna Kea in Hawaii, a dormant volcano with an altitude of 13,700 feet (4200 meters).

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. And, 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.

High and Dry Site.



Figure 6. 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 7 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.

Power of Adaptive Optics.



Figure 7. 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))

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.

New technologies for creating and supporting lightweight mirrors
have led to the construction of a number of large telescopes since 1990. The site for an astronomical observatory must be carefully chosen for clear weather, dark skies, low water vapor, and excellent atmospheric seeing (low atmospheric turbulence). The resolution of a visible-light or infrared telescope is degraded by turbulence in Earth's atmosphere. The technique of adaptive optics, however, can make corrections for this turbulence in real time and produce exquisitely detailed images.

Glossary

adaptive optics

systems used with telescopes that can compensate for distortions in an image introduced by the atmosphere, thus resulting in sharper images

resolution

detail in an image; specifically, the smallest angular (or linear) features that can be distinguished

seeing

unsteadiness of Earth's atmosphere, which blurs telescopic images; good seeing means the atmosphere is steady

Chapter 6 Section 6.3: Visible-Light Detectors and Instruments

6.3 Visible-Light Detectors and Instruments



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.

312 | Chapter 6 Section 6.3: Visible-Light Detectors and

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 on Radiation and in the chapter 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 **chargecoupled 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 <u>Figure 1</u>). 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).

Charge-Coupled Devices (CCDs).



Figure 1. (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 serves 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 (μ m), 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 2). 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.

Infrared Eyes.



Figure 2. 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.

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 3. 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.

Prism Spectrometer.



Figure 3. 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.

Visible-light detectors include the human eye, photographic film, and charge-coupled devices (CCDs). Detectors that are sensitive to infrared radiation must be cooled to very low temperatures since everything in and near the telescope gives off infrared waves. A spectrometer disperses the light into a spectrum to be recorded for detailed analysis.

Glossary

charge-coupled device (CCD)

array of high-sensitivity electronic detectors of electromagnetic radiation, used at the focus of a telescope (or camera lens) to record an image or spectrum

Chapter 6 Section 6.4: Radio Telescopes

6.4 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 1). 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.

First Radio Telescope.



Figure 1. 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 2 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.

Radio Image.



Figure 2. 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 3). Table lists some of the major radio telescopes of the world.

Robert C. Byrd Green Bank Telescope.



Figure 3. 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)

Major Radio Observatories of the World

Observatory	Location	Description	Website		
Individual Radio Dishes					
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	5				
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		
Very Large Array (VLA)	Socorro, New Mexico	27-element array of 25-m dishes (36-km baseline)	www.science.nrao.edu/ facilities/vla		

Major Radio Observatories of the World

Observatory	Location	Description	Website	
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	
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)	Mauna Kea, 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 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 4), 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.

Atacama Large Millimeter/Submillimeter Array (ALMA).



Figure 4. Located in the Atacama Desert of Northern Chile, ALMA currently provides the highest resolution for radio observations. (credit: ESO/S. Guisard)

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 5). 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.

Very Long Baseline Array.



Figure 5. 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.

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	Mauna Kea, 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

Visible-Light Interferometers

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. The most spectacular facility in the world for radar astronomy is the 1000-foot (305-meter) telescope at Arecibo in Puerto Rico (Figure <u>6</u>). The Arecibo telescope is too large to be pointed directly at different parts of the sky. Instead, it is constructed in a huge natural "bowl" (more than a mere dish) formed by several hills, and it is lined with reflecting metal panels. A limited ability to track astronomical sources is achieved by moving the receiver system, which is suspended on cables 100 meters above the surface of the bowl. An even larger (500-meter) radar telescope is currently under construction. It is the **Five-hundred-meter Aperture Spherical Telescope** (FAST) in China and is expected to be completed in 2016.

Largest Radio and Radar Dish.



Figure 6. The Arecibo Observatory, with its 1000-foot radio dishfilling valley in Puerto Rico, is part of the National Astronomy and Ionosphere Center, operated by SRI International, USRA, and UMET under a cooperative agreement with the National Science Foundation. (credit: National Astronomy and Ionosphere Center, Cornell U., NSF)

In the 1930s, radio astronomy was pioneered by Karl G. Jansky and Grote Reber. A radio telescope is basically a radio antenna (often a large, curved dish) connected to a receiver. Significantly enhanced resolution can be obtained with interferometers, including interferometer arrays like the 27-element VLA and the 66-element ALMA. Expanding to very long baseline interferometers, radio astronomers can achieve resolutions as precise as 0.0001 arcsecond. Radar astronomy involves transmitting as well as receiving. The largest radar telescope currently in operation is a 305-meter bowl at Arecibo.

Glossary

interference

process in which waves mix together such that their crests and troughs can alternately reinforce and cancel one another

interferometer

instrument that combines electromagnetic radiation from one or more telescopes to obtain a resolution equivalent to what would be obtained with a single telescope with a diameter equal to the baseline separating the individual separate telescopes

interferometer array

combination of multiple radio dishes to, in effect, work like a large number of two-dish interferometers

radar

technique of transmitting radio waves to an object and then detecting the radiation that the object reflects back to the transmitter; used to measure the distance to, and motion of, a target object or to form images of it

Chapter 6 Section 6.5: Observations outside Earth's Atmosphere

6.5 Observations outside Earth's Atmosphere



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 infared 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 1).

Stratospheric Observatory for Infrared

334 | Chapter 6 Section 6.5: Observations outside Earth's Atmosphere

Astronomy (SOFIA).



Figure 1. SOFIA allows observations to be made above most of Earth's atmospheric water vapor. (credit: NASA)

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 2. 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.

Observations from the Spitzer Space Telescope (SST).



Figure 2. 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 (Xray: 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 ([link])-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 3).

Hubble Ultra-Deep Field (HUDF).



Figure 3. 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 <u>4</u>). 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.

Chandra X-Ray Satellite.



Figure 4. 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. <u>Table</u> 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.

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. web/xmm-newton
International Gamma-Ray Astrophysics Laboratory (INTEGRAL)	2002	X- and gamma-rays	higher resolution gamma-ray images	http://sci.esa.int/int
Spitzer Space Telescope	2003	IR	0.85-m telescope	www.spitzer.caltech.
Fermi Gamma-ray Space Telescope	2008	gamma-rays	first high-energy gamma-ray observations	fermi.gsfc.nasa.gov
Kepler	2009	visible-light	planet finder	http://kepler.nasa.go
Wide-field Infrared Survey Explorer (WISE)	2009	IR	whole-sky map, asteroid searches	www.nasa.gov/ mission_pages/WISI
Gaia	2013	visible-light	Precise map of the Milky Way	http://sci.esa.int/gai

Infrared observations are made with telescopes aboard aircraft and in space, as well as from ground-based facilities on dry mountain peaks. Ultraviolet, X-ray, and gamma-ray observations must be made from above the atmosphere. Many orbiting observatories have been flown to observe in these bands of the spectrum in the last few decades. The largest-aperture telescope in space is the Hubble Space telescope (HST), the most significant infrared telescope is Spitzer, and Chandra and Fermi are the premier X-ray and gammaray observatories, respectively.

Chapter 6 Section 6.6: The Future of Large Telescopes

6.6 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 1), which (in a departure from tradition) is named after one of the early administrators of NASA instead of a scientist. This telescope will have a mirror 6 meters in diameter, made up, like the Keck telescopes, of 36 small hexagons. These will have to unfold into place once the telescope reaches its stable orbit point, some 1.5 million kilometers from Earth (where no astronauts can currently travel if it needs repair.) The telescope is scheduled for launch in 2018 and should have the sensitivity needed to detect the very first generation of stars, formed when the universe was only a few hundred million years old. With the ability to measure both visible and infrared wavelengths, it will serve as the successor to both HST and the Spitzer Space Telescope.

James Webb Space Telescope (JWST).



Figure 1. 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) 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.

On the ground, astronomers have started building the **Large Synoptic Survey Telescope** (LSST), 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. The LSST is expected to see first light in 2021.

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 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 onethird 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. These include the Thirty-Meter Telescope for which construction has begun at the top of Mauna Kea in Hawaii.

The most ambitious of these projects is the **European Extremely Large Telescope** (E-ELT) (<u>Figure 2</u>). (Astronomers try to outdo each other not only with the size of these telescopes, but also their names!) The design of the E-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. The E-ELT, along with the Thirty Meter Telescope and the Giant Magellan Telescope, which are being built by international consortia led by US astronomers, will combine lightgathering power with high-resolution imaging. These powerful new instruments will enable astronomers to tackle many important astronomical problems. For example, they should be able to tell us when, where, and how often planets form around other stars. They should even be able to provide us images and spectra of such planets and thus, perhaps, give us the first real evidence (from the chemistry of these planets' atmospheres) that life exists elsewhere.

Artist's Conception of the European Extremely Large Telescope.



Figure 2. 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
Check out this fun diagram comparing the sizes of the largest planned and existing telescopes to a regulation basketball and tennis court.

New and even larger telescopes are on the drawing boards. The James Webb Space Telescope, a 6-meter successor to Hubble, is currently scheduled for launch in 2018. Gamma-ray astronomers are planning to build the CTA to measure very energetic gamma rays. Astronomers are building the LSST to observe with an unprecedented field of view and a new generation of visible-light/ infrared telescopes with apertures of 24.5 to 39 meters in diameter.

For Further Exploration

Articles

Blades, J. C. "Fixing the Hubble One Last Time." Sky & *Telescope* (October 2008): 26. On the last Shuttle service mission and what the Hubble was then capable of doing.

Brown, A. "How Gaia will Map a Billion Stars." Astronomy (December 2014): 32. Nice review of the mission to do photometry and spectroscopy of all stars above a certain brightness.

Irion, R. "Prime Time." Astronomy (February 2001): 46. On how time is allotted on the major research telescopes.

Jedicke, Peter & Robert. "The Coming Giant Sky Patrols." Sky &

Telescope (September 2008): 30. About giant telescopes to survey the sky continuously.

Lazio, Joseph, et al. "Tuning in to the Universe: 21st Century Radio Astronomy." Sky & Telescope (July 2008): 21. About ALMA and the Square Kilometer Array.

Lowe, Jonathan. "Mirror, Mirror." Sky & Telescope (December 2007): 22. On the Large Binocular Telescope in Arizona.

Lowe, Jonathan. "Next Light: Tomorrow's Monster Telescopes." *Sky & Telescope* (April 2008): 20. About plans for extremely large telescopes on the ground.

Mason, Todd & Robin. "Palomar's Big Eye." Sky & Telescope (December 2008): 36. On the Hale 200-inch telescope.

Subinsky, Raymond. "Who Really Invented the Telescope." Astronomy (August 2008): 84. Brief historical introduction, focusing on Hans Lippershey.

Websites

Websites for major telescopes are given in [link], [link], [link], and [link].

Videos

Astronomy from the Stratosphere: SOFIA: https://www.youtube.com/watch?v=NV98BcBBA9c. A talk by Dr. Dana Backman (1:15:32)

Galaxies Viewed in Full Spectrum of Light: https://www.youtube.com/watch?v=368K0iQv8nE.

Scientists with the Spitzer Observatory show how a galaxy looks different at different wavelengths (6:22)

Lifting the Cosmic Veil: Highlights from a Decade of the Spitzer

Space Telescope: https://www.youtube.com/ watch?v=nkrNQcwkY78. A talk by Dr. Michael Bicay (1:42:44)

Collaborative Group Activities

- A. Most large telescopes get many more proposals for observing projects than there is night observing time available in a year. Suppose your group is the telescope time allocation committee reporting to an observatory director. What criteria would you use in deciding how to give out time on the telescope? What steps could you take to make sure all your colleagues thought the process was fair and people would still talk to you at future astronomy meetings?
- B. Your group is a committee of nervous astronomers about to make a proposal to the government ministers of your small European country to chip in with other countries to build the world's largest telescope in the high, dry desert of the Chilean Andes Mountains. You expect the government ministers to be very skeptical about supporting this project. What arguments would you make to convince them to participate?
- C. The same government ministers we met in the previous activity ask you to draw up a list of the pros and cons of having the world's largest telescope in the mountains of Chile (instead of a mountain in Europe). What would your group list in each column?
- D. Your group should discuss and make a list of all the ways in which an observing session at a large visible-light telescope and a large radio telescope might differ. (Hint: Bear in mind that because the Sun is not especially bright at many radio wavelengths, observations with radio telescopes can often be done during the day.)
- E. Another "environmental threat" to astronomy (besides light pollution) comes from the spilling of terrestrial

communications into the "channels"–wavelengths and frequencies–previously reserved for radio astronomy. For example, the demand for cellular phones means that more and more radio channels will be used for this purpose. The faint signals from cosmic radio sources could be drowned in a sea of earthly conversation (translated and sent as radio waves). Assume your group is a congressional committee being lobbied by both radio astronomers, who want to save some clear channels for doing astronomy, and the companies that stand to make a lot of money from expanding cellular phone use. What arguments would sway you to each side?

- F. When the site for the new Thirty-Meter Telescope on Hawaii's Mauna Kea was dedicated, a group of native Hawaiians announced opposition to the project because astronomers were building too many telescopes on a mountain that native Hawaiians consider a sacred site. You can read more about this controversy at http://www.nytimes.com/2015/12/04/ science/space/hawaii-court-rescinds-permit-to-build-thirty-meter-telescope.html?_r=0 and at http://www.nature.com/ news/the-mountain-top-battle-over-the-thirty-meter-telescope-1.18446. Once your group has the facts, discuss the claims of each side in the controversy. How do you think it should be resolved?
- G. If you could propose to use a large modern telescope, what would you want to find out? What telescope would you use and why?
- H. Light pollution (spilled light in the night sky making it difficult to see the planets and stars) used to be an issue that concerned mostly astronomers. Now spilled light at night is also of concern to environmentalists and those worrying about global warming. Can your group come up with some nonastronomical reasons to be opposed to light pollution?

Review Questions

What are the three basic components of a modern astronomical instrument? Describe each in one to two sentences.

Name the two spectral windows through which electromagnetic radiation easily reaches the surface of Earth and describe the largest-aperture telescope currently in use for each window.

List the largest-aperture single telescope currently in use in each of the following bands of the electromagnetic spectrum: radio, X-ray, gamma ray.

When astronomers discuss the apertures of their telescopes, they say bigger is better. Explain why.

The Hooker telescope at Palomar Observatory has a diameter of 5 m, and the Keck I telescope has a diameter of 10 m. How much more light can the Keck telescope collect than the Hooker telescope in the same amount of time?

What is meant by "reflecting" and "refracting" telescopes?

Why are the largest visible-light telescopes in the world made with mirrors rather than lenses?

Compare the eye, photographic film, and CCDs as detectors for light. What are the advantages and disadvantages of each?

What is a charge-coupled device (CCD), and how is it used in astronomy?

Why is it difficult to observe at infrared wavelengths? What do astronomers do to address this difficulty?

Radio and radar observations are often made with the same antenna, but otherwise they are very different techniques. Compare

and contrast radio and radar astronomy in terms of the equipment needed, the methods used, and the kind of results obtained.

Look back at [link] of Cygnus A and read its caption again. The material in the giant lobes at the edges of the image had to have been ejected from the center *at least* how many years ago?

Why do astronomers place telescopes in Earth's orbit? What are the advantages for the different regions of the spectrum?

What was the problem with the Hubble Space Telescope and how was it solved?

Describe the techniques radio astronomers use to obtain a resolution comparable to what astronomers working with visible light can achieve.

What kind of visible-light and infrared telescopes on the ground are astronomers planning for the future? Why are they building them on the ground and not in space?

Describe one visible-light or infrared telescope that astronomers are planning to launch into space in the future.

Thought Questions

What happens to the image produced by a lens if the lens is "stopped down" (the aperture reduced, thereby reducing the amount of light passing through the lens) with an iris diaphragm–a device that covers its periphery?

What would be the properties of an ideal astronomical detector? How closely do the actual properties of a CCD approach this ideal?

Many decades ago, the astronomers on the staff of Mount Wilson

and Palomar Observatories each received about 60 nights per year for their observing programs. Today, an astronomer feels fortunate to get 10 nights per year on a large telescope. Can you suggest some reasons for this change?

The largest observatory complex in the world is on Mauna Kea, the tallest mountain on Earth. What are some factors astronomers consider when selecting an observatory site? Don't forget practical ones. Should astronomers, for example, consider building an observatory on Denali (Mount McKinley) or Mount Everest?

Suppose you are looking for sites for a visible-light observatory, an infrared observatory, and a radio observatory. What are the main criteria of excellence for each? What sites are actually considered the best today?

Radio astronomy involves wavelengths much longer than those of visible light, and many orbiting observatories have probed the universe for radiation of very short wavelengths. What sorts of objects and physical conditions would you expect to be associated with emission of radiation at very long and very short wavelengths?

The dean of a university located near the ocean (who was not a science major in college) proposes building an infrared telescope right on campus and operating it in a nice heated dome so that astronomers will be comfortable on cold winter nights. Criticize this proposal, giving your reasoning.

Figuring for Yourself

What is the area, in square meters, of a 10-m telescope?

Approximately 9000 stars are visible to the naked eye in the whole sky (imagine that you could see around the entire globe and both the northern and southern hemispheres), and there are about 41,200 square degrees on the sky. How many stars are visible per square degree? Per square arcsecond?

Theoretically (that is, if seeing were not an issue), the resolution of a telescope is inversely proportional to its diameter. How much better is the resolution of the ALMA when operating at its longest baseline than the resolution of the Arecibo telescope?

In broad daylight, the size of your pupil is typically 3 mm. In dark situations, it expands to about 7 mm. How much more light can it gather?

How much more light can be gathered by a telescope that is 8 m in diameter than by your fully dark-adapted eye at 7 mm?

How much more light can the Keck telescope (with its 10-m diameter mirror) gather than an amateur telescope whose mirror is 25 cm (0.25 m) across?

People are often bothered when they discover that reflecting telescopes have a second mirror in the middle to bring the light out to an accessible focus where big instruments can be mounted. "Don't you lose light?" people ask. Well, yes, you do, but there is no better alternative. You can estimate how much light is lost by such an arrangement. The primary mirror (the one at the bottom in [link]) of the Gemini North telescope is 8 m in diameter. The secondary mirror at the top is about 1 m in diameter. Use the formula for the area of a circle to estimate what fraction of the light is blocked by the secondary mirror.

Telescopes can now be operated remotely from a warm room, but until about 25 years ago, astronomers worked at the telescope to guide it so that it remained pointed in exactly the right place. In a large telescope, like the Palomar 200-inch telescope, astronomers sat in a cage at the top of the telescope, where the secondary mirror is located, as shown in [link]. Assume for the purpose of your calculation that the diameter of this cage was 40 inches. What fraction of the light is blocked?

The HST cost about \$1.7 billion for construction and \$300 million for its shuttle launch, and it costs \$250 million per year to operate. If the telescope lasts for 20 years, what is the total cost per year? Per day? If the telescope can be used just 30% of the time for actual observations, what is the cost per hour and per minute for the astronomer's observing time on this instrument? What is the cost per person in the United States? Was your investment in the Hubble Space telescope worth it?

How much more light can the James Webb Space Telescope (with its 6-m diameter mirror) gather than the Hubble Space Telescope (with a diameter of 2.4 m)?

The Palomar telescope's 5-m mirror weighs 14.5 tons. If a 10-m mirror were constructed of the same thickness as Palomar's (only bigger), how much would it weigh?

Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System

Thinking Ahead

"Self-Portrait" of Mars.



Figure 1. This picture was taken by the *Curiosity* Rover on **Mars** in 2012. The image is reconstructed digitally from 55 different images taken by a camera on the rover's extended mast, so that the many positions of the mast (which acted like a selfie stick) are edited out. (credit: modification of work by NASA/JPL-Caltech/MSSS)

Surrounding the Sun is a complex system of worlds with a wide range of conditions: eight major planets, many dwarf planets, hundreds of moons, and countless smaller objects. Thanks largely to

356 | Chapter 7 The Other Worlds: An Introduction to the Solar System visits by spacecraft, we can now envision the members of the solar system as other worlds like our own, each with its own chemical and geological history, and unique sights that interplanetary tourists may someday visit. Some have called these past few decades the "golden age of planetary exploration," comparable to the golden age of exploration in the fifteenth century, when great sailing ships plied Earth's oceans and humanity became familiar with our own planet's surface.

In this chapter, we discuss our planetary system and introduce the idea of comparative planetology-studying how the planets work by comparing them with one another. We want to get to know the planets not only for what we can learn about them, but also to see what they can tell us about the origin and evolution of the entire solar system. In the upcoming chapters, we describe the betterknown members of the solar system and begin to compare them to the thousands of planets that have been discovered recently, orbiting other stars.

7.1 Overview of Our Planetary System



Model the solar system with distances from everyday life to better comprehend distances in space

The solar system¹ 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 and Itokawa, 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 (<u>Figure 1</u>). We have even discovered other places in our solar system that might be able to support some kind of life.

Astronauts on the Moon.



Figure 1. 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)

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 <u>Table</u>. It is an enormous ball about 1.4 million kilometers

Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System | 359 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.

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	0.000002 (estimate)
Cosmic dust	0.0000001 (estimate)

Mass of Members of the Solar System

Table 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 2). 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 1750 of these TNOs have been discovered.



Orbits of the Planets.

Figure 2. 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

Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System | 361 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 3).

Surface of Mercury.



Figure 3. 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

362 | Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System

Jupiter in mythology) or giant planets—a name they richly deserve (Figure 4). More than 1400 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.

The Four Giant Planets.



Figure 4. 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 5]). <u>Table</u> summarizes some of the main facts about the planets.

Pluto Close-up.



Figure 5. 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 temporarily being called the Sputnik Plain, after humanity's first spacecraft. (credit: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

Name	Distance from	Revolution Period	Diameter	Mass	Density
	(AU)2	(y)	(km)	(10 ²³ kg)	$(g/cm^3)^3$
Mercury	0.39	0.24	4,878	3.3	5.4
Venus	0.72	0.62	12,120	48.7	5.2
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

The Planets

364 | Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System

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. 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 km = 2×10^5 m).

Solution

The volume of Mimas is

 $\frac{4}{3} \times 3.14 \times (2 \times 10^5 m)^3 = 3.3 \times 10^{16} m^3$

Density is mass divided by volume:

$$\frac{4 \times 10^{19} kg}{3.3 \times 10^{16} m^3} = 1.2 \times 10^3 kg/m^3$$

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/m³?)

Check Your Learning

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 for data.

ANSWER:

For a sphere,

density=
$$\frac{mass}{(\frac{4}{3}\pi R^3)}kg/m^3$$

For Earth, then,

(

density
=
$$\frac{6 \times 10^{24} kg}{4.2 \times 2.6 \times 10^{20} m^3} = 5.5 \times 10^3 kg/m^3$$

This density is four to five times greater than Mimas'. In fact, Earth is the densest of the planets. Learn More

Learn more about NASA's mission to Pluto and see high-resolution images of Pluto's 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 180 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 6). 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.

Saturn and Its Rings.



Figure 6. 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 Z). 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.

Asteroid Eros.



Figure 7. This small Earth-crossing asteroid image was taken by

368 | Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System

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 <u>Figure 8</u>). 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.

Comet Churyumov-Gerasimenko (67P).



Figure 8. 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.)

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 9). 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.

Carl Sagan (1934–1996) and Neil deGrasse Tyson.



Figure 9. 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 closeup 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 mostwatched 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.

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 fictionoriented 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?).

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).

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 original 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

376 | Chapter 7 The Other Worlds: An Introduction to the Solar System Section 7.1: Overview of Our Planetary System

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.

Key Concepts and Summary

Our solar system currently consists of the Sun, eight planets, five dwarf planets, nearly 200 known moons, and a host of smaller objects. The planets can be divided into two groups: the inner terrestrial planets and the outer giant planets. Pluto, Eris, Haumea, and Makemake do not fit into either category; as icy dwarf planets, they exist in an ice realm on the fringes of the main planetary system. The giant planets are composed mostly of liquids and gases. Smaller members of the solar system include asteroids (including the dwarf planet Ceres), which are rocky and metallic objects found mostly between Mars and Jupiter; comets, which are made mostly of frozen gases and generally orbit far from the Sun; and countless smaller grains of cosmic dust. When a meteor survives its passage through our atmosphere and falls to Earth, we call it a meteorite.

Footnotes

• 1 The 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.

- 2 An AU (or astronomical unit) is the distance from Earth to the Sun.
- 3 We give densities in units where the density of water is 1 g/cm³. To get densities in units of kg/m³, multiply the given value by 1000.

Glossary

asteroid

a stony or metallic object orbiting the Sun that is smaller than a major planet but that shows no evidence of an atmosphere or of other types of activity associated with comets

comet

a small body of icy and dusty matter that revolves about the Sun; when a comet comes near the Sun, some of its material vaporizes, forming a large head of tenuous gas and often a tail

giant planet

any of the planets Jupiter, Saturn, Uranus, and Neptune in our solar system, or planets of roughly that mass and composition in other planetary systems

meteor

a small piece of solid matter that enters Earth's atmosphere and burns up, popularly called a *shooting star* because it is seen as a small flash of light

meteorite

a portion of a meteor that survives passage through an atmosphere and strikes the ground

terrestrial planet

any of the planets Mercury, Venus, Earth, or Mars; sometimes the Moon is included in the list

Chapter 7 Section 7.2: Composition and Structure of Planets

7.2 Composition and Structure of Planets



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 1). We must infer the existence of the denser core inside these planets from studies of each planet's gravity.

Jupiter.



Figure 1. This true-color image of Jupiter was taken from the Cassini spacecraft in 2000. (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 (H₂O). 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 [link]) 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 2).

Ganymede.



Figure 2. 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. 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.

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, 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 [link]). 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 <u>Figure</u> <u>3</u>) is one dramatic example of this process.

Comet Shoemaker–Levy 9.



Figure 3. 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 (STScl), E. Smith (STScl))

<u>Figure 4</u> shows the aftermath of these collisions, when debris clouds larger than Earth could be seen in **Jupiter**'s atmosphere.

Jupiter with Huge Dust Clouds.



Figure 4. 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 "Earthchauvinist" 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.

Key Concepts and Summary

The giant planets have dense cores roughly 10 times the mass of Earth, surrounded by layers of hydrogen and helium. The terrestrial planets consist mostly of rocks and metals. They were once molten, which allowed their structures to differentiate (that is, their denser materials sank to the center). The Moon resembles the terrestrial planets in composition, but most of the other moons–which orbit the giant planets–have larger quantities of frozen ice within them. In general, worlds closer to the Sun have higher surface temperatures. The surfaces of terrestrial planets have been modified by impacts from space and by varying degrees of geological activity.

Glossary

differentiation

gravitational separation of materials of different density into layers in the interior of a planet or moon

Chapter 7 Section 7.3: Dating Planetary Surfaces

7.3 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 1).



Figure 1. 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 2).

Radioactive Decay.



Figure 2. 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. <u>Table</u> summarizes the decay reactions used most often to date lunar and terrestrial rocks.

Radioactive Decay Reaction Used to Date Rocks¹

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

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.

When astronauts first flew to the Moon, one of their most important tasks was to bring back lunar rocks for radioactive agedating. 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).

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.

Key Concepts and Summary

The ages of the surfaces of objects in the solar system can be estimated by counting craters: on a given world, a more heavily cratered region will generally be older than one that is less cratered. We can also use samples of rocks with radioactive elements in them to obtain the time since the layer in which the rock formed last solidified. The half-life of a radioactive element is the time it takes for half the sample to decay; we determine how many half-lives have passed by how much of a sample remains the radioactive element and how much has become the decay product. In this way, we have estimated the age of the Moon and Earth to be roughly 4.5 billion years.

Footnotes

• 1 The 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.

Glossary

half-life

time required for half of the radioactive atoms in a sample to disintegrate

radioactivity

process by which certain kinds of atomic nuclei decay naturally, with the spontaneous emission of subatomic particles and gamma rays

Chapter 7 Section 7.4: Origin of the Solar System

7.4 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 hundreds 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 1).

Solar Nebula.



Figure 1.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 hydrogendominated 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 2).

Atlas of Planetary Nurseries.



Figure 2.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.)

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.

Key Concepts and Summary

Regularities among the planets have led astronomers to hypothesize that the Sun and the planets formed together in a giant, spinning cloud of gas and dust called the solar nebula. Astronomical observations show tantalizingly similar circumstellar disks around other stars. Within the solar nebula, material first coalesced into planetesimals; many of these gathered together to make the planets and moons. The remainder can still be seen as comets and asteroids. Probably all planetary systems have formed in similar ways, but many exoplanet systems have evolved along quite different paths, as we will see in Cosmic Samples and the Origin of the Solar System.

For Further Exploration

Collaborative Group Activities

- A. Discuss and make a list of the reasons why we humans might want to explore the other worlds in the solar system. Does your group think such missions of exploration are worth the investment? Why?
- B. Your instructor will assign each group a world. Your task is to think about what it would be like to be there. (Feel free to look ahead in the book to the relevant chapters.) Discuss where on or around your world we would establish a foothold and what we would need to survive there.
- C. In the There's No Place Like Home feature, we discuss briefly how human activity is transforming our planet's overall environment. Can you think of other ways that this is happening?
- D. Some scientists criticized Carl Sagan for "wasting his research

time" popularizing astronomy. To what extent do you think scientists should spend their time interpreting their field of research for the public? Why or why not? Are there ways that scientists who are not as eloquent or charismatic as Carl Sagan or Neil deGrasse Tyson can still contribute to the public understanding of science?

- E. Your group has been named to a special committee by the International Astronomical Union to suggest names of features (such as craters, trenches, and so on) on a newly explored asteroid. Given the restriction that any people after whom features are named must no longer be alive, what names or types of names would you suggest? (Keep in mind that you are not restricted to names of people, by the way.)
- F. A member of your group has been kidnapped by a little-known religious cult that worships the planets. They will release him only if your group can tell which of the planets are currently visible in the sky during the evening and morning. You are forbidden from getting your instructor involved. How and where else could you find out the information you need? (Be as specific as you can. If your instructor says it's okay, feel free to answer this question using online or library resources.)
- G. In the Carl Sagan: Solar System Advocate feature, you learned that science fiction helped spark and sustain his interest in astronomy. Did any of the members of your group get interested in astronomy as a result of a science fiction story, movie, or TV show? Did any of the stories or films you or your group members saw take place on the planets of our solar system? Can you remember any specific ones that inspired you? If no one in the group is into science fiction, perhaps you can interview some friends or classmates who are and report back to the group.
- H. A list of NASA solar system spacecraft missions can be found at http://www.nasa.gov/content/solar-missions-list. Your instructor will assign each group a mission. Look up when the mission was launched and executed, and describe the mission

goals, the basic characteristics of the spacecraft (type of instruments, propellant, size, and so on), and what was learned from the mission. If time allows, each group should present its findings to the rest of the class.

I. What would be some of the costs or risks of developing a human colony or base on another planetary body? What technologies would need to be developed? What would people need to give up to live on a different world in our solar system?

Review Questions

Venus rotates backward and Uranus and Pluto spin about an axis tipped nearly on its side. Based on what you learned about the motion of small bodies in the solar system and the surfaces of the planets, what might be the cause of these strange rotations?

What is the difference between a differentiated body and an undifferentiated body, and how might that influence a body's ability to retain heat for the age of the solar system?

What does a planet need in order to retain an atmosphere? How does an atmosphere affect the surface of a planet and the ability of life to exist?

Which type of planets have the most moons? Where did these moons likely originate?

What is the difference between a meteor and a meteorite?

Explain our ideas about why the terrestrial planets are rocky and have less gas than the giant planets.

Do all planetary systems look the same as our own?

What is comparative planetology and why is it useful to astronomers?

What changed in our understanding of the Moon and Moon-Earth system as a result of humans landing on the Moon's surface?

If Earth was to be hit by an extraterrestrial object, where in the solar system could it come from and how would we know its source region?

List some reasons that the study of the planets has progressed more in the past few decades than any other branch of astronomy.

Imagine you are a travel agent in the next century. An eccentric billionaire asks you to arrange a "Guinness Book of Solar System Records" kind of tour. Where would you direct him to find the following (use this chapter and Appendix F and Appendix G):

- A. the least-dense planet
- B. the densest planet
- C. the largest moon in the solar system
- D. excluding the jovian planets, the planet where you would weigh the most on its surface (Hint: Weight is directly proportional to surface gravity.)
- E. the smallest planet
- F. the planet that takes the longest time to rotate
- G. the planet that takes the shortest time to rotate
- H. the planet with a diameter closest to Earth's
- I. the moon with the thickest atmosphere
- J. the densest moon
- K. the most massive moon

What characteristics do the worlds in our solar system have in common that lead astronomers to believe that they all formed from the same "mother cloud" (solar nebula)?

How do terrestrial and giant planets differ? List as many ways as you can think of.

Why are there so many craters on the Moon and so few on Earth?

How do asteroids and comets differ?

How and why is Earth's Moon different from the larger moons of the giant planets?

Where would you look for some "original" planetesimals left over from the formation of our solar system?

Describe how we use radioactive elements and their decay products to find the age of a rock sample. Is this necessarily the age of the entire world from which the sample comes? Explain.

What was the solar nebula like? Why did the Sun form at its center?

Thought Questions

What can we learn about the formation of our solar system by studying other stars? Explain.

Earlier in this chapter, we modeled the solar system with Earth at a distance of about one city block from the Sun. If you were to make a model of the distances in the solar system to match your height, with the Sun at the top of your head and Pluto at your feet, which planet would be near your waist? How far down would the zone of the terrestrial planets reach?

Seasons are a result of the inclination of a planet's axial tilt being inclined from the normal of the planet's orbital plane. For example, Earth has an axis tilt of 23.4° (Appendix F). Using information about just the inclination alone, which planets might you expect to have

seasonal cycles similar to Earth, although different in duration because orbital periods around the Sun are different?

Again using Appendix F, which planet(s) might you expect not to have significant seasonal activity? Why?

Again using Appendix F, which planets might you expect to have extreme seasons? Why?

Using some of the astronomical resources in your college library or the Internet, find five names of features on each of three other worlds that are named after real people. In a sentence or two, describe each of these people and what contributions they made to the progress of science or human thought.

Explain why the planet Venus is differentiated, but asteroid Fraknoi, a very boring and small member of the asteroid belt, is not.

Would you expect as many impact craters per unit area on the surface of Venus as on the surface of Mars? Why or why not?

Interview a sample of 20 people who are not taking an astronomy class and ask them if they can name a living astronomer. What percentage of those interviewed were able to name one? Typically, the two living astronomers the public knows these days are Stephen Hawking and Neil deGrasse Tyson. Why are they better known than most astronomers? How would your result have differed if you had asked the same people to name a movie star or a professional basketball player?

Using Appendix G, complete the following table that describes the characteristics of the Galilean moons of Jupiter, starting from Jupiter and moving outward in distance.

Table A				
Moon	Semimajor Axis (km ³)	Diameter	Density (g/cm ³)	
Io				
Europa				
Ganymede				
Callisto				

This system has often been described as a mini solar system. Why might this be so? If Jupiter were to represent the Sun and the Galilean moons represented planets, which moons could be considered more terrestrial in nature and which ones more like gas/ice giants? Why? (Hint: Use the values in your table to help explain your categorization.)

Figuring for Yourself

Calculate the density of Jupiter. Show your work. Is it more or less dense than Earth? Why?

Calculate the density of Saturn. Show your work. How does it compare with the density of water? Explain how this can be.

What is the density of Jupiter's moon Europa (see Appendix G for data on moons)? Show your work.

Look at Appendix F and Appendix G and indicate the moon with a diameter that is the largest fraction of the diameter of the planet or dwarf planet it orbits.

Barnard's Star, the second closest star to us, is about 56 trillion (5.6 $\times 10^{12}$) km away. Calculate how far it would be using the scale model of the solar system given in Overview of Our Planetary System.

A radioactive nucleus has a half-life of 5×10^8 years. Assuming that

a sample of rock (say, in an asteroid) solidified right after the solar system formed, approximately what fraction of the radioactive element should be left in the rock today?

Glossary

planetesimals

objects, from tens to hundreds of kilometers in diameter, that formed in the solar nebula as an intermediate step between tiny grains and the larger planetary objects we see today; the comets and some asteroids may be leftover planetesimals

solar nebula

the cloud of gas and dust from which the solar system formed

Chapter 17 Analyzing Starlight Section 17.1: The Brightness of Stars

Star Colors.



Figure 1. This long time exposure shows the colors of the stars. The circular motion of the stars across the image is provided by Earth's rotation. The various colors of the stars are caused by their different temperatures. (credit: modification of work by ESO/ A.Santerne)

Everything we know about stars-how they are born, what they are made of, how far away they are, how long they live, and how they will die-we learn by decoding the messages contained in the light and radiation that reaches Earth. What questions should we ask, and how do we find the answers?

We can begin our voyage to the stars by looking at the night sky. It is obvious that stars do not all appear equally bright, nor are they all the same color. To understand the stars, we must first determine their basic properties, such as what their temperatures are, how much material they contain (their masses), and how much energy they produce. Since our Sun is a star, of course the same techniques, including spectroscopy, used to study the Sun can be used to find out what stars are like. As we learn more about the stars, we will use these characteristics to begin assembling clues to the main problems we are interested in solving: How do stars form? How long do they survive? What is their ultimate fate?

17.1 The Brightness of Stars



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 firstmagnitude star. Measurements showed that we receive about 100 times more light from a first-magnitude star than from a sixthmagnitude 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 $2.5 \times 2.5 = 6.25$ 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 1 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.

Apparent Magnitudes of Well-Known Objects.



Figure 1. The faintest magnitudes that can be detected by the

416 | Chapter 17 Analyzing Starlight Section 17.1: The Brightness of Stars

unaided eye, binoculars, and large telescopes are also shown.

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:

$$m_1 - m_2 = 2.5 \log(\frac{b_2}{b_1}) \operatorname{OR} \frac{b_2}{b_1} = 2.5^{m_1 - m_2}$$

Here is another way to write this equation:

$$\frac{b_2}{b_1} = (100^{0.2})^{m_1 - m_2}$$

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:

$$\frac{b_2}{b_1} = (100^{0.2})^{8.5 - (-1.5)} = (100^{0.2})^{10}$$

=(100)² = 100 X 100 = 10,000

Check Your Learning

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:

 $\frac{bSirius}{bPolaris} = (100^{0.2})^{2.0 - (-1.5)} = (100^{0.2})^{3.5} = 100^{0.7} = 25$

(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.

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} .

Key Concepts and Summary

The total energy emitted per second by a star is called its luminosity. How bright a star looks from the perspective of Earth is its apparent brightness. The apparent brightness of a star depends on both its luminosity and its distance from Earth. Thus, the determination of apparent brightness and measurement of the distance to a star provide enough information to calculate its luminosity. The apparent brightnesses of stars are often expressed in terms of magnitudes, which is an old system based on how human vision interprets relative light intensity.

Glossary

apparent brightness

a measure of the amount of light received by Earth from a star or other object-that is, how bright an object appears in the sky, as contrasted with its luminosity

luminosity

the rate at which a star or other object emits electromagnetic energy into space; the total power output of an object

magnitude

an older system of measuring the amount of light we receive from a star or other luminous object; the larger the magnitude, the less radiation we receive from the object

Chapter 17 Section 17.2: Colors of Stars

17.2 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 <u>Figure 1</u>. 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.

Sagittarius Star Cloud.



Figure 1. 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). 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.

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

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**. Go to this light and filters simulator for a demonstration of how different light sources and filters can combine to determine the observed spectrum. You can also see how the perceived colors are associated with the spectrum.

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.

Key Concepts and Summary

Stars have different colors, which are indicators of temperature. The hottest stars tend to appear blue or blue-white, whereas the coolest stars are red. A color index of a star is the difference in the magnitudes measured at any two wavelengths and is one way that astronomers measure and express the temperature of stars.

Glossary

color index

difference between the magnitudes of a star or other object measured in light of two different spectral regions-for example, blue minus visual (B-V) magnitudes

Chapter 17 Section 17.3: The Spectra of Stars (and Brown Dwarfs)

17.3 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 1) 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.

William Huggins (1824–1910) and Margaret Huggins (1848–1915).





Figure 1. 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.

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 <u>Figure 2</u> 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 classes is summarized in Table.

Absorption Lines in Stars of Different Temperatures.



^{6000-5200, 5200-3700,} and 3700-2400 on the right. Six curves are

plotted, each peaking as follows (from left to right): ionized helium peaks at spectral type O, neutral helium peaks at B, hydrogen peaks at about A, ionized metals peak between F and G, neutral metals peak at K, and molecules peak beyond M at right." width="700" height="516">

Figure 2. 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.

Spectral Classes for Stars

Spectral Class	Color	Approximate Temperature (K)	Principal Features	Examples
0	Blue	> 30,000	Neutral and ionized helium lines, weak hydrogen lines	10 Lacertae
В	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
К	Orange	3700-5200	Very weak hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of neutral metals	Arcturus, Aldebaran
М	Red	2400-3700	Strong lines of neutral metals and molecular bands of titanium oxide dominate	Betelgeuse , Antares

Spectral Classes for Stars

Spectral Class	Color	Approximate Temperature (K)	Principal Features	Examples
L	Red	1300-2400	Metal hydride lines, alkali metal lines (e.g., sodium, potassium, rubidium)	Teide 1
Т	Magenta	700–1300	Methane lines	Gliese 229B
Y	Infrared ¹	< 700	Ammonia lines	WISE 1828+2650

To see how spectral classification works, let's use Figure 2. 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 <u>Figure 3</u>, 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.

Spectra of Stars with Different Spectral Classes.



Figure 3. 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 2 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 2). (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.

ANNIE CANNON: CLASSIFIER OF THE STARS

Annie Jump **Cannon** was born in Delaware in 1863 (Figure 4). 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.

Annie Jump Cannon (1863–1941).



Figure 4. 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 twentiethcentury 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 5). 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 (CH4) 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 (NH₃), which are not seen in T dwarfs. A new spectral class, Y, was created for these objects. As of 2015, over two dozen brown dwarfs belonging to spectral class Y have been discovered, some with temperatures comparable to that of the human body (about 300 K).

Brown Dwarfs.



Y2 / 300-500 K for WISE1828, and no spectral type / 125 K for Jupiter. The main portion of the figure shows the stars to scale, with the Sun by far the largest, followed by Gliese 229A being about ½ the size of the Sun, then Teide 1, Gliese 229B, and WISE1828, each being about the size of Jupiter." width="845" height="683">

Figure 5. 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 MJ or 0.012 M_{Sun}), it is a brown dwarf. Objects with less than 13 MJ do not fuse deuterium and are usually considered planets.

Key Concepts and Summary

The differences in the spectra of stars are principally due to differences in temperature, not composition. The spectra of stars are described in terms of spectral classes. In order of decreasing temperature, these spectral classes are O, B, A, F, G, K, M, L, T, and Y. These are further divided into subclasses numbered from 0 to 9. The classes L, T, and Y have been added recently to describe newly discovered star-like objects-mainly brown dwarfs-that are cooler than M9. Our Sun has spectral type G2.

Footnotes

• 1 Absorption by sodium and potassium atoms makes Y dwarfs appear a bit less red than L dwarfs.

Glossary

brown dwarf

an object intermediate in size between a planet and a star; the approximate mass range is from about 1/100 of the mass of the Sun up to the lower mass limit for self-sustaining nuclear reactions, which is about 0.075 the mass of the Sun; brown dwarfs are capable of deuterium fusion, but not hydrogen fusion

spectral class

(or spectral type) the classification of stars according to their temperatures using the characteristics of their spectra; the types are O, B, A, F, G, K, and M with L, T, and Y added recently for cooler star-like objects that recent survey have revealed

For Further Exploration

Articles

Berman, B. "Magnitude Cum Laude." Astronomy (December 1998): 92. How we measure the apparent brightnesses of stars is discussed.

Dvorak, J. "The Women Who Created Modern Astronomy [including Annie Cannon]." Sky & Telescope (August 2013): 28.

Hearnshaw, J. "Origins of the Stellar Magnitude Scale." Sky & Telescope (November 1992): 494. A good history of how we have come to have this cumbersome system is discussed.

Hirshfeld, A. "The Absolute Magnitude of Stars." Sky & *Telescope* (September 1994): 35.

Kaler, J. "Stars in the Cellar: Classes Lost and Found." Sky & Telescope (September 2000): 39. An introduction is provided for spectral types and the new classes L and T.

Kaler, J. "Origins of the Spectral Sequence." Sky & Telescope (February 1986): 129.

Skrutskie, M. "2MASS: Unveiling the Infrared Universe." Sky & Telescope (July 2001): 34. This article focuses on an all-sky survey at 2 microns.

Sneden, C. "Reading the Colors of the Stars." Astronomy (April 1989): 36. This article includes a discussion of what we learn from spectroscopy.

Steffey, P. "The Truth about Star Colors." Sky & *Telescope* (September 1992): 266. The color index and how the eye and film "see" colors are discussed.

Tomkins, J. "Once and Future Celestial Kings." Sky & Telescope (April 1989): 59. Calculating the motion of stars and determining which stars were, are, and will be brightest in the sky are discussed.

Websites

Discovery of Brown Dwarfs: http://w.astro.berkeley.edu/~basri/bdwarfs/SciAm-book.pdf.

Listing of Nearby Brown Dwarfs: http://www.solstation.com/ stars/pc10bd.htm.

Spectral Types of Stars: http://www.skyandtelescope.com/astronomy-equipment/the-spectral-types-of-stars/.

Stellar Velocities https://www.e-education.psu.edu/astro801/ content/l4_p7.html.

Unheard Voices! The Contributions of Women to Astronomy: A Resource Guide: http://multiverse.ssl.berkeley.edu/ women and http://www.astrosociety.org/education/astronomyresource-guides/women-in-astronomy-an-introductoryresource-guide/.

Videos

When You Are Just Too Small to be a Star: https://www.youtube.com/watch?v=zXCDsb4n4KU. 2013 Public Talk on Brown Dwarfs and Planets by Dr. Gibor Basri of the University of California–Berkeley (1:32:52).

Collaborative Group Activities

A. The Voyagers in Astronomy feature on Annie Cannon: Classifier of the Stars discusses some of the difficulties women who wanted to do astronomy faced in the first half of the twentieth century. What does your group think about the situation for women today? Do men and women have an equal chance to become scientists? Discuss with your group whether, in your experience, boys and girls were equally encouraged to do science and math where you went to school.

- B. In the section on magnitudes in The Brightness of Stars, we discussed how this old system of classifying how bright different stars appear to the eye first developed. Your authors complained about the fact that this old system still has to be taught to every generation of new students. Can your group think of any other traditional systems of doing things in science and measurement where tradition rules even though common sense says a better system could certainly be found. Explain. (Hint: Try Daylight Savings Time, or metric versus English units.)
- C. Suppose you could observe a star that has only one spectral line. Could you tell what element that spectral line comes from? Make a list of reasons with your group about why you answered yes or no.
- D. A wealthy alumnus of your college decides to give \$50 million to the astronomy department to build a world-class observatory for learning more about the characteristics of stars. Have your group discuss what kind of equipment they would put in the observatory. Where should this observatory be located? Justify your answers. (You may want to refer back to the Astronomical Instruments chapter and to revisit this question as you learn more about the stars and equipment for observing them in future chapters.)
- E. For some astronomers, introducing a new spectral type for the stars (like the types L, T, and Y discussed in the text) is similar to introducing a new area code for telephone calls. No one likes to disrupt the old system, but sometimes it is simply necessary. Have your group make a list of steps an astronomer would have to go through to persuade colleagues that a new spectral class is needed.

Review Questions

What two factors determine how bright a star appears to be in the sky?

Explain why color is a measure of a star's temperature.

What is the main reason that the spectra of all stars are not identical? Explain.

What elements are stars mostly made of? How do we know this?

What did Annie Cannon contribute to the understanding of stellar spectra?

Name five characteristics of a star that can be determined by measuring its spectrum. Explain how you would use a spectrum to determine these characteristics.

How do objects of spectral types L, T, and Y differ from those of the other spectral types?

Do stars that look brighter in the sky have larger or smaller magnitudes than fainter stars?

The star Antares has an apparent magnitude of 1.0, whereas the star Procyon has an apparent magnitude of 0.4. Which star appears brighter in the sky?

Based on their colors, which of the following stars is hottest? Which is coolest? Archenar (blue), Betelgeuse (red), Capella (yellow).

Order the seven basic spectral types from hottest to coldest.

What is the defining difference between a brown dwarf and a true star?

Thought Questions

If the star Sirius emits 23 times more energy than the Sun, why does the Sun appear brighter in the sky?

How would two stars of equal luminosity—one blue and the other red—appear in an image taken through a filter that passes mainly blue light? How would their appearance change in an image taken through a filter that transmits mainly red light?

[link] lists the temperature ranges that correspond to the different spectral types. What part of the star do these temperatures refer to? Why?

Suppose you are given the task of measuring the colors of the brightest stars, listed in Appendix J, through three filters: the first transmits blue light, the second transmits yellow light, and the third transmits red light. If you observe the star Vega, it will appear equally bright through each of the three filters. Which stars will appear brighter through the blue filter than through the red filter? Which stars will appear brighter through the red filter? Which stars is likely to have colors most nearly like those of Vega?

Star X has lines of ionized helium in its spectrum, and star Y has bands of titanium oxide. Which is hotter? Why? The spectrum of star Z shows lines of ionized helium and also molecular bands of titanium oxide. What is strange about this spectrum? Can you suggest an explanation?

The spectrum of the Sun has hundreds of strong lines of nonionized iron but only a few, very weak lines of helium. A star of spectral type B has very strong lines of helium but very weak iron lines. Do these differences mean that the Sun contains more iron and less helium than the B star? Explain. What are the approximate spectral classes of stars with the following characteristics?

- A. Balmer lines of hydrogen are very strong; some lines of ionized metals are present.
- B. The strongest lines are those of ionized helium.
- C. Lines of ionized calcium are the strongest in the spectrum; hydrogen lines show only moderate strength; lines of neutral and metals are present.
- D. The strongest lines are those of neutral metals and bands of titanium oxide.

Look at the chemical elements in Appendix K. Can you identify any relationship between the abundance of an element and its atomic weight? Are there any obvious exceptions to this relationship?

Appendix I lists some of the nearest stars. Are most of these stars hotter or cooler than the Sun? Do any of them emit more energy than the Sun? If so, which ones?

Appendix J lists the stars that appear brightest in our sky. Are most of these hotter or cooler than the Sun? Can you suggest a reason for the difference between this answer and the answer to the previous question? (Hint: Look at the luminosities.) Is there any tendency for a correlation between temperature and luminosity? Are there exceptions to the correlation?

What star appears the brightest in the sky (other than the Sun)? The second brightest? What color is Betelgeuse? Use Appendix J to find the answers.

Suppose hominids one million years ago had left behind maps of the night sky. Would these maps represent accurately the sky that we see today? Why or why not?

Why can only a lower limit to the rate of stellar rotation be

determined from line broadening rather than the actual rotation rate? (Refer to Figure.)

Why do you think astronomers have suggested three different spectral types (L, T, and Y) for the brown dwarfs instead of M? Why was one not enough?

Sam, a college student, just bought a new car. Sam's friend Adam, a graduate student in astronomy, asks Sam for a ride. In the car, Adam remarks that the colors on the temperature control are wrong. Why did he say that?



(credit: modification of work by Michael Sheehan)

Would a red star have a smaller or larger magnitude in a red filter than in a blue filter?

Two stars have proper motions of one arcsecond per year. Star A is 20 light-years from Earth, and Star B is 10 light-years away from Earth. Which one has the faster velocity in space?

Suppose there are three stars in space, each moving at 100 km/s. Star A is moving across (i.e., perpendicular to) our line of sight, Star B is moving directly away from Earth, and Star C is moving away from Earth, but at a 30° angle to the line of sight. From which star will you observe the greatest Doppler shift? From which star will you observe the smallest Doppler shift?

What would you say to a friend who made this statement, "The visible-light spectrum of the Sun shows weak hydrogen lines and strong calcium lines. The Sun must therefore contain more calcium than hydrogen."?

Figuring for Yourself

In Appendix J, how much more luminous is the most luminous of the stars than the least luminous?

For Exercise through Exercise, use the equations relating magnitude and apparent brightness given in the section on the magnitude scale in The Brightness of Stars and [link].

Verify that if two stars have a difference of five magnitudes, this corresponds to a factor of 100 in the ratio (b2b1);(b2b1); that 2.5 magnitudes corresponds to a factor of 10; and that 0.75 magnitudes corresponds to a factor of 2.

As seen from Earth, the Sun has an apparent magnitude of about –26.7. What is the apparent magnitude of the Sun as seen from Saturn, about 10 AU away? (Remember that one AU is the distance from Earth to the Sun and that the brightness decreases as the inverse square of the distance.) Would the Sun still be the brightest star in the sky?

An astronomer is investigating a faint star that has recently been discovered in very sensitive surveys of the sky. The star has a magnitude of 16. How much less bright is it than Antares, a star with magnitude roughly equal to 1?

The center of a faint but active galaxy has magnitude 26. How much

452 | Chapter 17 Section 17.3: The Spectra of Stars (and Brown Dwarfs)

less bright does it look than the very faintest star that our eyes can see, roughly magnitude 6?

You have enough information from this chapter to estimate the distance to Alpha Centauri, the second nearest star, which has an apparent magnitude of 0. Since it is a G2 star, like the Sun, assume it has the same luminosity as the Sun and the difference in magnitudes is a result only of the difference in distance. Estimate how far away Alpha Centauri is. Describe the necessary steps in words and then do the calculation. (As we will learn in the Celestial Distances chapter, this method–namely, assuming that stars with identical spectral types emit the same amount of energy–is actually used to estimate distances to stars.) If you assume the distance to the Sun is in AU, your answer will come out in AU.

Do the previous problem again, this time using the information that the Sun is 150,000,000 km away. You will get a very large number of km as your answer. To get a better feeling for how the distances compare, try calculating the time it takes light at a speed of 299,338 km/s to travel from the Sun to Earth and from Alpha Centauri to Earth. For Alpha Centauri, figure out how long the trip will take in years as well as in seconds.

Star A and Star B have different apparent brightnesses but identical luminosities. If Star A is 20 light-years away from Earth and Star B is 40 light-years away from Earth, which star appears brighter and by what factor?

Star A and Star B have different apparent brightnesses but identical luminosities. Star A is 10 light-years away from Earth and appears 36 times brighter than Star B. How far away is Star B?

The star Sirius A has an apparent magnitude of -1.5. Sirius A has a dim companion, Sirius B, which is 10,000 times less bright than Sirius A. What is the apparent magnitude of Sirius B? Can Sirius B be seen with the naked eye?

Our Sun, a type G star, has a surface temperature of 5800 K. We know, therefore, that it is cooler than a type O star and hotter than a type M star. Given what you learned about the temperature ranges of these types of stars, how many times hotter than our Sun is the hottest type O star? How many times cooler than our Sun is the coolest type M star?

Glossary

giant

a star of exaggerated size with a large, extended photosphere

proper motion

the angular change per year in the direction of a star as seen from the Sun

radial velocity

motion toward or away from the observer; the component of relative velocity that lies in the line of sight

space velocity

the total (three-dimensional) speed and direction with which an object is moving through space relative to the Sun
Chapter 18 The Stars: A Celestial Census Section 18.1: A Stellar Census

Thinking Ahead

Variety of Stars.



Figure 1. Stars come in a variety of sizes, masses, temperatures, and luminosities. This image shows part of a cluster of stars in the Small Magellanic Cloud (catalog number NGC 290). Located about 200,000 light-years away, NGC 290 is about 65 light-years across. Because the stars in this cluster are all at about the same distance from us, the differences in apparent brightness correspond to differences in luminosity; differences in temperature account for the differences in color. The various colors and luminosities of these stars provide clues about their life stories. (credit: modification of work by E. Olszewski (University of Arizona), European Space Agency, NASA) How do stars form? How long do they live? And how do they die? Stop and think how hard it is to answer these questions.

Stars live such a long time that nothing much can be gained from staring at one for a human lifetime. To discover how stars evolve from birth to death, it was necessary to measure the characteristics of many stars (to take a celestial census, in effect) and then determine which characteristics help us understand the stars' life stories. Astronomers tried a variety of hypotheses about stars until they came up with the right approach to understanding their development. But the key was first making a thorough census of the stars around us.

18.1 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 shows an estimate of the number of stars of each spectral type¹ 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.

Spectral Type	Number of Stars
A	2
F	1
G	7
K	17
М	94
White dwarfs	8
Brown dwarfs	33

Stars within 21 Light-Years of the Sun

This table is based on data published through 2015, and it is likely that more faint objects remain to be discovered (see Figure 1). Along with the L and T brown dwarfs already observed in our neighborhood, astronomers expect to find perhaps hundreds of additional T 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.

Dwarf Simulation.



Figure 1. This computer simulation shows the stars in our neighborhood as they would be seen from a distance of 30 lightyears 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 (<u>Figure 2</u>). Why are we missing most of the brightest stars when we take our census of the local neighborhood?

The Closest Stars.



Figure 2. (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 6000 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.

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.

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.

Key Concepts and Summary

To understand the properties of stars, we must make wide-ranging surveys. We find the stars that appear brightest to our eyes are bright primarily because they are intrinsically very luminous, not because they are the closest to us. Most of the nearest stars are intrinsically so faint that they can be seen only with the aid of a telescope. Stars with low mass and low luminosity are much more common than stars with high mass and high luminosity. Most of the brown dwarfs in the local neighborhood have not yet been discovered.

Footnotes

• 1 The spectral types of stars were defined and discussed in Analyzing Starlight.

Glossary

selection effect

the selection of sample data in a nonrandom way, causing the sample data to be unrepresentative of the entire data set

Chapter 18 Section 18.2: Measuring Stellar Masses

18.2 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 mass**es 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 1. A binary star system in which both of the stars can be seen with a telescope is called a **visual binary**.

Revolution of a Binary Star.



Figure 1. 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 2).

Binary Star System.



Figure 2. 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 3 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 3.

Motions of Two Stars Orbiting Each Other and

What the Spectrum Shows.



Figure 3. 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.

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 <u>Figure 3</u> is shown in <u>Figure 4</u>.



Radial Velocities in a Spectroscopic Binary System.

Figure 4. 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 relative to the center of mass of the system, which in this case is 40 kilometers per second. Negative velocity means the star is moving toward us relative to the center of mass. The positions on the curve corresponding to the illustrations in <u>Figure 3</u> are marked with the diagram number (1–4).

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 4) 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 justdescribed 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:

$$M_1 + M_2 = \frac{8000}{2500} = 3.2$$

472 | Chapter 18 Section 18.2: Measuring Stellar Masses

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 5). Brown dwarfs are similar to **Jupiter** in radius but have masses from approximately 13 to 80 times larger than the mass of Jupiter.¹

Brown Dwarfs in Orion.



Figure 5. 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 6. 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.

Mass-Luminosity Relation.





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.

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:

$$L/L_{sun} = (M/M_{sun})^4$$

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:

$$M/M_{sun} = (L/L_{sun})^{0.25} = (L/L_{sun})^{0.25}$$

Check Your Learning

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:

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:

$$M/M_{sun} = 23^{0.25} = 2.2$$

The mass of the companion star to Sirius is then 3.2 - 2.2 = 1.0 solar mass.

Notice how good this mass-luminosity relationship is. Most stars (see Figure 6) 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 massluminosity relation. Later, we will explore why such a relationship exists and what we can learn from the roughly 10% of stars that "disobey" it.

Key Concepts and Summary

The masses of stars can be determined by analysis of the orbit of binary stars—two stars that orbit a common center of mass. In visual binaries, the two stars can be seen separately in a telescope, whereas in a spectroscopic binary, only the spectrum reveals the presence of two stars. Stellar masses range from about 1/12 to more than 100 times the mass of the Sun (in rare cases, going to 250 times the Sun's mass). Objects with masses between 1/12 and 1/100 that of the Sun are called brown dwarfs. Objects in which no nuclear reactions can take place are planets. The most massive stars are, in most cases, also the most luminous, and this correlation is known as the mass-luminosity relation.

Footnotes

 1 Exactly 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.

Glossary

binary stars

two stars that revolve about each other

brown dwarf

an object intermediate in size between a planet and a star; the approximate mass range is from about 1/100 of the mass of the Sun up to the lower mass limit for self-sustaining nuclear reactions, which is about 1/12 the mass of the Sun

mass-luminosity relation

the observed relation between the masses and luminosities of

many (90% of all) stars

spectroscopic binary

a binary star in which the components are not resolved but whose binary nature is indicated by periodic variations in radial velocity, indicating orbital motion

visual binary

a binary star in which the two components are telescopically resolved

Chapter 18 Section 18.3: Diameters of Stars

18.3 Diameters of Stars

Learning Objectives

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

- Describe the methods used to determine **star diameter**s
- 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 $1/2^{\circ}$. 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 1). When one star blocks the light of the other, preventing it from reaching Earth, the luminosity of the system decreases, and astronomers say that an eclipse has occurred.

Light Curve of an Eclipsing Binary.

482 | Chapter 18 Section 18.3: Diameters of Stars





Figure 1. 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 longstanding 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 1). But from our vantage point on Earth, only a few binary star systems are oriented in this way.

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."¹ 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 2. 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 (*thirdcontact*), 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 2. 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 2. 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.

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:

$$\frac{L_{Sirius}}{L_{companion}} = \frac{(A_{Sirius} \times F_{Sirius})}{(A_{companion} \times F_{companion})}$$
$$= \frac{A_{Sirius}}{A_{companion}} = \frac{4\pi R^2 Sirius}{4\pi R^2 companion} = \frac{R^2 Sirius}{R^2 companion}$$
$$\frac{L_{Sirius}}{L_{companion}} = 8200 = \frac{R^2 Sirius}{R^2 companion}$$

490 | Chapter 18 Section 18.3: Diameters of Stars
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 radium of its faint companion.

The method for determining the radius shown here requires both stars be visible, which is not always the case.

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.

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.

Key Concepts and Summary

The diameters of stars can be determined by measuring the time it takes an object (the Moon, a planet, or a companion star) to pass in front of it and block its light. Diameters of members of eclipsing binary systems (where the stars pass in front of each other) can be determined through analysis of their orbital motions.

Footnotes

• 1 Fans of Batman comic books and movies will recognize that this name was given to an archvillain in the series.

Glossary

eclipsing binary

a binary star in which the plane of revolution of the two stars is nearly edge-on to our line of sight, so that the light of one star is periodically diminished by the other passing in front of it

Chapter 18 Section 18.4: The H-R Diagram

18.4 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 <u>Table</u>. 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.

Measuring the Characteristics of Stars

Technique

Characteristic

	roomiquo
Surface temperature	 Determine the color (very rough). 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	 Measure the way a star's light is blocked by the Moon. 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 <u>Figure 1</u> 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. Height versus Weight.



Figure 1. 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 2).

Hertzsprung (1873–1967) and Russell (1877–1957).







(b)

Figure 2. (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.

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 3). 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.

H–R Diagram for a Selected Sample of Stars.



Figure 3. 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 1). 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 4 is a schematic H–R diagram for a large sample of stars, drawn to make the different types more apparent.

Schematic H–R Diagram for Many Stars.



Figure 4. 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 <u>Figure 3</u> 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 onethird 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 3). The characteristics of representative **mainsequence stars** (excluding brown dwarfs, which are not true stars) are listed in Table.

Spectral Type	Mass (Sun = 1)	Luminosity (Sun = 1)	Temperature	Radius (Sun = 1)
O5	40	7×10^5	40,000 K	18
BO	16	2.7×10^5	28,000 K	7
AO	3.3	55	10,000 K	2.5
FO	1.7	5	7500 K	1.4
G0	1.1	1.4	6000 K	1.1
К0	0.8	0.35	5000 K	0.8
M0	0.4	0.05	3500 K	0.6

Characteristics of Main-Sequence Stars

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 ([link]). 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 5). 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.



The Sun and a Supergiant.

Chapter 18 Section 18.4: The H-R Diagram | 507

Figure 5. 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 6). (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.)

Two Views of Sirius and Its White Dwarf Companion.



Figure 6. (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. [link] 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/cm³. 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.

Key Concepts and Summary

The Hertzsprung–Russell diagram, or H–R diagram, is a plot of stellar luminosity against surface temperature. Most stars lie on the main sequence, which extends diagonally across the H–R diagram from high temperature and high luminosity to low temperature and low luminosity. The position of a star along the main sequence is determined by its mass. High-mass stars emit more energy and are

hotter than low-mass stars on the main sequence. Main-sequence stars derive their energy from the fusion of protons to helium. About 90% of the stars lie on the main sequence. Only about 10% of the stars are white dwarfs, and fewer than 1% are giants or supergiants.

For Further Exploration

Articles

Croswell, K. "The Periodic Table of the Cosmos." *Scientific American* (July 2011):45–49. A brief introduction to the history and uses of the H–R diagram.

Davis, J. "Measuring the Stars." Sky & Telescope (October 1991): 361. The article explains direct measurements of stellar diameters.

DeVorkin, D. "Henry Norris Russell." Scientific American (May 1989): 126.

Kaler, J. "Journeys on the H-R Diagram." Sky & Telescope (May 1988): 483.

McAllister, H. "Twenty Years of Seeing Double." Sky & Telescope (November 1996): 28. An update on modern studies of binary stars.

Parker, B. "Those Amazing White Dwarfs." Astronomy (July 1984): 15. The article focuses on the history of their discovery.

Pasachoff, J. "The H–R Diagram's 100th Anniversary." Sky & Telescope (June 2014): 32.

Roth, J., and Sinnott, R. "Our Studies of Celestial Neighbors." Sky & Telescope (October 1996): 32. A discussion is provided on finding the nearest stars.

Websites

Eclipsing Binary Stars: http://www.midnightkite.com/ index.aspx?URL=Binary. Dan Bruton at Austin State University has created this collection of animations, articles, and links showing how astronomers use eclipsing binary light curves.

Henry Norris Russell: http://www.nasonline.org/publications/ biographical-memoirs/memoir-pdfs/russell-henry-n.pdf. A biographic memoir by Harlow Shapley.

Henry Norris Russell: http://www.phys-astro.sonoma.edu/ brucemedalists/russell/RussellBio.pdf. A Bruce Medal profile of Russell.

Hertzsprung–Russell Diagram: http://skyserver.sdss.org/dr1/ en/proj/advanced/hr/. This site from the Sloan Digital Sky Survey introduces the H–R diagram and gives you information for making your own. You can go step by step by using the menu at the left. Note that in the project instructions, the word "here" is a link and takes you to the data you need.

Stars of the Week: http://stars.astro.illinois.edu/sow/ sowlist.html. Astronomer James Kaler does "biographical summaries" of famous stars-not the Hollywood type, but ones in the real sky.

Videos

WISE Mission Surveys Nearby Stars: http://www.jpl.nasa.gov/ video/details.php?id=1089. Short video about the WISE telescope survey of brown dwarfs and M dwarfs in our immediate neighborhood (1:21).

Collaborative Group Activities

- A. Two stars are seen close together in the sky, and your group is given the task of determining whether they are a visual binary or whether they just happen to be seen in nearly the same direction. You have access to a good observatory. Make a list of the types of measurements you would make to determine whether they orbit each other.
- B. Your group is given information about five main sequence stars that are among the brightest-appearing stars in the sky and yet are pretty far away. Where would these stars be on the H-R diagram and why? Next, your group is given information about five main-sequence stars that are typical of the stars closest to us. Where would these stars be on the H-R diagram and why?
- C. A very wealthy (but eccentric) alumnus of your college donates a lot of money for a fund that will help in the search for more brown dwarfs. Your group is the committee in charge of this fund. How would you spend the money? (Be as specific as you can, listing instruments and observing programs.)
- D. Use the internet to search for information about the stars with the largest known diameter. What star is considered the record holder (this changes as new measurements are made)? Read about some of the largest stars on the web. Can your group list some reasons why it might be hard to know which star is the largest?
- E. Use the internet to search for information about stars with the largest mass. What star is the current "mass champion" among stars? Try to research how the mass of one or more of the most massive stars was measured, and report to the group or the whole class.

Review Questions

How does the mass of the Sun compare with that of other stars in our local neighborhood?

Name and describe the three types of binary systems.

Describe two ways of determining the diameter of a star.

What are the largest- and smallest-known values of the mass, luminosity, surface temperature, and diameter of stars (roughly)?

You are able to take spectra of both stars in an eclipsing binary system. List all properties of the stars that can be measured from their spectra and light curves.

Sketch an H–R diagram. Label the axes. Show where cool supergiants, white dwarfs, the Sun, and main-sequence stars are found.

Describe what a typical star in the Galaxy would be like compared to the Sun.

How do we distinguish stars from brown dwarfs? How do we distinguish brown dwarfs from planets?

Describe how the mass, luminosity, surface temperature, and radius of main-sequence stars change in value going from the "bottom" to the "top" of the main sequence.

One method to measure the diameter of a star is to use an object like the Moon or a planet to block out its light and to measure the time it takes to cover up the object. Why is this method used more often with the Moon rather than the planets, even though there are more planets?

We discussed in the chapter that about half of stars come in pairs,

or multiple star systems, yet the first eclipsing binary was not discovered until the eighteenth century. Why?

Thought Questions

Is the Sun an average star? Why or why not?

Suppose you want to determine the average educational level of people throughout the nation. Since it would be a great deal of work to survey every citizen, you decide to make your task easier by asking only the people on your campus. Will you get an accurate answer? Will your survey be distorted by a selection effect? Explain.

Why do most known visual binaries have relatively long periods and most spectroscopic binaries have relatively short periods?

[link] shows the light curve of a hypothetical eclipsing binary star in which the light of one star is completely blocked by another. What would the light curve look like for a system in which the light of the smaller star is only partially blocked by the larger one? Assume the smaller star is the hotter one. Sketch the relative positions of the two stars that correspond to various portions of the light curve.

There are fewer eclipsing binaries than spectroscopic binaries. Explain why.

Within 50 light-years of the Sun, visual binaries outnumber eclipsing binaries. Why?

Which is easier to observe at large distances–a spectroscopic binary or a visual binary?

The eclipsing binary Algol drops from maximum to minimum brightness in about 4 hours, remains at minimum brightness for 20 minutes, and then takes another 4 hours to return to maximum brightness. Assume that we view this system exactly edge-on, so that one star crosses directly in front of the other. Is one star much larger than the other, or are they fairly similar in size? (Hint: Refer to the diagrams of eclipsing binary light curves.)

Review this spectral data for five stars.

Table A			
Star	Spectrum		
1	G, main sequence		
2	K, giant		
3	K, main sequence		
4	O, main sequence		
5	M, main sequence		

Which is the hottest? Coolest? Most luminous? Least luminous? In each case, give your reasoning.

Which changes by the largest factor along the main sequence from spectral types O to M-mass or luminosity?

Suppose you want to search for brown dwarfs using a space telescope. Will you design your telescope to detect light in the ultraviolet or the infrared part of the spectrum? Why?

An astronomer discovers a type-M star with a large luminosity. How is this possible? What kind of star is it?

Approximately 6000 stars are bright enough to be seen without a telescope. Are any of these white dwarfs? Use the information given in this chapter to explain your reasoning.

Use the data in Appendix J to plot an H–R diagram for the brightest stars. Use the data from Table to show where the main sequence lies. Do 90% of the brightest stars lie on or near the main sequence? Explain why or why not.

Use the diagram you have drawn for Exercise to answer the following questions: Which star is more massive—Sirius or Alpha Centauri? Rigel and Regulus have nearly the same spectral type. Which is larger? Rigel and Betelgeuse have nearly the same luminosity. Which is larger? Which is redder?

Use the data in Appendix I to plot an H–R diagram for this sample of nearby stars. How does this plot differ from the one for the brightest stars in Exercise? Why?

If a visual binary system were to have two equal-mass stars, how would they be located relative to the center of the mass of the system? What would you observe as you watched these stars as they orbited the center of mass, assuming very circular orbits, and assuming the orbit was face on to your view?

Two stars are in a visual binary star system that we see face on. One star is very massive whereas the other is much less massive. Assuming circular orbits, describe their relative orbits in terms of orbit size, period, and orbital velocity.

Describe the spectra for a spectroscopic binary for a system comprised of an F-type and L-type star. Assume that the system is too far away to be able to easily observe the L-type star.

[link] shows the velocity of two stars in a spectroscopic binary system. Which star is the most massive? Explain your reasoning.

You go out stargazing one night, and someone asks you how far away the brightest stars we see in the sky without a telescope are. What would be a good, general response? (Use Appendix J for more information.)

If you were to compare three stars with the same surface temperature, with one star being a giant, another a supergiant, and the third a main-sequence star, how would their radii compare to one another? Are supergiant stars also extremely massive? Explain the reasoning behind your answer.

Consider the following data on four stars:

Table B

Star	Luminosity (in L_{Sun})	Туре
1	100	B, main sequence
2	1/100	B, white dwarf
3	1/100	M, main sequence
4	100	M, giant

Which star would have the largest radius? Which star would have the smallest radius? Which star is the most common in our area of the Galaxy? Which star is the least common?

Figuring for Yourself

If two stars are in a binary system with a combined mass of 5.5 solar masses and an orbital period of 12 years, what is the average distance between the two stars?

It is possible that stars as much as 200 times the Sun's mass or more exist. What is the luminosity of such a star based upon the mass-luminosity relation?

The lowest mass for a true star is 1/12 the mass of the Sun. What is the luminosity of such a star based upon the mass-luminosity relationship?

Spectral types are an indicator of temperature. For the first 10 stars in Appendix J, the list of the brightest stars in our skies, estimate their temperatures from their spectral types. Use information in the figures and/or tables in this chapter and describe how you made the estimates.

We can estimate the masses of most of the stars in Appendix J from the mass-luminosity relationship in [link]. However, remember this relationship works only for main sequence stars. Determine which of the first 10 stars in Appendix J are main sequence stars. Use one of the figures in this chapter. Make a table of stars' masses.

In Diameters of Stars, the relative diameters of the two stars in the Sirius system were determined. Let's use this value to explore other aspects of this system. This will be done through several steps, each in its own exercise. Assume the temperature of the Sun is 5800 K, and the temperature of Sirius A, the larger star of the binary, is

10,000 K. The luminosity of Sirius A can be found in Appendix J, and is given as about 23 times that of the Sun. Using the values provided, calculate the radius of Sirius A relative to that of the Sun.

Now calculate the radius of Sirius' white dwarf companion, Sirius B, to the Sun.

How does this radius of Sirius B compare with that of Earth?

From the previous calculations and the results from Diameters of Stars, it is possible to calculate the density of Sirius B relative to the Sun. It is worth noting that the radius of the companion is very similar to that of Earth, whereas the mass is very similar to the Sun's. How does the companion's density compare to that of the Sun? Recall that density = mass/volume, and the volume of a sphere = $(4/3)\pi R^3$. How does this density compare with that of water and other materials discussed in this text? Can you see why astronomers were so surprised and puzzled when they first determined the orbit of the companion to Sirius?

How much would you weigh if you were suddenly transported to the white dwarf Sirius B? You may use your own weight (or if don't want

to own up to what it is, assume you weigh 70 kg or 150 lb). In this case, assume that the companion to Sirius has a mass equal to that of the Sun and a radius equal to that of Earth. Remember Newton's law of gravity:

$$F = G \frac{M_1 M_2}{r^2}$$

and that your weight is proportional to the force that you feel. What kind of star should you travel to if you want to *lose* weight (and not gain it)?

The star Betelgeuse has a temperature of 3400 K and a luminosity of 13,200 L_{Sun} . Calculate the radius of Betelgeuse relative to the Sun.

Using the information provided in [link], what is the average stellar density in our part of the Galaxy? Use only the true stars (types O–M) and assume a spherical distribution with radius of 26 light-years.

Confirm that the angular diameter of the Sun of $1/2^{\circ}$ corresponds to a linear diameter of 1.39 million km. Use the average distance of the Sun and Earth to derive the answer. (Hint: This can be solved using a trigonometric function.)

An eclipsing binary star system is observed with the following contact times for the main eclipse:

Table C

Contact	Time	Date
First contact	12:00 p.m.	March 12
Second contact	4:00 p.m.	March 13
Third contact	9:00 a.m.	March 18
Fourth contact	1:00 p.m.	March 19

The orbital velocity of the smaller star relative to the larger is 62,000 km/h. Determine the diameters for each star in the system.

If a 100 solar mass star were to have a luminosity of 10⁷ times the Sun's luminosity, how would such a star's density compare when it is on the main sequence as an O-type star, and when it is a cool supergiant (M-type)? Use values of temperature from Figure or Figure and the relationship between luminosity, radius, and temperature as given in Exercise.

If Betelgeuse had a mass that was 25 times that of the Sun, how would its average density compare to that of the Sun? Use the definition of $denstiy = \frac{mass}{volume}$, where the volume is that of a sphere.

Glossary

H-R diagram

(Hertzsprung-Russell diagram) a plot of luminosity against surface temperature (or spectral type) for a group of stars

main sequence

a sequence of stars on the Hertzsprung–Russell diagram, containing the majority of stars, that runs diagonally from the upper left to the lower right

white dwarf

a low-mass star that has exhausted most or all of its nuclear fuel and has collapsed to a very small size; such a star is near its final state of life

Chapter 19 Celestial Distances Section 19.1: Fundamental Units of Distance

Thinking Ahead

Globular Cluster M80.



Figure 1. This beautiful image shows a giant cluster of stars called Messier 80, located about 28,000 light-years from Earth. Such crowded groups, which astronomers call globular clusters, contain hundreds of thousands of stars, including some of the RR Lyrae variables discussed in this chapter. Especially obvious in this picture are the bright red giants, which are stars similar to the Sun in mass that are nearing the ends of their lives. (credit: modification of work by The Hubble Heritage Team (AURA/ STScI/ NASA)) How large is the universe? What is the most distant object we can see? These are among the most fundamental questions astronomers can ask. But just as babies must crawl before they can take their first halting steps, so too must we start with a more modest question: How far away are the stars? And even this question proves to be very hard to answer. After all, stars are mere points of light. Suppose you see a point of light in the darkness when you are driving on a country road late at night. How can you tell whether it is a nearby firefly, an oncoming motorcycle some distance away, or the porchlight of a house much farther down the road? It's not so easy, is it? Astronomers faced an even more difficult problem when they tried to estimate how far away the stars are.

In this chapter, we begin with the fundamental definitions of distances on Earth and then extend our reach outward to the stars. We will also examine the newest satellites that are surveying the night sky and discuss the special types of stars that can be used as trail markers to distant galaxies.

19.1 Fundamental Units of Distance



the other members of the solar system

The first measures of distances were based on human dimensions—the inch as the distance between knuckles on the finger, or the yard as the span from the extended index finger to the nose of the British king. Later, the requirements of commerce led to some standardization of such units, but each nation tended to set up its own definitions. It was not until the middle of the eighteenth century that any real efforts were made to establish a uniform, international set of standards.

The Metric System

One of the enduring legacies of the era of the French emperor Napoleon is the establishment of the *metric* system of units, officially adopted in France in 1799 and now used in most countries around the world. The fundamental metric unit of length is the *meter*, originally defined as one ten-millionth of the distance along Earth's surface from the equator to the pole. French astronomers of the seventeenth and eighteenth centuries were pioneers in determining the dimensions of Earth, so it was logical to use their information as the foundation of the new system.

Practical problems exist with a definition expressed in terms of the size of Earth, since anyone wishing to determine the distance from one place to another can hardly be expected to go out and re-measure the planet. Therefore, an intermediate standard meter consisting of a bar of platinum-iridium metal was set up in Paris. In 1889, by international agreement, this bar was defined to be exactly one meter in length, and precise copies of the original meter bar were made to serve as standards for other nations.

Other units of length are derived from the **meter**. Thus, 1 kilometer (km) equals 1000 meters, 1 centimeter (cm) equals 1/100 meter, and so on. Even the old British and American units, such as the inch and the mile, are now defined in terms of the metric system.

Modern Redefinitions of the Meter

In 1960, the official definition of the meter was changed again. As a result of improved technology for generating spectral lines of precisely known wavelengths (see the chapter on Radiation and Spectra), the meter was redefined to equal 1,650,763.73 wavelengths of a particular atomic transition in the element krypton-86. The advantage of this redefinition is that anyone with a suitably equipped laboratory can reproduce a standard meter, without reference to any particular metal bar.

In 1983, the meter was defined once more, this time in terms of the velocity of light. Light in a vacuum can travel a distance of one meter in 1/299,792,458.6 second. Today, therefore, light travel time provides our basic unit of length. Put another way, a distance of *one light-second* (the amount of space light covers in one second) is defined to be 299,792,458.6 meters. That's almost 300 million meters that light covers in just one second; light really is *very* fast! We could just as well use the light-second as the fundamental unit of length, but for practical reasons (and to respect tradition), we have defined the meter as a small fraction of the light-second.

Distance within the Solar System

The work of Copernicus and Kepler established the *relative* distances of the planets—that is, how far from the Sun one planet is compared to another (see Observing the Sky: The Birth of Astronomy and Orbits and Gravity). But their work could not establish the *absolute* distances (in light-seconds or meters or other standard units of length). This is like knowing the height of all the students in your class only as compared to the height of your astronomy instructor, but not in inches or centimeters. Somebody's height has to be measured directly.

Similarly, to establish absolute distances, astronomers had to measure one distance in the solar system directly. Generally, the closer to us the object is, the easier such a measurement would be. Estimates of the distance to Venus were made as Venus crossed the face of the Sun in 1761 and 1769, and an international campaign was organized to estimate the distance to the asteroid Eros in the early 1930s, when its orbit brought it close to Earth. More recently, Venus crossed (or *transited*) the surface of the Sun in 2004 and 2012, and allowed us to make a modern distance estimate, although, as we will see below, by then it wasn't needed (Figure 1).

If you would like more information on just how the motion of Venus across the Sun helped us pin down distances in the solar system, you can turn to a nice explanation by a NASA astronomer.
Venus Transits the Sun, 2012.



Figure 1. This striking "picture" of Venus crossing the face of the Sun (it's the black dot at about 2 o'clock) is more than just an impressive image. Taken with the Solar Dynamics Observatory spacecraft and special filters, it shows a modern transit of Venus. Such events allowed astronomers in the 1800s to estimate the distance to Venus. They measured the time it took Venus to cross the face of the Sun from different latitudes on Earth. The differences in times can be used to estimate the distance to the planet. Today, radar is used for much more precise distance estimates. (credit: modification of work by NASA/SDO, AIA)

The key to our modern determination of solar system dimensions is **radar**, a type of radio wave that can bounce off solid objects (Figure 2). As discussed in several earlier chapters, by timing how long a radar beam (traveling at the speed of light) takes to reach another world and return, we can measure the distance involved very accurately. In 1961, radar signals were bounced off Venus for the first time, providing a direct measurement of the distance from Earth to Venus in terms of light-seconds (from the roundtrip travel time of the radar signal).

Subsequently, radar has been used to determine the distances to Mercury, Mars, the satellites of Jupiter, the rings of Saturn, and several asteroids. Note, by the way, that it is not possible to use radar to measure the distance to the Sun directly because the Sun does not reflect radar very efficiently. But we can measure the distance to many other solar system objects and use Kepler's laws to give us the distance to the Sun.

Radar Telescope.



Figure 2.This dish-shaped antenna, part of the NASA Deep Space Network in California's Mojave Desert, is 70 meters wide.
Nicknamed the "Mars antenna," this radar telescope can send and receive radar waves, and thus measure the distances to planets, satellites, and asteroids. (credit: NASA/JPL-Caltech)
From the various (related) solar system distances, astronomers

528 | Chapter 19 Celestial Distances Section 19.1: Fundamental Units of Distance

selected the average distance from Earth to the Sun as our standard "measuring stick" within the solar system. When Earth and the Sun are closest, they are about 147.1 million kilometers apart; when Earth and the Sun are farthest, they are about 152.1 million kilometers apart. The average of these two distances is called the **astronomical unit** (AU). We then express all the other distances in the solar system in terms of the AU. Years of painstaking analyses of radar measurements have led to a determination of the length of the AU to a precision of about one part in a billion. The length of 1 AU can be expressed in light travel time as 499.004854 light-seconds, or about 8.3 light-minutes. If we use the definition of the meter given previously, this is equivalent to 1 AU = 149,597,870,700 meters.

These distances are, of course, given here to a much higher level of precision than is normally needed. In this text, we are usually content to express numbers to a couple of significant places and leave it at that. For our purposes, it will be sufficient to round off these numbers:

speed of light $c = 3 \times 10^8 m/s = 3 \times 10^5 km/s$ length of light second: $1s = 3 \times 10^8 m = 3 \times 10^5 km$ astronomical unit:

 $AU = 1.50 \times 10^{11} m = 1.50 \times 10^8 km = 500 light seconds$

We now know the absolute distance scale within our own solar system with fantastic accuracy. This is the first link in the chain of cosmic distances.

The distances between the celestial bodies in our solar system are sometimes difficult to grasp or put into perspective. This interactive website provides a "map" that shows the distances by using a scale at the bottom of the screen and allows you to scroll (using your arrow keys) through screens of "empty space" to get to the next planet–all while your current distance from the Sun is visible on the scale.

Key Concepts and Summary

Early measurements of length were based on human dimensions, but today, we use worldwide standards that specify lengths in units such as the meter. Distances within the solar system are now determined by timing how long it takes radar signals to travel from Earth to the surface of a planet or other body and then return.

Chapter 19 Section 19.2: Surveying the Stars

19.2 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 study 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 this is written in 2016, Voyager 1 is 134 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 1). You set up two observing stations some distance apart. That distance (line AB in Figure 1) 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**.

Triangulation.



Figure 1. 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 3). 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 2), 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.

Friedrich Wilhelm Bessel (1784–1846), Thomas J. Henderson (1798–1844), and Friedrich Struve (1793-1864).



(a)

(c)

Figure 2. (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 3 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 3). The reason for this definition is just that they prefer to deal with a baseline of 1 AU instead of 2 AU.

Parallax.



Figure 3. 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.

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 *parallax* of one second." 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.

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 \times 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 days/year×86,400 s/day =31,536,000 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:

> distance=velocity \times time =300,000 km/s \times 31,536,000 s = 9.46 \times 10¹² km

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.

Check Your Learning

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:

 $\begin{array}{l} 1 light-year=9.46\times10^{12} km\\ =9.46\times10^{12} km\times\frac{1EarthDiameter}{12,700 km}\\ =7.45\times10^8 EarthDiameters \end{array}$

That means that 1 light-year is about 745 million times the diameter of Earth.

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 4). 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.

Objects in Orion.



Figure 4. (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 highspeed 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 bluishwhite, 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.

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:

 $Circumference = 2\pi \times 93,000,000 miles$

Then, the following two ratios are equal:

$$\frac{0.5}{360} = \frac{diameterofsun}{2\pi \times 93,000,000}$$

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.

$$\label{eq:DiameteroftheSun} \begin{split} Diameterof theSun &= \frac{0.5}{360} \times 2\pi \times 93,000,000 miles = 811,577 miles \\ \text{This is very close to the true value of about 848,000 miles.} \end{split}$$

Check Your Learning

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: 206,265 au

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 5 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 expected to measure the position and distances to almost one billion stars with an accuracy of a few ten-millionths of an arcsecond. Gaia's distance limit will extend 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 will also be able to measure proper motions² 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.

H–R Diagram of Stars Measured by *Gaia* and Hipparcos.



Figure 5. 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)

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 groundbased 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*. The Gaia mission is scheduled to last for 5 years. 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 threedimensional 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.

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.*

Key Concepts and Summary

For stars that are relatively nearby, we can "triangulate" the distances from a baseline created by Earth's annual motion around the Sun. Half the shift in a nearby star's position relative to very distant background stars, as viewed from opposite sides of Earth's orbit, is called the parallax of that star and is a measure of its distance. The units used to measure stellar distance are the light-year, the distance light travels in 1 year, and the parsec (pc), the distance of a star with a parallax of 1 arcsecond (1 parsec = 3.26 light-years). The closest star, a red dwarf, is over 1 parsec away. The first successful measurements of stellar parallaxes were reported in 1838. Parallax measurements are a fundamental link in the chain of cosmic distances. The Hipparcos satellite has allowed us to measure accurate parallaxes for stars out to about 300 light-years, and the Gaia mission will result in parallaxes out to 30,000 light-years.

Footnotes

- 1 To 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.
- 2 Proper motion (as discussed in Analyzing Starlight, is the motion of a star across the sky (perpendicular to our line of sight.)

Glossary

parallax

an apparent displacement of a nearby star that results from the motion of Earth around the Sun

parsec

a unit of distance in astronomy, equal to 3.26 light-years; at a distance of 1 parsec, a star has a parallax of 1 arcsecond

Chapter 19 Section 19.3: Variable Stars: One Key to Cosmic Distances

19.3 Variable Stars: One Key to Cosmic Distances



Let's briefly review the key reasons that measuring distances to the stars is such a struggle. As discussed in The Brightness of Stars, our problem is that stars come in a bewildering variety of intrinsic luminosities. (If stars were light bulbs, we'd say they come in a wide range of wattages.) Suppose, instead, that all stars had the same "wattage" or luminosity. In that case, the more distant ones would always look dimmer, and we could tell how far away a star is simply by how dim it appeared. In the real universe, however, when we look at a star in our sky (with eye or telescope) and measure its apparent brightness, we cannot know whether it looks dim because it's a lowwattage bulb or because it is far away, or perhaps some of each.

Astronomers need to discover something else about the star that allows us to "read off" its intrinsic luminosity—in effect, to know what the star's true wattage is. With this information, we can then attribute how dim it looks from Earth to its distance. Recall that the apparent brightness of an object decreases with the square of the distance to that object. If two objects have the same luminosity but one is three times farther than the other, the more distant one will look nine times fainter. Therefore, if we know the luminosity of a star and its apparent brightness, we can calculate how far away it is. Astronomers have long searched for techniques that would somehow allow us to determine the luminosity of a star–and it is to these techniques that we turn next.

Variable Stars

The breakthrough in measuring distances to remote parts of our Galaxy, and to other galaxies as well, came from the study of variable stars. Most stars are constant in their luminosity, at least to within a percent or two. Like the Sun, they generate a steady flow of energy from their interiors. However, some stars are seen to vary in brightness and, for this reason, are called *variable stars*. Many such stars vary on a regular cycle, like the flashing bulbs that decorate stores and homes during the winter holidays.

Let's define some tools to help us keep track of how a star varies. A graph that shows how the brightness of a variable star changes with time is called a light curve (<u>Figure 1</u>). The *maximum* is the point of the light curve where the star has its greatest brightness;

the *minimum* is the point where it is faintest. If the light variations repeat themselves periodically, the interval between the two maxima is called the *period* of the star. (If this kind of graph looks familiar, it is because we introduced it in Diameters of Stars.)

Cepheid Light Curve.



Figure 1. This light curve shows how the brightness changes with time for a typical cepheid variable, with a period of about 6 days.

Pulsating Variables

There are two special types of variable stars for which—as we will see—measurements of the light curve give us accurate distances. These are called cepheid and RR Lyrae variables, both of which are pulsating variable stars. Such a star actually changes its diameter with time—periodically expanding and contracting, as your chest does when you breathe. We now understand that these stars are going through a brief unstable stage late in their lives.

The expansion and contraction of pulsating variables can be measured by using the Doppler effect. The lines in the spectrum shift toward the blue as the surface of the star moves toward us and then shift to the red as the surface shrinks back. As the star pulsates, it also changes its overall color, indicating that its temperature is also varying. And, most important for our purposes, the luminosity of the pulsating variable also changes in a regular way as it expands and contracts.

Cepheid Variables

Cepheids are large, yellow, pulsating stars named for the firstknown star of the group, Delta Cephei. This, by the way, is another example of how confusing naming conventions get in astronomy; here, a whole class of stars is named after the constellation in which the first one happened to be found. (We textbook authors can only apologize to our students for the whole mess!)

The variability of Delta Cephei was discovered in 1784 by the young English astronomer John Goodricke (see John Goodricke). The star rises rather rapidly to maximum light and then falls more slowly to minimum light, taking a total of 5.4 days for one cycle. The curve in <u>Figure 1</u> represents a simplified version of the light curve of Delta Cephei.

Several hundred cepheid variables are known in our Galaxy. Most cepheids have periods in the range of 3 to 50 days and luminosities that are about 1000 to 10,000 times greater than that of the Sun. Their variations in luminosity range from a few percent to a factor of 10.

Polaris, the North Star, is a cepheid variable that, for a long time, varied by one tenth of a magnitude, or by about 10% in visual luminosity, in a period of just under 4 days. Recent measurements indicate that the amount by which the brightness of Polaris changes is decreasing and that, sometime in the future, this star will no longer be a pulsating variable. This is just one more piece of evidence that stars really do evolve and change in fundamental ways as they age, and that being a cepheid variable represents a stage in the life of the star.

The Period-Luminosity Relation

The importance of cepheid variables lies in the fact that their periods and average luminosities turn out to be directly related. The longer the period (the longer the star takes to vary), the greater the luminosity. This period-luminosity relation was a remarkable discovery, one for which astronomers still (pardon the expression) thank their lucky stars. The period of such a star is easy to measure: a good telescope and a good clock are all you need. Once you have the period, the relationship (which can be put into precise mathematical terms) will give you the luminosity of the star.

Let's be clear on what that means. The relation allows you to essentially "read off" how bright the star really is (how much energy it puts out). Astronomers can then compare this intrinsic brightness with the apparent brightness of the star. As we saw, the difference between the two allows them to calculate the distance.

The relation between period and luminosity was discovered in 1908 by Henrietta Leavitt (Figure 2), a staff member at the Harvard College Observatory (and one of a number of women working for low wages assisting Edward Pickering, the observatory's director; see Annie Cannon: Classifier of the Stars). Leavitt discovered hundreds of variable stars in the Large Magellanic Cloud and Small Magellanic Cloud, two great star systems that are actually neighboring galaxies (although they were not known to be galaxies then). A small fraction of these variables were cepheids (Figure 3).

Henrietta Swan Leavitt (1868–1921).



Figure 2. Leavitt worked as an astronomer at the Harvard College Observatory. While studying photographs of the Magellanic Clouds, she found over 1700 variable stars, including 20 cepheids. Since all the cepheids in these systems were at roughly the same distance, she was able to compare their luminosities and periods of variation.

She thus discovered a fundamental relationship between these characteristics that led to a new and much better way of estimating

cosmic distances. (credit: modification of work by AIP) These systems presented a wonderful opportunity to study the behavior of variable stars independent of their distance. For all practical purposes, the Magellanic Clouds are so far away that astronomers can assume that all the stars in them are at roughly the same distance from us. (In the same way, all the suburbs of Los Angeles are roughly the same distance from New York City. Of course, if you are *in* Los Angeles, you will notice annoying distances between the suburbs, but compared to how far away New York City is, the differences seem small.) If all the variable stars in the Magellanic Clouds are at roughly the same distance, then any difference in their apparent brightnesses must be caused by differences in their intrinsic luminosities.

Large Magellanic Cloud.



Figure 3. The Large Magellanic Cloud (so named because Magellan's crew were the first Europeans to record it) is a small, irregularly shaped galaxy near our own Milky Way. It was in this galaxy that Henrietta Leavitt discovered the cepheid periodluminosity relation. (credit: ESO)

Leavitt found that the brighter-appearing cepheids always have the longer periods of light variation. Thus, she reasoned, the period must be related to the luminosity of the stars. When Leavitt did this work, the distance to the Magellanic Clouds was not known, so she was only able to show that luminosity was related to period. She could not determine exactly what the relationship is.

To define the period-luminosity relation with actual numbers (to *calibrate* it), astronomers first had to measure the actual distances to a few nearby cepheids in another way. (This was accomplished by finding cepheids associated in clusters with other stars whose distances could be estimated from their spectra, as discussed in the next section of this chapter.) But once the relation was thus defined, it could give us the distance to any cepheid, wherever it might be located (Figure 4).

How to Use a Cepheid to Measure Distance.



Figure 4. (a) Find a cepheid variable star and measure its period.(b) Use the period-luminosity relation to calculate the star's luminosity. (c) Measure the star's apparent brightness. (d) Compare the luminosity with the apparent brightness to calculate the distance.

Here at last was the technique astronomers had been searching for to break the confines of distance that parallax imposed on them. Cepheids can be observed and monitored, it turns out, in many parts of our own Galaxy and in other nearby galaxies as well. Astronomers, including Ejnar Hertzsprung and Harvard's Harlow Shapley, immediately saw the potential of the new technique; they and many others set to work exploring more distant reaches of space using cepheids as signposts. In the 1920s, Edwin Hubble made one of the most significant astronomical discoveries of all time using cepheids, when he observed them in nearby galaxies and discovered the expansion of the universe. As we will see, this work still continues, as the Hubble Space Telescope and other modern instruments try to identify and measure individual cepheids in galaxies farther and farther away. The most distant known variable stars are all cepheids, with some about 60 million light-years away.

JOHN GOODRICKE

The brief life of John Goodricke (<u>Figure 5</u>) is a testament to the human spirit under adversity. Born deaf and unable to speak, Goodricke nevertheless made a number of pioneering discoveries in astronomy through patient and careful observations of the heavens.

John Goodricke (1764–1786).



Figure 5. This portrait of Goodricke by artist J. Scouler hangs in the Royal Astronomical Society in London. There is some controversy about whether this is actually what Goodricke looked like or whether the painting was much retouched to please his family. (credit: James Scouler)
Born in Holland, where his father was on a diplomatic mission, Goodricke was sent back to England at age eight to study at a special school for the deaf. He did sufficiently well to enter Warrington Academy, a secondary school that offered no special assistance for students with handicaps. His mathematics teacher there inspired an interest in astronomy, and in 1781, at age 17, Goodricke began observing the sky at his family home in York, England. Within a year, he had discovered the brightness variations of the star Algol (discussed in The Stars: A Celestial Census) and suggested that an unseen companion star was causing the changes, a theory that waited over 100 years for proof. His paper on the subject was read before the Royal Society (the main British group of scientists) in 1783 and won him a medal from that distinguished group.

In the meantime, Goodricke had discovered two other stars that varied regularly, Beta Lyrae and Delta Cephei, both of which continued to interest astronomers for years to come. Goodricke shared his interest in observing with his older cousin, Edward Pigott, who went on to discover other variable stars during his much longer life. But Goodricke's time was quickly drawing to a close; at age 21, only 2 weeks after he was elected to the Royal Society, he caught a cold while making astronomical observations and never recovered.

Today, the University of York has a building named Goodricke Hall and a plaque that honors his contributions to science. Yet if you go to the churchyard cemetery where he is buried, an overgrown tombstone has only the initials "J. G." to show where he lies. Astronomer Zdenek Kopal, who looked carefully into Goodricke's life, speculated on why the marker is so modest: perhaps the rather staid Goodricke relatives were ashamed of having a "deaf-mute" in the family and could not sufficiently appreciate how much a man who could not hear could nevertheless see.

RR Lyrae Stars

A related group of stars, whose nature was understood somewhat later than that of the cepheids, are called RR Lyrae variables, named for the star RR Lyrae, the best-known member of the group. More common than the cepheids, but less luminous, thousands of these pulsating variables are known in our Galaxy. The periods of RR Lyrae stars are always less than 1 day, and their changes in brightness are typically less than about a factor of two.

Astronomers have observed that the RR Lyrae stars occurring in any particular cluster all have about the same apparent brightness. Since stars in a cluster are all at approximately the same distance, it follows that RR Lyrae variables must all have nearly the same intrinsic luminosity, which turns out to be about 50 L_{Sun} . In this sense, RR Lyrae stars are a little bit like standard light bulbs and can also be used to obtain distances, particularly within our Galaxy. <u>Figure 6</u> displays the ranges of periods and luminosities for both the cepheids and the RR Lyrae stars.



Period-Luminosity Relation for Cepheid Variables.

Figure 6. In this class of variable stars, the time the star takes to go through a cycle of luminosity changes is related to the average luminosity of the star. Also shown are the period and luminosity for RR Lyrae stars.

Key Concepts and Summary

Cepheids and RR Lyrae stars are two types of pulsating variable stars. Light curves of these stars show that their luminosities vary with a regularly repeating period. RR Lyrae stars can be used as standard bulbs, and cepheid variables obey a period-luminosity relation, so measuring their periods can tell us their luminosities. Then, we can calculate their distances by comparing their

Chapter 19 Section 19.3: Variable Stars: One Key to Cosmic Distances | 565

luminosities with their apparent brightnesses, and this can allow us to measure distances to these stars out to over 60 million lightyears.

Glossary

cepheid

a star that belongs to a class of yellow supergiant pulsating stars; these stars vary periodically in brightness, and the relationship between their periods and luminosities is useful in deriving distances to them

light curve

a graph that displays the time variation of the light from a variable or eclipsing binary star or, more generally, from any other object whose radiation output changes with time

period-luminosity relation

an empirical relation between the periods and luminosities of certain variable stars

pulsating variable star

a variable star that pulsates in size and luminosity

RR Lyrae

one of a class of giant pulsating stars with periods shorter than 1 day, useful for finding distances

Chapter 19 Section 19.4: The H-R Diagram and Cosmic Distances

19.4 The H–R Diagram and Cosmic Distances



Variable stars are not the only way that we can estimate the luminosity of stars. Another way involves the H–R diagram, which shows that the intrinsic brightness of a star can be estimated if we know its spectral type.

Distances from Spectral Types

As satisfying and productive as variable stars have been for distance measurement, these stars are rare and are not found near all the objects to which we wish to measure distances. Suppose, for example, we need the distance to a star that is not varying, or to a group of stars, none of which is a variable. In this case, it turns out the H–R diagram can come to our rescue.

If we can observe the spectrum of a star, we can estimate its distance from our understanding of the H–R diagram. As discussed in Analyzing Starlight, a detailed examination of a stellar spectrum allows astronomers to classify the star into one of the spectral typesindicating surface temperature. (The types are O, B, A, F, G, K, M, L, T, and Y; each of these can be divided into numbered subgroups.) In general, however, the spectral type alone is not enough to allow us to estimate luminosity. Look again at [link]. A G2 star could be a main-sequence star with a luminosity of $1 L_{Sun}$, or it could be a giant with a luminosity of $100 L_{Sun}$, or even a supergiant with a still higher luminosity.

We can learn more from a star's spectrum than just its temperature. Remember, for example, that we can detect pressure differences in stars from the details of the spectrum. This knowledge is very useful because giant stars are larger (and have lower pressures) than main-sequence stars, and supergiants are still larger than giants. If we look in detail at the spectrum of a star, we can determine whether it is a main-sequence star, a giant, or a supergiant.

Suppose, to start with the simplest example, that the spectrum, color, and other properties of a distant G2 star match those of the Sun exactly. It is then reasonable to conclude that this distant star is likely to be a main-sequence star just like the Sun and to have the same luminosity as the Sun. But if there are subtle differences between the solar spectrum and the spectrum of the distant star, then the distant star may be a giant or even a supergiant.

The most widely used system of star classification divides stars of a given spectral class into six categories called luminosity classes. These luminosity classes are denoted by Roman numbers as follows:

- Ia: Brightest supergiants
- Ib: Less luminous supergiants
- II: Bright giants
- III: Giants
- IV: Subgiants (intermediate between giants and main-sequence stars)
- V: Main-sequence stars

The full spectral specification of a star includes its luminosity class. For example, a main-sequence star with spectral class F3 is written as F3 V. The specification for an M2 giant is M2 III. Figure 1 illustrates the approximate position of stars of various luminosity classes on the H–R diagram. The dashed portions of the lines represent regions with very few or no stars.

Luminosity Classes.



Figure 1. Stars of the same temperature (or spectral class) can fall into different luminosity classes on the Hertzsprung-Russell diagram. By studying details of the spectrum for each star, astronomers can determine which luminosity class they fall in (whether they are main-sequence stars, giant stars, or supergiant stars).

With both its spectral and luminosity classes known, a star's position on the H–R diagram is uniquely determined. Since the diagram plots luminosity versus temperature, this means we can now read off the star's luminosity (once its spectrum has helped us place it on the diagram). As before, if we know how luminous the star really is and see how dim it looks, the difference allows us to calculate its distance. (For historical reasons, astronomers sometimes call this method of distance determination *spectroscopic parallax*, even though the method has nothing to do with parallax.)

The H–R diagram method allows astronomers to estimate distances to nearby stars, as well as some of the most distant stars in our Galaxy, but it is anchored by measurements of parallax. The

distances measured using parallax are the gold standard for distances: they rely on no assumptions, only geometry. Once astronomers take a spectrum of a nearby star for which we also know the parallax, we know the luminosity that corresponds to that spectral type. Nearby stars thus serve as benchmarks for more distant stars because we can assume that two stars with identical spectra have the same intrinsic luminosity.

A Few Words about the Real World

Introductory textbooks such as ours work hard to present the material in a straightforward and simplified way. In doing so, we sometimes do our students a disservice by making scientific techniques seem too clean and painless. In the real world, the techniques we have just described turn out to be messy and difficult, and often give astronomers headaches that last long into the day.

For example, the relationships we have described such as the period-luminosity relation for certain variable stars aren't exactly straight lines on a graph. The points representing many stars scatter widely when plotted, and thus, the distances derived from them also have a certain built-in scatter or uncertainty.

The distances we measure with the methods we have discussed are therefore only accurate to within a certain percentage of error-sometimes 10%, sometimes 25%, sometimes as much as 50% or more. A 25% error for a star estimated to be 10,000 light-years away means it could be anywhere from 7500 to 12,500 light-years away. This would be an unacceptable uncertainty if you were loading fuel into a spaceship for a trip to the star, but it is not a bad first figure to work with if you are an astronomer stuck on planet Earth.

Nor is the construction of H–R diagrams as easy as you might think at first. To make a good diagram, one needs to measure the

characteristics and distances of many stars, which can be a timeconsuming task. Since our own solar neighborhood is already well mapped, the stars astronomers most want to study to advance our knowledge are likely to be far away and faint. It may take hours of observing to obtain a single spectrum. Observers may have to spend many nights at the telescope (and many days back home working with their data) before they get their distance measurement. Fortunately, this is changing because surveys like Gaia will study billions of stars, producing public datasets that all astronomers can use.

Despite these difficulties, the tools we have been discussing allow us to measure a remarkable range of distances-parallaxes for the nearest stars, RR Lyrae variable stars; the H-R diagram for clusters of stars in our own and nearby galaxies; and cepheids out to distances of 60 million light-years. <u>Table</u> describes the distance limits and overlap of each method.

Each technique described in this chapter builds on at least one other method, forming what many call the *cosmic distance ladder*. Parallaxes are the foundation of all stellar distance estimates, spectroscopic methods use nearby stars to calibrate their H–R diagrams, and RR Lyrae and cepheid distance estimates are grounded in H–R diagram distance estimates (and even in a parallax measurement to a nearby cepheid, Delta Cephei).

This chain of methods allows astronomers to push the limits when looking for even more distant stars. Recent work, for example, has used RR Lyrae stars to identify dim companion galaxies to our own Milky Way out at distances of 300,000 light-years. The H–R diagram method was recently used to identify the two most distant stars in the Galaxy: red giant stars way out in the halo of the Milky Way with distances of almost 1 million light-years.

We can combine the distances we find for stars with measurements of their composition, luminosity, and temperature-made with the techniques described in Analyzing Starlight and <u>The Stars: A Celestial Census</u>. Together, these make up the arsenal of information we need to trace the evolution of stars

from birth to death, the subject to which we turn in the chapters that follow.

Method	Distance Range
Trigonometric parallax	4–30,000 light-years when the Gaia mission is complete
RR Lyrae stars	Out to 300,000 light-years
H–R diagram and spectroscopic distances	Out to 1,200,000 light-years
Cepheid stars	Out to 60,000,000 light-years

Distance Range of Celestial Measurement Methods

Key Concepts and Summary

Stars with identical temperatures but different pressures (and diameters) have somewhat different spectra. Spectral classification can therefore be used to estimate the luminosity class of a star as well as its temperature. As a result, a spectrum can allow us to pinpoint where the star is located on an H–R diagram and establish its luminosity. This, with the star's apparent brightness, again yields its distance. The various distance methods can be used to check one against another and thus make a kind of distance ladder which allows us to find even larger distances.

For Further Exploration

Articles

Adams, A. "The Triumph of Hipparcos." Astronomy (December 1997): 60. Brief introduction.

Dambeck, T. "Gaia's Mission to the Milky Way." Sky & Telescope (March 2008): 36–39. An introduction to the mission to measure distances and positions of stars with unprecedented accuracy.

Hirshfeld, A. "The Absolute Magnitude of Stars." Sky & Telescope (September 1994): 35. Good review of how we measure luminosity, with charts.

Hirshfeld, A. "The Race to Measure the Cosmos." Sky & *Telescope* (November 2001): 38. On parallax.

Trefil, J. Puzzling Out Parallax." Astronomy (September 1998): 46. On the concept and history of parallax.

Turon, C. "Measuring the Universe." Sky & Telescope (July 1997): 28. On the Hipparcos mission and its results.

Zimmerman, R. "Polaris: The Code-Blue Star." Astronomy (March 1995): 45. On the famous cepheid variable and how it is changing.

Websites

ABCs of Distance: http://www.astro.ucla.edu/~wright/ distance.htm. Astronomer Ned Wright (UCLA) gives a concise primer on many different methods of obtaining distances. This site is at a higher level than our textbook, but is an excellent review for those with some background in astronomy.

American Association of Variable Star Observers (AAVSO): https://www.aavso.org/. This organization of amateur astronomers helps to keep track of variable stars; its site has some background material, observing instructions, and links.

Friedrich Wilhelm Bessel: http://messier.seds.org/xtra/Bios/ bessel.html. A brief site about the first person to detect stellar parallax, with references and links.

Gaia: http://sci.esa.int/gaia/. News from the Gaia mission, including images and a blog of the latest findings.

Hipparchos: http://sci.esa.int/hipparcos/. Background, results, catalogs of data, and educational resources from the Hipparchos

574 | Chapter 19 Section 19.4: The H-R Diagram and Cosmic Distances

mission to observe parallaxes from space. Some sections are technical, but others are accessible to students.

John Goodricke: The Deaf Astronomer: http://www.bbc.com/ news/magazine-20725639. A biographical article from the BBC.

Women in Astronomy: http://www.astrosociety.org/education/ astronomy-resource-guides/women-in-astronomy-anintroductory-resource-guide/. More about Henrietta Leavitt's and other women's contributions to astronomy and the obstacles they faced.

Videos

Gaia's Mission: Solving the Celestial Puzzle: https://www.youtube.com/watch?v=oGri4YNggoc.

Describes the Gaia mission and what scientists hope to learn, from Cambridge University (19:58).

Hipparcos: Route Map to the Stars: http://www.esa.int/ spaceinvideos/Videos/1997/05/

Hipparcos_Route_Maps_to_the_Stars_May_97. This ESA video describes the mission to measure parallax and its results (14:32)

How Big Is the Universe: https://www.youtube.com/ watch?v=K_xZuopg4Sk. Astronomer Pete Edwards from the British Institute of Physics discusses the size of the universe and gives a step-by-step introduction to the concepts of distances (6:22)

Search for Miss Leavitt: http://perimeterinstitute.ca/videos/ search-miss-leavitt. Video of talk by George Johnson on his search for Miss Leavitt (55:09).

Women in Astronomy: http://www.youtube.com/ watch?v=5vMR7su4fi8. Emily Rice (CUNY) gives a talk on the contributions of women to astronomy, with many historical and contemporary examples, and an analysis of modern trends (52:54).

Collaborative Group Activities

- A. In this chapter, we explain the various measurements that have been used to establish the size of a standard meter. Your group should discuss why we have changed the definitions of our standard unit of measurement in science from time to time. What factors in our modern society contribute to the growth of technology? Does technology "drive" science, or does science "drive" technology? Or do you think the two are so intertwined that it's impossible to say which is the driver?
- B. Cepheids are scattered throughout our own Milky Way Galaxy, but the period-luminosity relation was discovered from observations of the Magellanic Clouds, a satellite galaxy now known to be about 160,000 light-years away. What reasons can you give to explain why the relation was not discovered from observations of cepheids in our own Galaxy? Would your answer change if there were a small cluster in our own Galaxy that contained 20 cepheids? Why or why not?
- C. You want to write a proposal to use the Hubble Space Telescope to look for the brightest cepheids in galaxy M100 and estimate their luminosities. What observations would you need to make? Make a list of all the reasons such observations are harder than it first might appear.
- D. Why does your group think so many different ways of naming stars developed through history? (Think back to the days before everyone connected online.) Are there other fields where things are named confusingly and arbitrarily? How do stars differ from other phenomena that science and other professions tend to catalog?
- E. Although cepheids and RR Lyrae variable stars tend to change their brightness pretty regularly (while they are in that stage of their lives), some variable stars are unpredictable or change their their behavior even during the course of a single human lifetime. Amateur astronomers all over the world follow such

variable stars patiently and persistently, sending their nightly observations to huge databases that are being kept on the behavior of many thousands of stars. None of the hobbyists who do this get paid for making such painstaking observations. Have your group discuss why they do it. Would you ever consider a hobby that involves so much work, long into the night, often on work nights? If observing variable stars doesn't pique your interest, is there something you think you could do as a volunteer after college that does excite you? Why?

- F. In [link], the highest concentration of stars occurs in the middle of the main sequence. Can your group give reasons why this might be so? Why are there fewer very hot stars and fewer very cool stars on this diagram?
- G. In this chapter, we discuss two astronomers who were differently abled than their colleagues. John Goodricke could neither hear nor speak, and Henrietta Leavitt struggled with hearing impairment for all of her adult life. Yet they each made fundamental contributions to our understanding of the universe. Does your group know people who are handling a disability? What obstacles would people with different disabilities face in trying to do astronomy and what could be done to ease their way? For a set of resources in this area, see http://astronomerswithoutborders.org/gam2013/programs/ 1319-people-with-disabilities-astronomy-resources.html.

Review Questions

Explain how parallax measurements can be used to determine distances to stars. Why can we not make accurate measurements of parallax beyond a certain distance?

Suppose you have discovered a new cepheid variable star. What steps would you take to determine its distance?

Explain how you would use the spectrum of a star to estimate its distance.

Which method would you use to obtain the distance to each of the following?

- A. An asteroid crossing Earth's orbit
- B. A star astronomers believe to be no more than 50 light-years from the Sun
- C. A tight group of stars in the Milky Way Galaxy that includes a significant number of variable stars
- D. A star that is not variable but for which you can obtain a clearly defined spectrum

What are the luminosity class and spectral type of a star with an effective temperature of 5000 K and a luminosity of 100 L_{Sun} ?

Thought Questions

The meter was redefined as a reference to Earth, then to krypton, and finally to the speed of light. Why do you think the reference point for a meter continued to change?

While a meter is the fundamental unit of length, most distances traveled by humans are measured in miles or kilometers. Why do you think this is?

Most distances in the Galaxy are measured in light-years instead of meters. Why do you think this is the case?

The AU is defined as the *average* distance between Earth and the Sun, not the distance between Earth and the Sun. Why does this need to be the case?

What would be the advantage of making parallax measurements from Pluto rather than from Earth? Would there be a disadvantage?

Parallaxes are measured in fractions of an arcsecond. One arcsecond equals 1/60 arcmin; an arcminute is, in turn, 1/60th of a degree (°). To get some idea of how big 1° is, go outside at night and find the Big Dipper. The two pointer stars at the ends of the bowl are 5.5° apart. The two stars across the top of the bowl are 10° apart. (Ten degrees is also about the width of your fist when held at arm's length and projected against the sky.) Mizar, the second star from the end of the Big Dipper's handle, appears double. The fainter star, Alcor, is about 12 arcmin from Mizar. For comparison, the diameter of the full moon is about 30 arcmin. The belt of Orion is about 3° long. Keeping all this in mind, why did it take until 1838 to make parallax measurements for even the nearest stars?

For centuries, astronomers wondered whether comets were true celestial objects, like the planets and stars, or a phenomenon that occurred in the atmosphere of Earth. Describe an experiment to determine which of these two possibilities is correct.

The Sun is much closer to Earth than are the nearest stars, yet it is not possible to measure accurately the diurnal parallax of the Sun relative to the stars by measuring its position relative to background objects in the sky directly. Explain why.

Parallaxes of stars are sometimes measured relative to the positions of galaxies or distant objects called quasars. Why is this a good technique?

Estimating the luminosity class of an M star is much more important than measuring it for an O star if you are determining the distance to that star. Why is that the case?

[link] is the light curve for the prototype cepheid variable Delta Cephei. How does the luminosity of this star compare with that of the Sun? Which of the following can you determine about a star without knowing its distance, and which can you not determine: radial velocity, temperature, apparent brightness, or luminosity? Explain.

A G2 star has a luminosity 100 times that of the Sun. What kind of star is it? How does its radius compare with that of the Sun?

A star has a temperature of 10,000 K and a luminosity of $10^{-2}\,L_{Sun}.$ What kind of star is it?

What is the advantage of measuring a parallax distance to a star as compared to our other distance measuring methods?

What is the disadvantage of the parallax method, especially for studying distant parts of the Galaxy?

Luhman 16 and WISE 0720 are brown dwarfs, also known as failed stars, and are some of the new closest neighbors to Earth, but were only discovered in the last decade. Why do you think they took so long to be discovered?

Most stars close to the Sun are red dwarfs. What does this tell us about the average star formation event in our Galaxy?

Why would it be easier to measure the characteristics of intrinsically less luminous cepheids than more luminous ones?

When Henrietta Leavitt discovered the period-luminosity relationship, she used cepheid stars that were all located in the Small Magellanic Cloud. Why did she need to use stars in another galaxy and not cepheids located in the Milky Way?

Figuring for Yourself

A radar astronomer who is new at the job claims she beamed radio

waves to Jupiter and received an echo exactly 48 min later. Should you believe her? Why or why not?

The New Horizons probe flew past Pluto in July 2015. At the time, Pluto was about 32 AU from Earth. How long did it take for communication from the probe to reach Earth, given that the speed of light in km/hr is 1.08×10^9 ?

Estimate the maximum and minimum time it takes a radar signal to make the round trip between Earth and Venus, which has a semimajor axis of 0.72 AU.

The Apollo program (not the lunar missions with astronauts) being conducted at the Apache Point Observatory uses a 3.5-m telescope to direct lasers at retro-reflectors left on the Moon by the Apollo astronauts. If the Moon is 384,472 km away, approximately how long do the operators need to wait to see the laser light return to Earth?

In 1974, the Arecibo Radio telescope in Puerto Rico was used to transmit a signal to M13, a star cluster about 25,000 light-years away. How long will it take the message to reach M13, and how far has the message travelled so far (in light-years)?

Demonstrate that 1 pc equals 3.09×10^{13} km and that it also equals 3.26 light-years. Show your calculations.

The best parallaxes obtained with Hipparcos have an accuracy of 0.001 arcsec. If you want to measure the distance to a star with an accuracy of 10%, its parallax must be 10 times larger than the typical error. How far away can you obtain a distance that is accurate to 10% with Hipparcos data? The disk of our Galaxy is 100,000 light-years in diameter. What fraction of the diameter of the Galaxy's disk is the distance for which we can measure accurate parallaxes?

Astronomers are always making comparisons between measurements in astronomy and something that might be more familiar. For example, the Hipparcos web pages tell us that the measurement accuracy of 0.001 arcsec is equivalent to the angle made by a golf ball viewed from across the Atlantic Ocean, or to the angle made by the height of a person on the Moon as viewed from Earth, or to the length of growth of a human hair in 10 sec as seen from 10 meters away. Use the ideas in [link] to verify one of the first two comparisons.

Gaia will have greatly improved precision over the measurements of Hipparcos. The average uncertainty for most *Gaia*parallaxes will be about 50 microarcsec, or 0.00005 arcsec. How many times better than Hipparcos (see Exercise) is this precision?

Using the same techniques as used in Exercise, how far away can *Gaia* be used to measure distances with an uncertainty of 10%? What fraction of the Galactic disk does this correspond to?

The human eye is capable of an angular resolution of about one arcminute, and the average distance between eyes is approximately 2 in. If you blinked and saw something move about one arcmin across, how far away from you is it? (Hint: You can use the setup in [link] as a guide.)

How much better is the resolution of the *Gaia* spacecraft compared to the human eye (which can resolve about 1 arcmin)?

The most recently discovered system close to Earth is a pair of brown dwarfs known as Luhman 16. It has a distance of 6.5 lightyears. How many parsecs is this?

What would the parallax of Luhman 16 (see Exercise) be as measured from Earth?

The New Horizons probe that passed by Pluto during July 2015 is one of the fastest spacecraft ever assembled. It was moving at about 14 km/s when it went by Pluto. If it maintained this speed, how long would it take New Horizons to reach the nearest star, Proxima Centauri, which is about 4.3 light-years away? (Note: It isn't headed in that direction, but you can pretend that it is.)

What physical properties are different for an M giant with a luminosity of 1000 L_{Sun} and an M dwarf with a luminosity of 0.5 L_{Sun} ? What physical properties are the same?

Glossary

luminosity class

a classification of a star according to its luminosity within a given spectral class; our Sun, a G2V star, has luminosity class V, for example

Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System

21.4 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

584 | Chapter 21 The Birth of Stars and the Discovery of Planets outside

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 mainsequence 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 onethousandth 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 10 seconds of arc, and that of the Sun's orbit is 0.010 seconds of arc. (Remember, 1 second of arc is 1/3600 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 0.010 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 1).

Doppler Method of Detecting Planets.



Figure 1. 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 2) 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.)

Planet Discoverers.



Figure 2. In 1995, Didier Queloz and Michel Mayor of the Geneva Observatory were the first to discover a planet around a regular star (51 Pegasi). 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 3). 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 Jupiter¹, making it clearly a jovian and not a terrestrial-type planet.

Hot Jupiter.



Figure 3. 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. Hundreds of 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 Jupiterstake 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.

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

Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System | 591 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 edgeon, 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 <u>4</u> 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 4. 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

592 | Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System

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.

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:

$$\frac{\pi R^2 Planet}{\pi R^2 Star} = \frac{R^2 Planet}{R^2 Star} = \left(\frac{^R Planet}{^R Star}\right)^2$$

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 $\left(\frac{^{R}Planet}{^{R}Star}\right)^{2} = \left(\frac{71,400km}{695,700km}\right)^{2} = 0.01$ or 1%, which can easily be detected with the instruments on board the Kepler spacecraft.

Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System | 593



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.

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 5). 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

Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System | 595 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. A new NASA mission called TESS (Transiting Exoplanet Survey Satellite) will carry out a survey all over the sky of the nearer (and therefore brighter) stars, starting in 2018.

Kepler's Field of View.



Figure 5. 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 (<u>Figure 6</u>). 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.



Exoplanets around HR 8799.

Figure 6. 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 hydrogenrich 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

Chapter 21 The Birth of Stars and the Discovery of Planets outside the Solar System Section 21.4: Planets Beyond the Solar System | 599 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.

Key Concepts and Summary

Several observational techniques have successfully detected planets orbiting other stars. These techniques fall into two general categories-direct and indirect detection. The Doppler and transit techniques are our most powerful indirect tools for finding exoplanets. Some planets are also being found by direct imaging.

Footnotes

 1 The 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.

Glossary

exoplanet

a planet orbiting a star other than our Sun

transit

when one astronomical object moves in front of another

Chapter 21 Section 21.5: Exoplanets Everywhere: What We Are Learning

21.5 Exoplanets Everywhere: What We Are Learning



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 1, 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.

Exoplanet Discoveries through 2015.



Figure 1. 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 and red dots show planets discovered by transits, and the blue dots are the discoveries by the radial velocity (Doppler) technique. (credit: modification of work by NASA/Kepler mission)

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.

604 | Chapter 21 Section 21.5: Exoplanets Everywhere: What We Are Learning

Figure 2 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.



Kepler Discoveries.

Figure 2. 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 2. 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).

Size Distribution of Planets for Stars Similar to the Sun.



Figure 3. 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 onequarter 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).

<u>Figure 4</u> 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.

Exoplanets with Known Densities.



Figure 4. 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 4, 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 4 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 4, 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 4 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 (about 2600 as of 2016). 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 2018, astronomers gathered data on nearly 3000 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 5). 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).

Exoplanet System Kepler-62, with the Solar System Shown to the Same Scale.



Figure 5. 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.

Key Concepts and Summary

Although the Kepler mission is finding thousands of new exoplanets, these are limited to orbital periods of less than 400 days and sizes larger than Mars. Still, we can use the Kepler discoveries to extrapolate the distribution of planets in our Galaxy. The data so far imply that planets like Earth are the most common type of planet, and that there may be 100 billion Earth-size planets around Sun-like stars in the Galaxy. About 2600 planetary systems have been discovered around other stars. In many of them, planets are arranged differently than in our solar system.

Glossary

super-Earth

a planet larger than Earth, generally between 1.4 and 2.8 times the size of our planet

mini-Neptune

a planet that is intermediate between the largest terrestrial planet in our solar system (Earth) and the smallest jovian planet (Neptune); generally, mini-Neptunes have sizes between 2.8 and 4 times Earth's size

For Further Exploration

Articles

Star Formation

Blaes, O. "A Universe of Disks." *Scientific American* (October 2004): 48. On accretion disks and jets around young stars and black holes.

Croswell, K. "The Dust Belt Next Door [Tau Ceti]." Scientific American (January 2015): 24. Short intro to recent observations of planets and a wide dust belt.

Frank, A. "Starmaker: The New Story of Stellar Birth." Astronomy (July 1996): 52.

Jayawardhana, R. "Spying on Stellar Nurseries." Astronomy (November 1998): 62. On protoplanetary disks.

O'Dell, C. R. "Exploring the Orion Nebula." Sky & Telescope (December 1994): 20. Good review with Hubble results.

Ray, T. "Fountains of Youth: Early Days in the Life of a Star." *Scientific American* (August 2000): 42. On outflows from young stars.

Young, E. "Cloudy with a Chance of Stars." *Scientific American* (February 2010): 34. On how clouds of interstellar matter turn into star systems.

Young, Monica "Making Massive Stars." Sky & Telescope (October 2015): 24. Models and observations on how the most massive stars form.

Exoplanets

Billings, L. "In Search of Alien Jupiters." *Scientific American* (August 2015): 40–47. The race to image jovian planets with current instruments and why a direct image of a terrestrial planet is still in the future.

Heller, R. "Better Than Earth." Scientific American (January 2015): 32–39. What kinds of planets may be habitable; super-Earths and jovian planet moons should also be considered.

Laughlin, G. "How Worlds Get Out of Whack." Sky & Telescope (May 2013): 26. On how planets can migrate from the places they form in a star system.

Marcy, G. "The New Search for Distant Planets." Astronomy (October 2006): 30. Fine brief overview. (The same issue has a dramatic fold-out visual atlas of extrasolar planets, from that era.)

Redd, N. "Why Haven't We Found Another Earth?" Astronomy (February 2016): 25. Looking for terrestrial planets in the habitable zone with evidence of life.

Seager, S. "Exoplanets Everywhere." Sky & Telescope (August 2013): 18. An excellent discussion of some of the frequently asked questions about the nature and arrangement of planets out there.

Seager, S. "The Hunt for Super-Earths." Sky & Telescope (October

2010): 30. The search for planets that are up to 10 times the mass of Earth and what they can teach us.

Villard, R. "Hunting for Earthlike Planets." Astronomy (April 2011): 28. How we expect to find and characterize super-Earth (planets somewhat bigger than ours) using new instruments and techniques that could show us what their atmospheres are made of.

Websites

Exoplanet Exploration: http://planetquest.jpl.nasa.gov/. PlanetQuest (from the Navigator Program at the Jet Propulsion Lab) is probably the best site for students and beginners, with introductory materials and nice illustrations; it focuses mostly on NASA work and missions.

Exoplanets: http://www.planetary.org/exoplanets/. Planetary Society's exoplanets pages with a dynamic catalog of planets found and good explanations.

Exoplanets: The Search for Planets beyond Our Solar System: http://www.iop.org/publications/iop/2010/

page_42551.html. From the British Institute of Physics in 2010.

Extrasolar Planets Encyclopedia: http://exoplanet.eu/. Maintained by Jean Schneider of the Paris Observatory, has the largest catalog of planet discoveries and useful background material (some of it more technical).

Formation of Stars: https://www.spacetelescope.org/science/ formation_of_stars/. Star Formation page from the Hubble Space Telescope, with links to images and information.

Kepler Mission: http://kepler.nasa.gov/. The public website for the remarkable telescope in space that is searching planets using the transit technique and is our best hope for finding earthlike planets.

Proxima Centauri Planet Discovery: http://www.eso.org/public/ news/eso1629/. Exoplanet: http://itunes.apple.com/us/app/exoplanet/ id327702034?mt=8. Allows you to browse through a regularly updated visual catalog of exoplanets that have been found so far.

Journey to the Exoplanets: http://itunes.apple.com/us/app/ journey-to-the-exoplanets/id463532472?mt=8. Produced by the staff of *Scientific American*, with input from scientists and space artists; gives background information and visual tours of the nearer star systems with planets.

Videos

A Star Is Born: http://www.discovery.com/tv-shows/othershows/videos/how-the-universe-works-a-star-is-born/. Discovery Channel video with astronomer Michelle Thaller (2:25).

Are We Alone: An Evening Dialogue with the Kepler Mission Leaders: http://www.youtube.com/watch?v=O7ItAXfl0Lw. A nontechnical panel discussion on Kepler results and ideas about planet formation with Bill Borucki, Natalie Batalha, and Gibor Basri (moderated by Andrew Fraknoi) at the University of California, Berkeley (2:07:01).

Finding the Next Earth: The Latest Results from Kepler: https://www.youtube.com/watch?v=ZbijeR_AALo. Natalie Batalha (San Jose State University & NASA Ames) public talk in the Silicon Valley Astronomy Lecture Series (1:28:38).

From Hot Jupiters to Habitable Worlds: https://vimeo.com/ 37696087 (Part 1) and https://vimeo.com/37700700 (Part 2). Debra Fischer (Yale University) public talk in Hawaii sponsored by the Keck Observatory (15:20 Part 1, 21:32 Part 2).

Search for Habitable Exoplanets: http://www.youtube.com/ watch?v=RLWb_T9yaDU. Sara Seeger (MIT) public talk at the SETI Institute, with Kepler results (1:10:35). Strange Planetary Vistas: http://www.youtube.com/ watch?v=_8ww9eLRSCg. Josh Carter (CfA) public talk at Harvard's Center for Astrophysics with a friendly introduction to exoplanets for non-specialists (46:35).

Collaborative Group Activities

- A. Your group is a subcommittee of scientists examining whether any of the "hot Jupiters" (giant planets closer to their stars than Mercury is to the Sun) could have life on or near them. Can you come up with places on, in, or near such planets where life could develop or where some forms of life might survive?
- B. A wealthy couple (who are alumni of your college or university and love babies) leaves the astronomy program several million dollars in their will, to spend in the best way possible to search for "infant stars in our section of the Galaxy." Your group has been assigned the task of advising the dean on how best to spend the money. What kind of instruments and search programs would you recommend, and why?
- C. Some people consider the discovery of any planets (even hot Jupiters) around other stars one of the most important events in the history of astronomical research. Some astronomers have been surprised that the public is not more excited about the planet discoveries. One reason that has been suggested for this lack of public surprise and excitement is that science fiction stories have long prepared us for there being planets around other stars. (The Starship Enterprise on the 1960s Star Trek TV series found some in just about every weekly episode.) What does your group think? Did you know about the discovery of planets around other stars before taking this course? Do you consider it exciting? Were you surprised to hear about it? Are science fiction movies and books good or

bad tools for astronomy education in general, do you think?

- D. What if future space instruments reveal an earthlike exoplanet with significant amounts of oxygen and methane in its atmosphere? Suppose the planet and its star are 50 light-years away. What does your group suggest astronomers do next? How much effort and money would you recommend be put into finding out more about this planet and why?
- E. Discuss with your group the following question: which is easier to find orbiting a star with instruments we have today: a jovian planet or a proto-planetary disk? Make a list of arguments for each side of this question.
- F. (This activity should be done when your group has access to the internet.) Go to the page which indexes all the publicly released Hubble Space Telescope images by subject: http://hubblesite.org/newscenter/archive/browse/image/. Under "Star," go to "Protoplanetary Disk" and find a system—not mentioned in this chapter—that your group likes, and prepare a short report to the class about why you find it interesting. Then, under "Nebula," go to "Emission" and find a region of star formation not mentioned in this chapter, and prepare a short report to the class about what you find interesting about it.
- G. There is a "citizen science" website called Planet Hunters (http://www.planethunters.org/) where you can participate in identifying exoplanets from the data that Kepler provided. Your group should access the site, work together to use it, and classify two light curves. Report back to the class on what you have done.
- H. Yuri Milner, a Russian-American billionaire, recently pledged \$100 million to develop the technology to send many miniaturized probes to a star in the Alpha Centauri triple star system (which includes Proxima Centauri, the nearest star to us, now known to have at least one planet.) Each tiny probe will be propelled by powerful lasers at 20% the speed of light, in the hope that one or more might arrive safely and be able to

send back information about what it's like there. Your group should search online for more information about this project (called "Breakthrough: Starshot") and discuss your reactions to this project. Give specific reasons for your arguments.

Review Questions

Give several reasons the Orion molecular cloud is such a useful "laboratory" for studying the stages of star formation.

Why is star formation more likely to occur in cold molecular clouds than in regions where the temperature of the interstellar medium is several hundred thousand degrees?

Why have we learned a lot about star formation since the invention of detectors sensitive to infrared radiation?

Describe what happens when a star forms. Begin with a dense core of material in a molecular cloud and trace the evolution up to the time the newly formed star reaches the main sequence.

Describe how the T Tauri star stage in the life of a low-mass star can lead to the formation of a Herbig-Haro (H-H) object.

Look at the four stages shown in [link]. In which stage(s) can we see the star in visible light? In infrared radiation?

The evolutionary track for a star of 1 solar mass remains nearly vertical in the H–R diagram for a while (see [link]). How is its luminosity changing during this time? Its temperature? Its radius?

Two protostars, one 10 times the mass of the Sun and one half the mass of the Sun are born at the same time in a molecular cloud. Which one will be first to reach the main sequence stage, where it is stable and getting energy from fusion?

Compare the scale (size) of a typical dusty disk around a forming star with the scale of our solar system.

Why is it so hard to see planets around other stars and so easy to see them around our own?

Why did it take astronomers until 1995 to discover the first exoplanet orbiting another star like the Sun?

Which types of planets are most easily detected by Doppler measurements? By transits?

List three ways in which the exoplanets we have detected have been found to be different from planets in our solar system.

List any similarities between discovered exoplanets and planets in our solar system.

What revisions to the theory of planet formation have astronomers had to make as a result of the discovery of exoplanets?

Why are young Jupiters easier to see with direct imaging than old Jupiters?

Thought Questions

A friend of yours who did not do well in her astronomy class tells you that she believes all stars are old and none could possibly be born today. What arguments would you use to persuade her that stars are being born somewhere in the Galaxy during your lifetime?

Observations suggest that it takes more than 3 million years for the dust to begin clearing out of the inner regions of the disks surrounding protostars. Suppose this is the minimum time required to form a planet. Would you expect to find a planet around a $10-M_{Sun}$ star? (Refer to [link].)

Suppose you wanted to observe a planet around another star with direct imaging. Would you try to observe in visible light or in the infrared? Why? Would the planet be easier to see if it were at 1 AU or 5 AU from its star?

Why were giant planets close to their stars the first ones to be discovered? Why has the same technique not been used yet to discover giant planets at the distance of Saturn?

Exoplanets in eccentric orbits experience large temperature swings during their orbits. Suppose you had to plan for a mission to such a planet. Based on Kepler's second law, does the planet spend more time closer or farther from the star? Explain.

Figuring for Yourself

When astronomers found the first giant planets with orbits of only a few days, they did not know whether those planets were gaseous and liquid like Jupiter or rocky like Mercury. The observations of HD 209458 settled this question because observations of the transit of the star by this planet made it possible to determine the radius of the planet. Use the data given in the text to estimate the density of this planet, and then use that information to explain why it must be a gas giant.

An exoplanetary system has two known planets. Planet X orbits in 290 days and Planet Y orbits in 145 days. Which planet is closest to its host star? If the star has the same mass as the Sun, what is the semi-major axis of the orbits for Planets X and Y?

Kepler's third law says that the orbital period (in years) is proportional to the square root of the cube of the mean distance (in

622 | Chapter 21 Section 21.5: Exoplanets Everywhere: What We Are Learning

AU) from the Sun ($P \propto a^{1.5}$). For mean distances from 0.1 to 32 AU, calculate and plot a curve showing the expected Keplerian period. For each planet in our solar system, look up the mean distance from the Sun in AU and the orbital period in years and overplot these data on the theoretical Keplerian curve.

Calculate the transit depth for an M dwarf star that is 0.3 times the radius of the Sun with a gas giant planet the size of Jupiter.

If a transit depth of 0.00001 can be detected with the Kepler spacecraft, what is the smallest planet that could be detected around a 0.3 $R_{\rm sun}$ M dwarf star?

What fraction of gas giant planets seems to have inflated radii?

Chapter 23 The Death of Stars Section 23.1: The Death of Low-Mass Stars

Thinking Ahead

Stellar Life Cycle.



Figure 1. This remarkable picture of NGC 3603, a nebula in the Milky Way Galaxy, was taken with the Hubble Space Telescope. This image illustrates the life cycle of stars. In the bottom half of the image, we see clouds of dust and gas, where it is likely that star formation will take place in the near future. Near the center, there is a cluster of massive, hot young stars that are only a few million years old. Above and to the right of the cluster, there is an isolated star surrounded by a ring of gas. Perpendicular to the ring and on either side of it, there are two bluish blobs of gas. The ring and the blobs were ejected by the star, which is nearing the end of its life. (credit: modification of work by NASA, Wolfgang Brandner (JPL/

IPAC), Eva K. Grebel (University of Washington), You-Hua Chu (University of Illinois Urbana-Champaign))

Do stars die with a bang or a whimper? In the preceding two chapters, we followed the life story of stars, from the process of birth to the brink of death. Now we are ready to explore the ways that stars end their lives. Sooner or later, each star exhausts its store of nuclear energy. Without a source of internal pressure to balance the weight of the overlying layers, every star eventually gives way to the inexorable pull of gravity and collapses under its own weight.

Following the rough distinction made in the last chapter, we will discuss the end-of-life evolution of stars of lower and higher mass separately. What determines the outcome-bang or whimper-is the mass of the star *when it is ready to die*, not the mass it was born with. As we noted in the last chapter, stars can lose a significant amount of mass in their middle and old age.

23.1 The Death of Low-Mass Stars



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 electrondegenerate 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 (<u>Figure 1</u>).

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.

Relating Masses and Radii of White Dwarfs.



Figure 1. Models of white-dwarf structure predict that as the mass of the star increases (toward the right), its radius gets smaller and smaller.

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 2). 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

 $630\,\mid$ Chapter 23 The Death of Stars Section 23.1: The Death of Low-Mass Stars

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 Russellrefused 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.

S. Chandrasekhar (1910–1995).



Figure 2.

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."

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 <u>3</u> shows the path of a star like the Sun on the H–R diagram during its final stages.

Evolutionary Track for a Star Like the Sun.



Figure 3. 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

634 | Chapter 23 The Death of Stars Section 23.1: The Death of Low-Mass Stars
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 4). 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.

Visible Light and X-Ray Images of the Sirius Star System.





Figure 4. (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 fifthclosest 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.

> Chapter 23 The Death of Stars Section 23.1: The Death of Low-Mass Stars | 635

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!

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} .

Key Concepts and Summary

During the course of their evolution, stars shed their outer layers and lose a significant fraction of their initial mass. Stars with masses of 8 M_{Sun} or less can lose enough mass to become white dwarfs, which have masses less than the Chandrasekhar limit (about 1.4 M_{Sun}). The pressure exerted by degenerate electrons keeps white dwarfs from contracting to still-smaller diameters. Eventually, white dwarfs cool off to become black dwarfs, stellar remnants made mainly of carbon, oxygen, and neon.

Glossary

Chandrasekhar limit

the upper limit to the mass of a white dwarf (equals 1.4 times the mass of the Sun)

degenerate gas

a gas that resists further compression because no two electrons can be in the same place at the same time doing the same thing (Pauli exclusion principle)

Chapter 23 Section 23.2: Evolution of Massive Stars: An Explosive Finish

23.2 Evolution of Massive Stars: An Explosive Finish



Thanks to mass loss, then, stars with starting masses up to 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 stillheavier 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 (<u>Figure 1</u>).

Structure of an Old Massive Star.



Figure 1. 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 [link] 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/cm³ (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 2). 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).

Five Supernova Explosions in Other Galaxies.



HST04Sas

HST04Yow

HST04Zwi

HST05Lan

HST05Str

Figure 2. 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))

Chapter 23 Section 23.2: Evolution of Massive Stars: An Explosive Finish | 645 <u>Table</u> 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.

The Ultimate Fate of Stars and Substella	r Objects with Different Masses
--	---------------------------------

Initial Mass (Mass of Sun = 1) ¹	Final State at the End of Its
< 0.01	Planet
0.01 to 0.08	Brown dwarf
0.08 to 0.25	White dwarf made mostly of
0.25 to 8	White dwarf made mostly of
8 to 10	White dwarf made of oxyger
10 to 40	Supernova explosion that lea
> 40	Supernova explosion that lea

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.

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 3).

Kepler Supernova Remant.



Figure 3. 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 multiwavelength 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 highenergy *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 highenergy 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.

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 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

$$F = M_1 \times a = GM_1M_2/R^2$$

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 × 10⁶ 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

 $g(whitedwarf) = \frac{_{G \times ^M sun}}{_{R_{earth} 2^2}} = \frac{_{6.67 \times 10^{-11} m^2/kgs^2 \times 2 \times 10^{30} kg}}{_{(6.4 \times 10^6 m)^2}} = 3.26 \times 10^6 m/s^2$

Compare this to g on the surface of Earth, which is 9.8 m/s².

Check Your Learning

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:

$$g(whitedwarf) = \frac{(G \times 2M_{sun})}{(0.5R_{Earth})^2} = \frac{(6.67 \times ^{-11}m^2/kgs^2 \times 4 \times 10^{30}kg)}{(3.2 \times 10^6)^2} = 2.61 \times 10^7 m/s^2$$

Key Concepts and Summary

In a massive star, hydrogen fusion in the core is followed by several other fusion reactions involving heavier elements. Just before it exhausts all sources of energy, a massive star has an iron core surrounded by shells of silicon, sulfur, oxygen, neon, carbon, helium, and hydrogen. The fusion of iron requires energy (rather than releasing it). If the mass of a star's iron core exceeds the Chandrasekhar limit (but is less than $3 M_{Sun}$), the core collapses until its density exceeds that of an atomic nucleus, forming a neutron star with a typical diameter of 20 kilometers. The core rebounds and transfers energy outward, blowing off the outer layers of the star in a type II supernova explosion.

Footnotes

• 1 Stars in the mass ranges 0.25-8 and 8-10 may later produce a

 $652\ |\ Chapter$ 23 Section 23.2: Evolution of Massive Stars: An Explosive Finish

type of supernova different from the one we have discussed so far. These are discussed in The Evolution of Binary Star Systems.

Glossary

neutron star

a compact object of extremely high density composed almost entirely of neutrons

type II supernova

a stellar explosion produced at the endpoint of the evolution of stars whose mass exceeds roughly 10 times the mass of the Sun

Chapter 23 Section 23.3: Supernova Observations

23.3 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.

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 1). 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.

Supernova 1006 Remnant.



Figure 1. 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 ([link]). 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 ([link]). 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 2); 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.

Supernova 2014].



Figure 2. 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 3). 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.

Hubble Space Telescope Image of SN 1987A.



Figure 3. 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 4). 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.

Ring around Supernova 1987A.



Figure 4. 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 <u>Table</u>. 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.

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

Evolution of the Star That Exploded as SN 1987A

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 5, 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 magnitude 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.

Change in the Brightness of SN 1987A over Time.



Figure 5. 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 5, 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 10⁵⁸ 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.

Key Concepts and Summary

A supernova occurs on average once every 25 to 100 years in the Milky Way Galaxy. Despite the odds, no supernova in our Galaxy has been observed from Earth since the invention of the telescope. However, one nearby supernova (SN 1987A) has been observed in a neighboring galaxy, the Large Magellanic Cloud. The star that evolved to become SN 1987A began its life as a blue supergiant, evolved to become a red supergiant, and returned to being a blue supergiant at the time it exploded. Studies of SN 1987A have detected neutrinos from the core collapse and confirmed theoretical calculations of what happens during such explosions, including the formation of elements beyond iron. Supernovae are a main source of high-energy cosmic rays and can be dangerous for any living organisms in nearby star systems.

Chapter 23 Section 23.4: Pulsars and the Discovery of Neutron Stars

23.4 Pulsars and the Discovery of Neutron Star



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 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.

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	$8 \times 10^5 \mathrm{g/cm^3}$	$10^{14} { m g/cm^3}$

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

Chapter 23 Section 23.4: Pulsars and the Discovery of Neutron Stars | 671

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, more than 2500 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 1). 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.

Crab Nebula.



Figure 1. This image shows X-ray emmisions 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.

View 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 2). 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.

Lighthouse.



674 | Chapter 23 Section 23.4: Pulsars and the Discovery of Neutron Stars

Figure 2. 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. of Recall the principle 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 3). 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.

Model of a Pulsar.



Figure 3. 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 3 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 4).

Speeding Pulsar.



Figure 4. 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)

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.

Key Concepts and Summary

At least some supernovae leave behind a highly magnetic, rapidly rotating neutron star, which can be observed as a pulsar if its beam of escaping particles and focused radiation is pointing toward us. Pulsars emit rapid pulses of radiation at regular intervals; their periods are in the range of 0.001 to 10 seconds. The rotating neutron star acts like a lighthouse, sweeping its beam in a circle and giving us a pulse of radiation when the beam sweeps over Earth. As pulsars age, they lose energy, their rotations slow, and their periods increase.

Glossary

pulsar

a variable radio source of small physical size that emits very rapid radio pulses in very regular periods that range from fractions of a second to several seconds; now understood to be a rotating, magnetic neutron star that is energetic enough to produce a detectable beam of radiation and particles

Chapter 23 Section 23.5: The Evolution of Binary Star Systems

23.5 The Evolution of Binary Star Systems



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.¹ 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 1. 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.

Evolution of a Binary System.



Figure 1. 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.

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.")

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.

Key Concepts and Summary

When a white dwarf or neutron star is a member of a close binary star system, its companion star can transfer mass to it. Material falling *gradually* onto a white dwarf can explode in a sudden burst of fusion and make a nova. If material falls *rapidly* onto a white dwarf, it can push it over the Chandrasekhar limit and cause it to explode completely as a type Ia supernova. Another possible mechanism for a type Ia supernova is the merger of two white dwarfs. Material falling onto a neutron star can cause powerful bursts of X-ray radiation. Transfer of material and angular momentum can speed up the rotation of pulsars until their periods are just a few thousandths of a second.

Footnotes

 1 We 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 lowmass 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.

Glossary

nova

the cataclysmic explosion produced in a binary system, temporarily increasing its luminosity by hundreds to thousands of times

millisecond pulsar

a pulsar that rotates so quickly that it can give off hundreds of pulses per second (and its period is therefore measured in milliseconds)

Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts

23.6 The Mystery of the Gamma-Ray Bursts

Learning Objectives

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

- Give a brief history of how gamma-ray bursts were discovered and what instruments made the discovery possible
- Explain why astronomers think that gamma-ray bursts beam their energy rather than it radiating uniformly in all directions
- Describe how the radiation from a gamma-ray burst and its afterglow is produced
- Explain how short-duration gamma-ray bursts differ from longer ones, and describe the process that makes short-duration gamma-ray bursts
- Explain why gamma-ray bursts may help us understand the early universe

Everybody loves a good mystery, and astronomers are no exception. The mystery we will discuss in this section was first discovered in the mid-1960s, not via astronomical research, but as a result of a search for the tell-tale signs of nuclear weapon explosions. The US Defense Department launched a series of *Vela* satellites to make sure that no country was violating a treaty that banned the detonation of nuclear weapons in space.

Since nuclear explosions produce the most energetic form of electromagnetic waves called *gamma rays* (see Radiation and Spectra), the Vela satellites contained detectors to search for this type of radiation. The satellites did not detect any confirmed events from human activities, but they did-to everyone's surprise-detect short bursts of gamma rays coming from random directions in the sky. News of the discovery was first published in 1973; however, the origin of the bursts remained a mystery. No one knew what produced the brief flashes of gamma rays or how far away the sources were.

From a Few Bursts to Thousands

With the launch of the Compton Gamma-Ray Observatory by NASA in 1991, astronomers began to identify many more bursts and to learn more about them (<u>Figure 1</u>). Approximately once per day, the NASA satellite detected a flash of gamma rays somewhere in the sky that lasted from a fraction of a second to several hundred seconds. Before the Compton measurements, astronomers had expected that the most likely place for the bursts to come from was the main disk of our own (pancake-shaped) Galaxy. If this had been the case, however, more bursts would have been seen in the crowded plane of the Milky Way than above or below it. Instead, the sources of the bursts were distributed *isotropically*; that is, they could appear anywhere in the sky with no preference for one region over another. Almost never did a second burst come from the same location.

Compton Detects Gamma-Ray Bursts.



Figure 1. (a) In 1991, the Compton Gamma-Ray Observatory was deployed by the Space Shuttle Atlantis. Weighing more than 16 tons, it was one of the largest scientific payloads ever launched into space. (b) This map of gamma-ray burst positions measured by the Compton Gamma-Ray Observatory shows the isotropic (same in all directions), uniform distribution of bursts on the sky. The map is oriented so that the disk of the Milky Way would stretch across the center line (or equator) of the oval. Note that the bursts show no preference at all for the plane of the Milky Way, as many other types of objects in the sky do. Colors indicate the total energy in the burst: red dots indicate long-duration, bright bursts; blue and purple dots show short, weaker bursts. (credit a: modification of work by NASA; credit b: modification of work by NASA/GSFC)

To get a good visual sense of the degree to which the bursts come from all over the sky, watch this short animated NASA video showing the location of the first 500 bursts found by the later *Swift* satellite.

For several years, astronomers actively debated whether the burst sources were relatively nearby or very far away-the two

Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts | 695

possibilities for bursts that are isotropically distributed. Nearby locations might include the cloud of comets that surrounds the solar system or the halo of our Galaxy, which is large and spherical, and also surrounds us in all directions. If, on the other hand, the bursts occurred at very large distances, they could come from faraway galaxies, which are also distributed uniformly in all directions.

Both the very local and the very distant hypotheses required something strange to be going on. If the bursts were coming from the cold outer reaches of our own solar system or from the halo of our Galaxy, then astronomers had to hypothesize some new kind of physical process that could produce unpredictable flashes of highenergy gamma rays in these otherwise-quiet regions of space. And if the bursts came from galaxies millions or billions of light-years away, then they must be extremely powerful to be observable at such large distances; indeed they had to be the among the biggest explosions in the universe.

The First Afterglows

The problem with trying to figure out the source of the gamma-ray bursts was that our instruments for detecting gamma rays could not pinpoint the exact place in the sky where the burst was happening. Early gamma-ray telescopes did not have sufficient resolution. This was frustrating because astronomers suspected that if they could pinpoint the exact position of one of these rapid bursts, then they would be able to identify a counterpart (such as a star or galaxy) at other wavelengths and learn much more about the burst, including where it came from. This would, however, require either major improvements in gamma-ray detector technology to provide better resolution or detection of the burst at some other wavelength. In the end, both techniques played a role.

The breakthrough came with the launch of the Italian

Dutch BeppoSAX satellite in 1996. BeppoSAX included a new type of gamma-ray telescope capable of identifying the position of a source much more accurately than previous instruments, to within a few minutes of arc on the sky. By itself, however, it was still not sophisticated enough to determine the exact source of the gamma-ray burst. After all, a box a few minutes of arc on a side could still contain many stars or other celestial objects.

However, the angular resolution of *BeppoSAX* was good enough to tell astronomers where to point other, more precise telescopes in the hopes of detecting longer-lived electromagnetic emission from the bursts at other wavelengths. Detection of a burst at visiblelight or radio wavelengths could provide a position accurate to a few seconds of arc and allow the position to be pinpointed to an individual star or galaxy. *BeppoSAX* carried its own X-ray telescope onboard the spacecraft to look for such a counterpart, and astronomers using visible-light and radio facilities on the ground were eager to search those wavelengths as well.

Two crucial BeppoSAX burst observations in 1997 helped to resolve the mystery of the gamma-ray bursts. The first burst came in February from the direction of the constellation Orion. Within 8 hours, astronomers working with the satellite had identified the position of the burst, and reoriented the spacecraft to focus BeppoSAX's X-ray detector on the source. To their excitement, they detected a slowly fading X-ray source 8 hours after the event—the first successful detection of an afterglow from a gammaray burst. This provided an even-better location of the burst (accurate to about 40 seconds of arc), which was then distributed to astronomers across the world to try to detect it at even longer wavelengths.

That very night, the 4.2-meter William Herschel Telescope on the Canary Islands found a fading visible-light source at the same position as the X-ray afterglow, confirming that such an afterglow could be detected in visible light as well. Eventually, the afterglow faded away, but left behind at the location of the original gammaray burst was a faint, fuzzy source right where the fading point of light had been–a distant galaxy (Figure 2). This was the first piece of evidence that gamma-ray bursts were indeed very energetic objects from very far away. However, it also remained possible that the burst source was much closer to us and just happened to align with a more distant galaxy, so this one observation alone was not a conclusive demonstration of the extragalactic origin of gamma-ray bursts.

Gamma-Ray Burst.



Figure 2. This false-color Hubble Space Telescope image, taken in September 1997, shows the fading afterglow of the gamma-ray burst of February 28, 1997 and the host galaxy in which the burst originated. The left view shows the region of the burst. The enlargement shows the burst source and what appears to be its host galaxy. Note that the gamma-ray source is not in the center of the galaxy. (credit: modification of work by Andrew Fruchter

(STScI), Elena Pian (ITSRE-CNR), and NASA, ESA) On May 8 of the same year, a burst came from the direction of the constellation Camelopardalis. In a coordinated international effort, *BeppoSAX* again fixed a reasonably precise position, and almost immediately a telescope on Kitt Peak in Arizona was able to catch the visible-light afterglow. Within 2 days, the largest telescope in the world (the Keck in Hawaii) collected enough light to record a spectrum of the burst. The May gamma-ray burst afterglow spectrum showed absorption features from a fuzzy object that was 4 billion light-years from the Sun, meaning that the location of the burst had to be at least this far away-and possibly even farther. (How astronomers can get the distance of such an object from the Doppler shift in the spectrum is something we will discuss in Galaxies.) What that spectrum showed was clear evidence that the gamma-ray burst had taken place in a distant galaxy.

Networking to Catch More Bursts

After initial observations showed that the precise locations and afterglows of gamma-ray bursts could be found, astronomers set up a system to catch and pinpoint bursts on a regular basis. But to respond as quickly as needed to obtain usable results, astronomers realized that they needed to rely on automated systems rather than human observers happening to be in the right place at the right time.

Now, when an orbiting high-energy telescope discovers a burst, its rough location is immediately transmitted to a *Gamma-Ray Coordinates* Network based at NASA's Goddard Space Flight Center, alerting observers on the ground within a few seconds to look for the visible-light afterglow.

The first major success with this system was achieved by a team of astronomers from the University of Michigan, Lawrence Livermore National Laboratory, and Los Alamos National Laboratories, who designed an automated device they called the Robotic Optical Transient Search Experiment (ROTSE), which detected a very bright visible-light counterpart in 1999. At peak, the burst was almost as bright as Neptune–despite a distance (measured later by spectra from larger telescopes) of 9 billion lightyears. More recently, astronomers have been able to take this a step further, using wide-field-of-view telescopes to stare at large fractions of the sky in the hope that a gamma-ray burst will occur at the right place and time, and be recorded by the telescope's camera. These wide-field telescopes are not sensitive to faint sources, but ROTSE showed that gamma-ray burst afterglows could sometimes be very bright.

Astronomers' hopes were vindicated in March 2008, when an extremely bright gamma-ray burst occurred and its light was captured by two wide-field camera systems in Chile: the Polish "Pi of the Sky" and the Russian-Italian TORTORA [Telescopio Ottimizzato per la Ricerca dei Transienti Ottici Rapidi (Italian for Telescope Optimized for the Research of Rapid Optical Transients)] (see Figure 3). According to the data taken by these telescopes, for a period of about 30 seconds, the light from the gamma-ray burst was bright enough that it could have been seen by the unaided eye had a person been looking in the right place at the right time. Adding to our amazement, later observations by larger telescopes demonstrated that the burst occurred at a distance of 8 billion light-years from Earth!

Gamma-Ray Burst Observed in March 2008.



700 | Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts

Figure 3. The extremely luminous afterglow of GRB 080319B was imaged by the Swift Observatory in X-rays (left) and visible light/ ultraviolet (right). (credit: modification of work by NASA/Swift/ Stefan Immler, et al.)

To Beam or Not to Beam

The enormous distances to these events meant they had to have been astoundingly energetic to appear as bright as they were across such an enormous distance. In fact, they required so much energy that it posed a problem for gamma-ray burst models: if the source was radiating energy in all directions, then the energy released in gamma rays alone during a bright burst (such as the 1999 or 2008 events) would have been equivalent to the energy produced if the entire mass of a Sun-like star were suddenly converted into pure radiation.

For a source to produce this much energy this quickly (in a burst) is a real challenge. Even if the star producing the gamma-ray burst was much more massive than the Sun (as is probably the case), there is no known means of converting so much mass into radiation within a matter of seconds. However, there is one way to reduce the power required of the "mechanism" that makes gamma-ray bursts. So far, our discussion has assumed that the source of the gamma rays gives off the same amount of energy in all directions, like an incandescent light bulb.

But as we discuss in Pulsars and the Discovery of Neutron Stars, not all sources of radiation in the universe are like this. Some produce thin beams of radiation that are concentrated into only one or two directions. A laser pointer and a lighthouse on the ocean are examples of such beamed sources on Earth (Figure 4). If, when a burst occurs, the gamma rays come out in only one or two narrow beams, then our estimates of the luminosity of the source can be reduced, and the bursts may be easier to explain. In that case,

however, the beam has to point toward Earth for us to be able to see the burst. This, in turn, would imply that for every burst we see from Earth, there are probably many others that we never detect because their beams point in other directions.

Burst That Is Beamed.



Figure 4. This artist's conception shows an illustration of one kind of gamma-ray burst. The collapse of the core of a massive star into a black hole has produced two bright beams of light originating from the star's poles, which an observer pointed along one of these axes would see as a gamma-ray burst. The hot blue stars and gas clouds in the vicinity are meant to show that the

Long-Duration Gamma-Ray Bursts: Exploding Stars

After identifying and following large numbers of gamma-ray bursts, astronomers began to piece together clues about what kind of event is thought to be responsible for producing the gamma-ray burst. Or, rather, what kind of *events*, because there are at least two distinct types of gamma-ray bursts. The two-like the different types of supernovae-are produced in completely different ways.

Observationally, the crucial distinction is how long the burst lasts. Astronomers now divide gamma-ray bursts into two categories: short-duration ones (defined as lasting less than 2 seconds, but typically a fraction of a second) and long-duration ones (defined as lasting more than 2 seconds, but typically about a minute).

All of the examples we have discussed so far concern the longduration gamma-ray bursts. These constitute most of the gammaray bursts that our satellites detect, and they are also brighter and easier to pinpoint. Many hundreds of long-duration gamma-ray bursts, and the properties of the galaxies in which they occurred, have now been studied in detail. Long-duration gamma-ray bursts are universally observed to come from distant galaxies that are still actively making stars. They are usually found to be located in regions of the galaxy with strong star-formation activity (such as spiral arms). Recall that the more massive a star is, the less time it spends in each stage of its life. This suggests that the bursts come from a young and short-lived, and therefore massive type of star.

Furthermore, in several cases when a burst has occurred in a galaxy relatively close to Earth (within a few billion light-years), it has been possible to search for a supernova at the same position-and in nearly all of these cases, astronomers have found evidence of a supernova of type Ic going off. A type Ic is a particular type of supernova, which we did not discuss in the earlier parts of this chapter; these are produced by a massive star that has been stripped of its outer hydrogen layer. However, only a tiny fraction of type Ic supernovae produce gamma-ray bursts.

Why would a massive star with its outer layers missing sometimes produce a gamma-ray burst at the same time that it explodes as a supernova? The explanation astronomers have in mind for the extra energy is the collapse of the star's core to form a spinning, magnetic black hole or neutron star. Because the star corpse is both magnetic and spinning rapidly, its sudden collapse is complex and can produce swirling jets of particles and powerful beams of radiation–just like in a quasar or active galactic nucleus (objects you will learn about Active Galaxies, Quasars, and Supermassive Black Holes), but on a much faster timescale. A small amount of the infalling mass is ejected in a narrow beam, moving at speeds close to that of light. Collisions among the particles in the beam can produce intense bursts of energy that we see as a gamma-ray burst.

Within a few minutes, the expanding blast from the fireball plows into the interstellar matter in the dying star's neighborhood. This matter might have been ejected from the star itself at earlier stages in its evolution. Alternatively, it could be the gas out of which the massive star and its neighbors formed.

As the high-speed particles from the blast are slowed, they transfer their energy to the surrounding matter in the form of a shock wave. That shocked material emits radiation at longer wavelengths. This accounts for the afterglow of X-rays, visible light, and radio waves—the glow comes at longer and longer wavelengths as the blast continues to lose energy.

Short-Duration Gamma-Ray Bursts: Colliding Stellar Corpses

What about the shorter gamma-ray bursts? The gamma-ray emission from these events lasts less than 2 seconds, and in some cases may last only milliseconds—an amazingly short time. Such a timescale is difficult to achieve if they are produced in the same way as long-duration gamma-ray bursts, since the collapse of the stellar interior onto the black hole should take at least a few seconds.

Astronomers looked fruitlessly for afterglows from shortduration gamma-ray bursts found by *BeppoSAX* and other satellites. Evidently, the afterglows fade away too quickly. Fast-responding visible-light telescopes like ROTSE were not helpful either: no matter how fast these telescopes responded, the bursts were not bright enough at visible wavelengths to be detected by these small telescopes.

Once again, it took a new satellite to clear up the mystery. In this case, it was the *Swift Gamma-Ray Burst Satellite*, launched in 2004 by a collaboration between NASA and the Italian and UK space agencies (Figure 5). The design of *Swift* is similar to that of *BeppoSAX*. However, *Swift* is much more agile and flexible: after a gamma-ray burst occurs, the X-ray and UV telescopes can be repointed automatically within a few minutes (rather than a few hours). Thus, astronomers can observe the afterglow much earlier, when it is expected to be much brighter. Furthermore, the X-ray telescope is far more sensitive and can provide positions that are 30 times more precise than those provided by *BeppoSAX*, allowing bursts to be identified even without visible-light or radio observations.

Artist's Illustration of Swift.



Figure 5. The US/UK/Italian spacecraft *Swift* contains on-board gamma-ray, X-ray, and ultraviolet detectors, and has the ability to automatically reorient itself to a gamma-ray burst detected by the

gamma-ray instrument. Since its launch in 2005, Swift has detected and observed over a thousand bursts, including dozens of

short-duration bursts. (credit: NASA, Spectrum Astro)

On May 9, 2005, *Swift* detected a flash of gamma rays lasting 0.13 seconds in duration, originating from the constellation Coma Berenices. Remarkably, the galaxy at the X-ray position looked completely different from any galaxy in which a long-duration burst had been seen to occur. The afterglow originated from the halo of a giant elliptical galaxy 2.7 billion light-years away, with no signs of any young, massive stars in its spectrum. Furthermore, no supernova was ever detected after the burst, despite extensive searching.

What could produce a burst less than a second long, originating from a region with no star formation? The leading model involves

706 | Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts
the *merger* of two compact stellar corpses: two neutron stars, or perhaps a neutron star and a black hole. Since many stars come in binary or multiple systems, it's possible to have systems where two such star corpses orbit one another. According to general relativity (which will be discussed in Black Holes and Curved Spacetime), the orbits of a binary star system composed of such objects should slowly decay with time, eventually (after millions or billions of years) causing the two objects to slam together in a violent but brief explosion. Because the decay of the binary orbit is so slow, we would expect more of these mergers to occur in old galaxies in which star formation has long since stopped.

To learn more about the merger of two neutron stars and how they can produce a burst that lasts less than a second, check out this computer simulation by NASA.

While it was impossible to be sure of this model based on only a single event (it is possible this burst actually came from a background galaxy and lined up with the giant elliptical only by chance), several dozen more short-duration gamma-ray bursts have since been located by *Swift*, many of which also originate from galaxies with very low star-formation rates. This has given astronomers greater confidence that this model is the correct one. Still, to be fully convinced, astronomers are searching for a "smoking gun" signature for the merger of two ultra-dense stellar remnants.

Astronomers identified two observations that would provide more direct evidence. Theoretical calculations indicate that when two neutron stars collide there will be a very special kind of explosion; neutrons stripped from the neutron stars during the violent final phase of the merger will fuse together into heavy elements and then release heat due to radioactivity, producing a short-lived but red supernova sometimes called a *kilonova*. (The term is used because it is about a thousand times brighter than an ordinary nova, but not quite as "super" as a traditional supernova.) Hubble observations of one short-duration gamma-ray burst in 2013 showed suggestive evidence of such a signature, but needed to be confirmed by future observations.

The second "smoking gun" is the detection of gravitational waves. As will be discussed in Black Holes and Curved Spacetime, gravitational waves are ripples in the fabric of spacetime that general relativity predicts should be produced by the acceleration of extremely massive and dense objects-such as two neutron stars or black holes spiraling toward each other and colliding. The construction of instruments to detect gravitational waves is very challenging technically, and gravitational wave astronomy became feasible only in 2015. The first few detected gravitational wave events were produced by mergers of black holes. In 2017, however, gravitational waves were observed from a source that was coincident in time and space with a gamma-ray burst. The source consisted of two objects with the masses of neutron stars. A red supernova was also observed at this location, and the ejected material was rich in heavy elements. This observation not only confirms the theory of the origin of short gamma-ray bursts, but also is a spectacular demonstration of the validity of Einstein's theory of general relativity.

Probing the Universe with Gamma-Ray Bursts

The story of how astronomers came to explain the origin of the different kinds of bursts is a good example of how the scientific process sometimes resembles good detective work. While the mystery of short-duration gamma-ray bursts is still being unraveled, the focus of studies for long-duration gamma-ray bursts has begun to change from understanding the origin of the bursts

themselves (which is now fairly well-established) to using them as tools to understand the broader universe.

The reason that long-duration gamma-ray bursts are useful has to do with their extreme luminosities, if only for a short time. In fact, long-duration gamma-ray bursts are so bright that they could easily be seen at distances that correspond to a few hundred million years after the expansion of the universe began, which is when theorists think that the first generation of stars formed. Some theories predict that the first stars are likely to be massive and complete their evolution in only a million years or so. If this turns out to be the case, then gamma-ray bursts (which signal the death of some of these stars) may provide us with the best way of probing the universe when stars and galaxies first began to form.

So far, the most distant gamma-ray burst found (on April 29, 2009) was in a galaxy with a redshift that corresponds to a remarkable 13.2 billion light years-meaning it happened only 600 million years after the Big Bang itself. This is comparable to the earliest and most distant galaxies found by the Hubble Space Telescope. It is not quite old enough to expect that it formed from the first generation of stars, but its appearance at this distance still gives us useful information about the production of stars in the early universe. Astronomers continue to scan the skies, looking for even more distant events signaling the deaths of stars from even further back in time.

Key Concepts and Summary

Gamma-ray bursts last from a fraction of a second to a few minutes. They come from all directions and are now known to be associated with very distant objects. The energy is most likely beamed, and, for the ones we can detect, Earth lies in the direction of the beam. Long-duration bursts (lasting more than a few seconds) come from massive stars with their outer hydrogen layers missing that explode as supernovae. Short-duration bursts are believed to be mergers of stellar corpses (neutron stars or black holes).

For Further Exploration

Articles

Death of Stars

Hillebrandt, W., et al. "How To Blow Up a Star." Scientific American (October 2006): 42. On supernova mechanisms.

Irion, R. "Pursuing the Most Extreme Stars." Astronomy (January 1999): 48. On pulsars.

Kalirai, J. "New Light on Our Sun's Fate." Astronomy (February 2014): 44. What will happen to stars like our Sun between the main sequence and the white dwarf stages.

Kirshner, R. "Supernova 1987A: The First Ten Years." Sky & Telescope (February 1997): 35.

Maurer, S. "Taking the Pulse of Neutron Stars." Sky & Telescope (August 2001): 32. Review of recent ideas and observations of pulsars.

Zimmerman, R. "Into the Maelstrom." Astronomy (November 1998): 44. About the Crab Nebula.

Gamma-Ray Bursts

Fox, D. & Racusin, J. "The Brightest Burst." Sky & Telescope (January 2009): 34. Nice summary of the brightest burst observed so far, and what we have learned from it.

Nadis, S. "Do Cosmic Flashes Reveal Secrets of the Infant

Universe?" Astronomy (June 2008): 34. On different types of gammaray bursts and what we can learn from them.

Naeye, R. "Dissecting the Bursts of Doom." Sky & Telescope (August 2006): 30. Excellent review of gamma-ray bursts-how we discovered them, what they might be, and what they can be used for in probing the universe.

Zimmerman, R. "Speed Matters." Astronomy (May 2000): 36. On the quick-alert networks for finding afterglows.

Zimmerman, R. "Witness to Cosmic Collisions." Astronomy (July 2006): 44. On the Swift mission and what it is teaching astronomers about gamma-ray bursts.

Websites

Death of Stars

Crab Nebula: http://chandra.harvard.edu/xray_sources/crab/ crab.html. A short, colorfully written introduction to the history and science involving the best-known supernova remant.

Introduction to Neutron Stars: https://www.astro.umd.edu/~miller/nstar.html. Coleman Miller of the University of Maryland maintains this site, which goes from easy to hard as you get into it, but it has lots of good information about corpses of massive stars.

Introduction to Pulsars (by Maryam Hobbs at the Australia National Telescope Facility): http://www.atnf.csiro.au/outreach/education/everyone/pulsars/index.html.

Magnetars, Soft Gamma Repeaters, and Very Strong Magnetic Fields: http://solomon.as.utexas.edu/magnetar.html. Robert Duncan, one of the originators of the idea of magnetars, assembled this site some years ago. Gamma-Ray Bursts

Brief Intro to Gamma-Ray Bursts (from PBS' Seeing in the Dark): http://www.pbs.org/seeinginthedark/astronomy-topics/gamma-ray-bursts.html.

Discovery of Gamma-ray Bursts: http://science.nasa.gov/ science-news/science-at-nasa/1997/ast19sep97_2/.

Gamma-Ray Bursts: Introduction to a Mystery (at NASA's Imagine the Universe site): http://imagine.gsfc.nasa.gov/docs/science/know_ll/bursts.html.

Introduction from the Swift Satellite Site: http://swift.sonoma.edu/about_swift/grbs.html.

Missions to Detect and Learn More about Gamma-ray Bursts:

- Fermi Space Telescope: http://fermi.gsfc.nasa.gov/public/.
- INTEGRAL Spacecraft: http://www.esa.int/science/integral.
- SWIFT Spacecraft: http://swift.sonoma.edu/.

Videos

Death of Stars

BBC interview with Antony Hewish: http://www.bbc.co.uk/archive/scientists/10608.shtml. (40:54).

Black Widow Pulsars: The Vengeful Corpses of Stars: https://www.youtube.com/watch?v=Fn-3G_N0hy4. A public talk in the Silicon Valley Astronomy Lecture Series by Dr. Roger Romani (Stanford University) (1:01:47).

Hubblecast 64: It all ends with a bang!: http://www.spacetelescope.org/videos/hubblecast64a/. HubbleCast Program introducing Supernovae with Dr. Joe Liske

(9:48).

Space Movie Reveals Shocking Secrets of the Crab

712 | Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts

Pulsar: http://hubblesite.org/newscenter/archive/releases/

2002/24/video/c/. A sequence of Hubble and Chandra Space Telescope images of the central regions of the Crab Nebula have been assembled into a very brief movie accompanied by animation showing how the pulsar affects its environment; it comes with some useful background material (40:06).

Gamma-Ray Bursts

Gamma-Ray Bursts: The Biggest Explosions Since the Big Bang!: https://www.youtube.com/watch?v=ePo_EdgV764. Edo Berge in a popular-level lecture at Harvard (58:50).

Gamma-Ray Bursts: Flashes in the Sky: https://www.youtube.com/watch?v=23EhcAP3O8Q. American Museum of Natural History Science Bulletin on the Swift satellite (5:59).

Overview Animation of Gamma-Ray Burst: http://news.psu.edu/ video/296729/2013/11/27/overview-animation-gamma-ray-burst. Brief Animation of what causes a long-duration gamma-ray burst (0:55).

Collaborative Group Activities

- A. Someone in your group uses a large telescope to observe an expanding shell of gas. Discuss what measurements you could make to determine whether you have discovered a planetary nebula or the remnant of a supernova explosion.
- B. The star Sirius (the brightest star in our northern skies) has a white-dwarf companion. Sirius has a mass of about 2 M_{Sun} and is still on the main sequence, while its companion is already a star corpse. Remember that a white dwarf can't have a mass greater than 1.4 M_{Sun} . Assuming that the two stars formed at

the same time, your group should discuss how Sirius could have a white-dwarf companion. Hint: Was the initial mass of the white-dwarf star larger or smaller than that of Sirius?

- C. Discuss with your group what people today would do if a brilliant star suddenly became visible during the daytime?What kind of fear and superstition might result from a supernova that was really bright in our skies? Have your group invent some headlines that the tabloid newspapers and the less responsible web news outlets would feature.
- D. Suppose a supernova exploded only 40 light-years from Earth. Have your group discuss what effects there may be on Earth when the radiation reaches us and later when the particles reach us. Would there be any way to protect people from the supernova effects?
- E. When pulsars were discovered, the astronomers involved with the discovery talked about finding "little green men." If you had been in their shoes, what tests would you have performed to see whether such a pulsating source of radio waves was natural or the result of an alien intelligence? Today, several groups around the world are actively searching for possible radio signals from intelligent civilizations. How might you expect such signals to differ from pulsar signals?
- F. Your little brother, who has not had the benefit of an astronomy course, reads about white dwarfs and neutron stars in a magazine and decides it would be fun to go near them or even try to land on them. Is this a good idea for future tourism? Have your group make a list of reasons it would not be safe for children (or adults) to go near a white dwarf and a neutron star.
- G. A lot of astronomers' time and many instruments have been devoted to figuring out the nature of gamma-ray bursts. Does your group share the excitement that astronomers feel about these mysterious high-energy events? What are some reasons that people outside of astronomy might care about learning about gamma-ray bursts?

Review Questions

How does a white dwarf differ from a neutron star? How does each form? What keeps each from collapsing under its own weight?

Describe the evolution of a star with a mass like that of the Sun, from the main-sequence phase of its evolution until it becomes a white dwarf.

Describe the evolution of a massive star (say, 20 times the mass of the Sun) up to the point at which it becomes a supernova. How does the evolution of a massive star differ from that of the Sun? Why?

How do the two types of supernovae discussed in this chapter differ? What kind of star gives rise to each type?

A star begins its life with a mass of 5 M_{Sun} but ends its life as a white dwarf with a mass of 0.8 M_{Sun} . List the stages in the star's life during which it most likely lost some of the mass it started with. How did mass loss occur in each stage?

If the formation of a neutron star leads to a supernova explosion, explain why only three of the hundreds of known pulsars are found in supernova remnants.

How can the Crab Nebula shine with the energy of something like 100,000 Suns when the star that formed the nebula exploded almost 1000 years ago? Who "pays the bills" for much of the radiation we see coming from the nebula?

How is a nova different from a type Ia supernova? How does it differ from a type II supernova?

Apart from the masses, how are binary systems with a neutron star different from binary systems with a white dwarf?

What observations from SN 1987A helped confirm theories about supernovae?

Describe the evolution of a white dwarf over time, in particular how the luminosity, temperature, and radius change.

Describe the evolution of a pulsar over time, in particular how the rotation and pulse signal changes over time.

How would a white dwarf that formed from a star that had an initial mass of $1 M_{Sun}$ be different from a white dwarf that formed from a star that had an initial mass of $9 M_{Sun}$?

What do astronomers think are the causes of longer-duration gamma-ray bursts and shorter-duration gamma-ray bursts?

How did astronomers finally solve the mystery of what gamma-ray bursts were? What instruments were required to find the solution?

Thought Questions

Arrange the following stars in order of their evolution:

- A. A star with no nuclear reactions going on in the core, which is made primarily of carbon and oxygen.
- B. A star of uniform composition from center to surface; it contains hydrogen but has no nuclear reactions going on in the core.
- C. A star that is fusing hydrogen to form helium in its core.
- D. A star that is fusing helium to carbon in the core and hydrogen to helium in a shell around the core.
- E. A star that has no nuclear reactions going on in the core but is fusing hydrogen to form helium in a shell around the core.

Would you expect to find any white dwarfs in the Orion Nebula?

716 | Chapter 23 Section 23.6: The Mystery of the Gamma-Ray Bursts

(See The Birth of Stars and the Discovery of Planets outside the Solar System to remind yourself of its characteristics.) Why or why not?

Suppose no stars more massive than about 2 M_{Sun} had ever formed. Would life as we know it have been able to develop? Why or why not?

Would you be more likely to observe a type II supernova (the explosion of a massive star) in a globular cluster or in an open cluster? Why?

Astronomers believe there are something like 100 million neutron stars in the Galaxy, yet we have only found about 2000 pulsars in the Milky Way. Give several reasons these numbers are so different. Explain each reason.

Would you expect to observe every supernova in our own Galaxy? Why or why not?

The Large Magellanic Cloud has about one-tenth the number of stars found in our own Galaxy. Suppose the mix of high- and low-mass stars is exactly the same in both galaxies. Approximately how often does a supernova occur in the Large Magellanic Cloud?

Look at the list of the nearest stars in Appendix I. Would you expect any of these to become supernovae? Why or why not?

If most stars become white dwarfs at the ends of their lives and the formation of white dwarfs is accompanied by the production of a planetary nebula, why are there more white dwarfs than planetary nebulae in the Galaxy?

If a 3 and 8 $\ensuremath{M_{Sun}}$ star formed together in a binary system, which star would:

A. Evolve off the main sequence first?

- B. Form a carbon- and oxygen-rich white dwarf?
- C. Be the location for a nova explosion?

You have discovered two star clusters. The first cluster contains mainly main-sequence stars, along with some red giant stars and a few white dwarfs. The second cluster also contains mainly mainsequence stars, along with some red giant stars, and a few neutron stars—but no white dwarf stars. What are the relative ages of the clusters? How did you determine your answer?

A supernova remnant was recently discovered and found to be approximately 150 years old. Provide possible reasons that this supernova explosion escaped detection.

Based upon the evolution of stars, place the following elements in order of least to most common in the Galaxy: gold, carbon, neon. What aspects of stellar evolution formed the basis for how you ordered the elements?

What observations or types of telescopes would you use to distinguish a binary system that includes a main-sequence star and a white dwarf star from one containing a main-sequence star and a neutron star?

How would the spectra of a type II supernova be different from a type Ia supernova? Hint: Consider the characteristics of the objects that are their source.

Figuring for Yourself

The ring around SN 1987A ([link]) initially became illuminated when energetic photons from the supernova interacted with the material in the ring. The radius of the ring is approximately 0.75 light-year from the supernova location. How long after the supernova did the ring become illuminated?

What is the acceleration of gravity (*g*) at the surface of the Sun? (See Appendix E for the Sun's key characteristics.) How much greater is this than *g* at the surface of Earth? Calculate what you would weigh on the surface of the Sun. Your weight would be your Earth weight multiplied by the ratio of the acceleration of gravity on the Sun to the acceleration of gravity on Earth. (Okay, we know that the Sun does not have a solid surface to stand on and that you would be vaporized if you were at the Sun's photosphere. Humor us for the sake of doing these calculations.)

What is the escape velocity from the Sun? How much greater is it than the escape velocity from Earth?

What is the average density of the Sun? How does it compare to the average density of Earth?

Say that a particular white dwarf has the mass of the Sun but the radius of Earth. What is the acceleration of gravity at the surface of the white dwarf? How much greater is this than *g* at the surface of Earth? What would you weigh at the surface of the white dwarf (again granting us the dubious notion that you could survive there)?

What is the escape velocity from the white dwarf in Exercise? How much greater is it than the escape velocity from Earth?

What is the average density of the white dwarf in Exercise? How does it compare to the average density of Earth?

Now take a neutron star that has twice the mass of the Sun but a radius of 10 km. What is the acceleration of gravity at the surface of the neutron star? How much greater is this than *g* at the surface of Earth? What would you weigh at the surface of the neutron star (provided you could somehow not become a puddle of protoplasm)?

What is the escape velocity from the neutron star in Exercise? How much greater is it than the escape velocity from Earth?

What is the average density of the neutron star in Exercise? How does it compare to the average density of Earth?

One way to calculate the radius of a star is to use its luminosity and temperature and assume that the star radiates approximately like a blackbody. Astronomers have measured the characteristics of central stars of planetary nebulae and have found that a typical central star is 16 times as luminous and 20 times as hot (about 110,000 K) as the Sun. Find the radius in terms of the Sun's. How does this radius compare with that of a typical white dwarf?

According to a model described in the text, a neutron star has a radius of about 10 km. Assume that the pulses occur once per rotation. According to Einstein's theory of relatively, nothing can move faster than the speed of light. Check to make sure that this pulsar model does not violate relativity. Calculate the rotation speed of the Crab Nebula pulsar at its equator, given its period of 0.033 s. (Remember that distance equals velocity \times time and that the circumference of a circle is given by $2\pi R$).

Do the same calculations as in Exercise but for a pulsar that rotates 1000 times per second.

If the Sun were replaced by a white dwarf with a surface temperature of 10,000 K and a radius equal to Earth's, how would its luminosity compare to that of the Sun?

A supernova can eject material at a velocity of 10,000 km/s. How long would it take a supernova remnant to expand to a radius of 1 AU? How long would it take to expand to a radius of 1 lightyears? Assume that the expansion velocity remains constant and use the relationship: expansion time=distanceexpansion velocity.expansion time=distanceexpansion velocity. A supernova remnant was observed in 2007 to be expanding at a velocity of 14,000 km/s and had a radius of 6.5 light-years. Assuming a constant expansion velocity, in what year did this supernova occur?

The ring around SN 1987A ([link]) started interacting with material propelled by the shockwave from the supernova beginning in 1997 (10 years after the explosion). The radius of the ring is approximately 0.75 light-year from the supernova location. How fast is the supernova material moving, assume a constant rate of motion in km/s?

Before the star that became SN 1987A exploded, it evolved from a red supergiant to a blue supergiant while remaining at the same luminosity. As a red supergiant, its surface temperature would have been approximately 4000 K, while as a blue supergiant, its surface temperature was 16,000 K. How much did the radius change as it evolved from a red to a blue supergiant?

What is the radius of the progenitor star that became SN 1987A? Its luminosity was 100,000 times that of the Sun, and it had a surface temperature of 16,000 K.

What is the acceleration of gravity at the surface of the star that became SN 1987A? How does this g compare to that at the surface of Earth? The mass was 20 times that of the Sun and the radius was 41 times that of the Sun.

What was the escape velocity from the surface of the SN 1987A progenitor star? How much greater is it than the escape velocity from Earth? The mass was 20 times that of the Sun and the radius was 41 times that of the Sun.

What was the average density of the star that became SN 1987A? How does it compare to the average density of Earth? The mass was 20 times that of the Sun and the radius was 41 times that of the Sun. If the pulsar shown in [link] is rotating 100 times per second, how many pulses would be detected in one minute? The two beams are located along the pulsar's equator, which is aligned with Earth.

Chapter 24 Black Holes and Curved Spacetime Section 24.1: Introduction General Relativity

Stellar Mass Black Hole.



Figure 1. On the left, a visible-light image shows a region of the sky in the constellation of Cygnus; the red box marks the position of the X-ray source Cygnus X-1. It is an example of a black hole created when a massive star collapses at the end of its life. Cygnus X-1 is in a binary star system, and the artist's illustration on the right shows the black hole pulling material away from a massive blue companion star. This material forms a disk (shown in red and orange) that rotates around the black hole before falling into it or being redirected away from the black hole in the form of powerful jets. The material in the disk (before it falls into the black hole) is so hot that it glows with X-rays, explaining why this object is an X-ray source. (credit left: modification of work by DSS; credit right: modification of work by NASA/CXC/M.Weiss)

For most of the twentieth century, black holes seemed the stuff of science fiction, portrayed either as monster vacuum cleaners consuming all the matter around them or as tunnels from one universe to another. But the truth about black holes is almost stranger than fiction. As we continue our voyage into the universe, we will discover that black holes are the key to explaining many mysterious and remarkable objects-including collapsed stars and the active centers of giant galaxies.

24.1 Introducing General Relativity

Learning Objectives

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

- Discuss some of the key ideas of the theory of general relativity
- Recognize that one's experiences of gravity and acceleration are interchangeable and indistinguishable
- Distinguish between Newtonian ideas of gravity and Einsteinian ideas of gravity
- Recognize why the theory of general relativity is necessary for understanding the nature of black holes

Most stars end their lives as white dwarfs or neutron stars. When a very massive star collapses at the end of its life, however, not even the mutual repulsion between densely packed neutrons can support the core against its own weight. If the remaining mass of the star's core is more than about three times that of the Sun (M_{Sun}), our theories predict that no known force can stop it from collapsing forever! Gravity simply overwhelms all other forces and crushes the core until it occupies an infinitely small volume. A star in which this occurs may become one of the strangest objects ever predicted by theory–a black hole.

To understand what a black hole is like and how it influences its surroundings, we need a theory that can describe the action of gravity under such extreme circumstances. To date, our best theory of gravity is the general theory of relativity, which was put forward in 1916 by Albert Einstein.

General relativity was one of the major intellectual achievements of the twentieth century; if it were music, we would compare it to the great symphonies of Beethoven or Mahler. Until recently, however, scientists had little need for a better theory of gravity; Isaac Newton's ideas that led to his law of universal gravitation (see Orbits and Gravity) are perfectly sufficient for most of the objects we deal with in everyday life. In the past half century, however, general relativity has become more than just a beautiful idea; it is now essential in understanding pulsars, quasars (which will be discussed in Active Galaxies, Quasars, and Supermassive Black Holes), and many other astronomical objects and events, including the black holes we will discuss here.

We should perhaps mention that this is the point in an astronomy course when many students start to feel a little nervous (and perhaps wish they had taken botany or some other earthbound course to satisfy the science requirement). This is because in popular culture, Einstein has become a symbol for mathematical brilliance that is simply beyond the reach of most people (<u>Figure 1</u>).

Albert Einstein (1879–1955).



Figure 1. This famous scientist, seen here younger than in the usual photos, has become a symbol for high intellect in popular culture. (credit: NASA)

So, when we wrote that the theory of general relativity was Einstein's work, you may have worried just a bit, convinced that anything Einstein did must be beyond your understanding. This popular view is unfortunate and mistaken. Although the detailed calculations of general relativity do involve a good deal of higher mathematics, the basic ideas are not difficult to understand (and are, in fact, almost poetic in the way they give us a new perspective on the world). Moreover, general relativity goes beyond Newton's famous "inverse-square" law of gravity; it helps *explain* how matter interacts with other matter in space and time. This explanatory power is one of the requirements that any successful scientific theory must meet.

The Principle of Equivalence

The fundamental insight that led to the formulation of the general theory of relativity starts with a very simple thought: if you were able to jump off a high building and fall freely, you would not feel your own weight. In this chapter, we will describe how Einstein built on this idea to reach sweeping conclusions about the very fabric of space and time itself. He called it the "happiest thought of my life."

Einstein himself pointed out an everyday example that illustrates this effect (see Figure 2). Notice how your weight seems to be reduced in a high-speed elevator when it accelerates from a stop to a rapid descent. Similarly, your weight seems to increase in an elevator that starts to move quickly upward. This effect is not just a feeling you have: if you stood on a scale in such an elevator, you could measure your weight changing (you can actually perform this experiment in some science museums).

Your Weight in an Elevator.



Figure 2. In an elevator at rest, you feel your normal weight. In an elevator that accelerates as it descends, you would feel lighter than normal. In an elevator that accelerates as it ascends, you would feel heavier than normal. If an evil villain cut the elevator cable, you

would feel weightless as you fell to your doom.

In a *freely falling* elevator, with no air friction, you would lose your weight altogether. We generally don't like to cut the cables holding elevators to try this experiment, but near-weightlessness can be achieved by taking an airplane to high altitude and then dropping rapidly for a while. This is how NASA trains its astronauts for the experience of free fall in space; the scenes of weightlessness in the 1995 movie *Apollo* 13 were filmed in the same way. (Moviemakers have since devised other methods using underwater filming, wire stunts, and computer graphics to create the appearance of weightlessness seen in such movies as *Gravity* and *The Martian*.)

Watch how NASA uses a "weightless" environment to help train astronauts.

Another way to state Einstein's idea is this: suppose we have a spaceship that contains a windowless laboratory equipped with all the tools needed to perform scientific experiments. Now, imagine that an astronomer wakes up after a long night celebrating some scientific breakthrough and finds herself sealed into this laboratory. She has no idea how it happened but notices that she is weightless. This could be because she and the laboratory are far away from any source of gravity, and both are either at rest or moving at some steady speed through space (in which case she has plenty of time to wake up). But it could also be because she and the laboratory are falling freely toward a planet like Earth (in which case she might first want to check her distance from the surface before making coffee).

What Einstein postulated is that there is no experiment she can perform inside the sealed laboratory to determine whether she is floating in space or falling freely in a gravitational field.¹ As far as she is concerned, the two situations are completely *equivalent*. This idea that free fall is indistinguishable from, and hence equivalent to, zero gravity is called the equivalence principle.

Gravity or Acceleration?

Einstein's simple idea has big consequences. Let's begin by considering what happens if two foolhardy people jump from opposite banks into a bottomless chasm (Figure 3). If we ignore air friction, then we can say that while they freely fall, they both accelerate downward at the same rate and feel no external force acting on them. They can throw a ball back and forth, always aiming it straight at each other, as if there were no gravity. The ball falls at the same rate that they do, so it always remains in a line between them. Free Fall.



732 | Chapter 24 Black Holes and Curved Spacetime Section 24.1: Introduction General Relativity

Figure 3. Two people play catch as they descend into a bottomless abyss. Since the people and ball all fall at the same speed, it appears to them that they can play catch by throwing the ball in a straight line between them. Within their frame of

reference, there appears to be no gravity.

Such a game of catch is very different on the surface of Earth. Everyone who grows up feeling gravity knows that a ball, once thrown, falls to the ground. Thus, in order to play catch with someone, you must aim the ball upward so that it follows an arc-rising and then falling as it moves forward-until it is caught at the other end.

Now suppose we isolate our falling people and ball inside a large box that is falling with them. No one inside the box is aware of any gravitational force. If they let go of the ball, it doesn't fall to the bottom of the box or anywhere else but merely stays there or moves in a straight line, depending on whether it is given any motion.

Astronauts in the International Space Station (ISS) that is orbiting Earth live in an environment just like that of the people sealed in a freely falling box (Figure 4). The orbiting ISS is actually "falling" freely around Earth. While in free fall, the astronauts live in a strange world where there seems to be no gravitational force. One can give a wrench a shove, and it moves at constant speed across the orbiting laboratory. A pencil set in midair remains there as if no force were acting on it.

Astronauts aboard the Space Shuttle.



Figure 4. Shane Kimbrough and Sandra Magnus are shown aboard the Endeavour in 2008 with various fruit floating freely. Because the shuttle is in free fall as it orbits Earth, everything–including astronauts–stays put or moves uniformly relative to the walls of the spacecraft. This free-falling state produces a lack of apparent gravity inside the spacecraft. (credit: NASA)

In the "weightless" environment of the International Space Station, moving takes very little effort. Watch astronaut Karen Nyberg demonstrate how she can propel herself with the force of a single human hair.

Appearances are misleading, however. There is a force in this

734 | Chapter 24 Black Holes and Curved Spacetime Section 24.1: Introduction General Relativity situation. Both the ISS and the astronauts continually fall around Earth, pulled by its gravity. But since all fall together–shuttle, astronauts, wrench, and pencil–inside the ISS all gravitational forces appear to be absent.

Thus, the orbiting ISS provides an excellent example of the principle of equivalence-how local effects of gravity can be completely compensated by the right acceleration. To the astronauts, falling around Earth creates the same effects as being far off in space, remote from all gravitational influences.

The Paths of Light and Matter

Einstein postulated that the equivalence principle is a fundamental fact of nature, and that there is *no* experiment inside any spacecraft by which an astronaut can ever distinguish between being weightless in remote space and being in free fall near a planet like Earth. This would apply to experiments done with beams of light as well. But the minute we use light in our experiments, we are led to some very disturbing conclusions—and it is these conclusions that lead us to general relativity and a new view of gravity.

It seems apparent to us, from everyday observations, that beams of light travel in straight lines. Imagine that a spaceship is moving through empty space far from any gravity. Send a laser beam from the back of the ship to the front, and it will travel in a nice straight line and land on the front wall exactly opposite the point from which it left the rear wall. If the equivalence principle really applies universally, then this same experiment performed in free fall around Earth should give us the same result.

Now imagine that the astronauts again shine a beam of light along the length of their ship. But, as shown in Figure 5, this time the orbiting space station falls a bit between the time the light leaves the back wall and the time it hits the front wall. (The amount of the fall is grossly exaggerated in Figure 5 to illustrate the effect.) Therefore, if the beam of light follows a straight line but the ship's path curves downward, then the light should strike the front wall at a point higher than the point from which it left.

Curved Light Path.



Figure 5. In a spaceship moving to the left (in this figure) in its orbit about a planet, light is beamed from the rear, A, toward the front, B. Meanwhile, the ship is falling out of its straight path (exaggerated here). We might therefore expect the light to strike at B', above the target in the ship. Instead, the light follows a curved path and strikes at C. In order for the principle of equivalence to be correct, gravity must be able to curve the path of a light beam just as it curves the path of the spaceship.

However, this would violate the principle of equivalence-the two experiments would give different results. We are thus faced with giving up one of our two assumptions. Either the principle of equivalence is not correct, or light does not always travel in straight lines. Instead of dropping what probably seemed at the time like a ridiculous idea, Einstein worked out what happens if light sometimes does *not* follow a straight path.

Let's suppose the principle of equivalence is right. Then the light beam must arrive directly opposite the point from which it started in the ship. The light, like the ball thrown back and forth, *must fall with the ship* that is in orbit around Earth (see Figure 5). This would make its path curve downward, like the path of the ball, and thus the light would hit the front wall exactly opposite the spot from which it came.

Thinking this over, you might well conclude that it doesn't seem like such a big problem: why *can*'t light fall the way balls do? But, as discussed in Radiation and Spectra, light is profoundly different from balls. Balls have mass, while light does not.

Here is where Einstein's intuition and genius allowed him to make a profound leap. He gave physical meaning to the strange result of our thought experiment. Einstein suggested that the light curves down to meet the front of the shuttle because Earth's gravity actually bends the *fabric of space and time*. This radical idea–which we will explain next–keeps the behavior of light the same in both empty space and free fall, but it changes some of our most basic and cherished ideas about space and time. The reason we take Einstein's suggestion seriously is that, as we will see, experiments now clearly show his intuitive leap was correct.

Key Concepts and Summary

Einstein proposed the equivalence principle as the foundation of the theory of general relativity. According to this principle, there is no way that anyone or any experiment in a sealed environment can distinguish between free fall and the absence of gravity.

Footnotes

• 1 Strictly speaking, this is true only if the laboratory is infinitesimally small. Different locations in a real laboratory that is falling freely due to gravity cannot all be at identical distances from the object(s) responsible for producing the

Chapter 24 Black Holes and Curved Spacetime Section 24.1: Introduction General Relativity | 737 gravitational force. In this case, objects in different locations will experience slightly different accelerations. But this point does not invalidate the principle of equivalence that Einstein derived from this line of thinking.

Glossary

equivalence principle

concept that a gravitational force and a suitable acceleration are indistinguishable within a sufficiently local environment

general theory of relativity

Einstein's theory relating gravity and the structure (geometry) of space and time

Chapter 24 Section 24.2: Spacetime and Gravity

24.2 Spacetime and Gravity

Learning Objectives

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

- Describe Einstein's view of gravity as the warping of spacetime in the presence of massive objects
- Understand that Newton's concept of the gravitational force between two massive objects and Einstein's concept of warped spacetime are different explanations for the same observed accelerations of one massive object in the presence of another massive object

Is light actually bent from its straight-line path by the mass of Earth? How can light, which has no mass, be affected by gravity? Einstein preferred to think that it is *space and time* that are affected by the presence of a large mass; light beams, and everything else that travels through space and time, then find their paths affected. Light always follows the shortest path-but that path may not always be straight. This idea is true for human travel on the curved surface of planet Earth, as well. Say you want to fly from Chicago to Rome. Since an airplane can't go through the solid body of the Earth, the shortest distance is not a straight line but the arc of a *great circle*.

Linkages: Mass, Space, and Time

To show what Einstein's insight really means, let's first consider how we locate an event in space and time. For example, imagine you have to describe to worried school officials the fire that broke out in your room when your roommate tried cooking shish kebabs in the fireplace. You explain that your dorm is at 6400 College Avenue, a street that runs in the left-right direction on a map of your town; you are on the fifth floor, which tells where you are in the up-down direction; and you are the sixth room back from the elevator, which tells where you are in the forward-backward direction. Then you explain that the fire broke out at 6:23 p.m. (but was soon brought under control), which specifies the event in time. Any event in the universe, whether nearby or far away, can be pinpointed using the three dimensions of space and the one dimension of time.

Newton considered space and time to be completely independent, and that continued to be the accepted view until the beginning of the twentieth century. But Einstein showed that there is an intimate connection between space and time, and that only by considering the two together—in what we call spacetime—can we build up a correct picture of the physical world. We examine spacetime a bit more closely in the next subsection.

The gist of Einstein's general theory is that the presence of matter curves or warps the fabric of spacetime. This curving of spacetime is identified with gravity. When something else–a beam of light, an electron, or the starship *Enterprise*–enters such a region of distorted spacetime, its path will be different from what it would have been in the absence of the matter. As American physicist John Wheeler summarized it: "Matter tells spacetime how to curve; spacetime tells matter how to move."

The amount of distortion in spacetime depends on the mass of material that is involved and on how concentrated and compact it is. Terrestrial objects, such as the book you are reading, have far too little mass to introduce any significant distortion. Newton's view of gravity is just fine for building bridges, skyscrapers, or amusement park rides. General relativity does, however, have some practical applications. The GPS (Global Positioning System) in every smartphone can tell you where you are within 5 to 10 meters only because the effects of general and special relativity on the GPS satellites in orbit around the Earth are taken into account.

Unlike a book or your roommate, stars produce measurable distortions in spacetime. A white dwarf, with its stronger surface gravity, produces more distortion just above its surface than does a red giant with the same mass. So, you see, we *are* eventually going to talk about collapsing stars again, but not before discussing Einstein's ideas (and the evidence for them) in more detail.

Spacetime Examples

How can we understand the distortion of spacetime by the presence of some (significant) amount of mass? Let's try the following analogy. You may have seen maps of New York City that squeeze the full three dimensions of this towering metropolis onto a flat sheet of paper and still have enough information so tourists will not get lost. Let's do something similar with diagrams of spacetime.

Figure 1, for example, shows the progress of a motorist driving east on a stretch of road in Kansas where the countryside is absolutely flat. Since our motorist is traveling only in the eastwest direction and the terrain is flat, we can ignore the other two dimensions of space. The amount of time elapsed since he left home is shown on the *y*-axis, and the distance traveled eastward is shown on the *x*-axis. From A to B he drove at a uniform speed; unfortunately, it was too fast a uniform speed and a police car spotted him. From B to C he stopped to receive his ticket and made no progress through space, only through time. From C to D he drove more slowly because the police car was behind him.



Spacetime Diagram.

Figure 1. This diagram shows the progress of a motorist traveling east across the flat Kansas landscape. Distance traveled is plotted along the horizontal axis. The time elapsed since the motorist left the starting point is plotted along the vertical axis. Now let's try illustrating the distortions of spacetime in two
dimensions. In this case, we will (in our imaginations) use a rubber sheet that can stretch or warp if we put objects on it.

Let's imagine stretching our rubber sheet taut on four posts. To complete the analogy, we need something that normally travels in a straight line (as light does). Suppose we have an extremely intelligent ant-a friend of the comic book superhero Ant-Man, perhaps-that has been trained to walk in a straight line.

We begin with just the rubber sheet and the ant, simulating empty space with no mass in it. We put the ant on one side of the sheet and it walks in a beautiful straight line over to the other side (Figure 2). We next put a small grain of sand on the rubber sheet. The sand does distort the sheet a tiny bit, but this is not a distortion that we or the ant can measure. If we send the ant so it goes close to, but not on top of, the sand grain, it has little trouble continuing to walk in a straight line.

Now we grab something with a little more mass—say, a small pebble. It bends or distorts the sheet just a bit around its position. If we send the ant into this region, it finds its path slightly altered by the distortion of the sheet. The distortion is not large, but if we follow the ant's path carefully, we notice it deviating slightly from a straight line.

The effect gets more noticeable as we increase the mass of the object that we put on the sheet. Let's say we now use a massive paperweight. Such a heavy object distorts or warps the rubber sheet very effectively, putting a good sag in it. From our point of view, we can see that the sheet near the paperweight is no longer straight.

Three-Dimensional Analogy for Spacetime.



Figure 2. On a flat rubber sheet, a trained ant has no trouble walking in a straight line. When a massive object creates a big depression in the sheet, the ant, which must walk where the sheet

takes it, finds its path changed (warped) dramatically.

Now let's again send the ant on a journey that takes it close to, but not on top of, the paperweight. Far away from the paperweight, the ant has no trouble doing its walk, which looks straight to us. As it nears the paperweight, however, the ant is forced down into the sag. It must then climb up the other side before it can return to walking on an undistorted part of the sheet. All this while, the ant is following the shortest path it can, but through no fault of its own (after all, ants can't fly, so it has to stay on the sheet) this path is curved by the distortion of the sheet itself.

In the same way, according to Einstein's theory, light always follows the shortest path through spacetime. But the mass associated with large concentrations of matter distorts spacetime, and the shortest, most direct paths are no longer straight lines, but curves.

How large does a mass have to be before we can measure a change in the path followed by light? In 1916, when Einstein first proposed his theory, no distortion had been detected at the surface of Earth (so Earth might have played the role of the grain of sand in our analogy). Something with a mass like our Sun's was necessary to detect the effect Einstein was describing (we will discuss how this effect was measured using the Sun in the next section). The paperweight in our analogy might be a white dwarf or a neutron star. The distortion of spacetime is greater near the surfaces of these compact, massive objects than near the surface of the Sun. And when, to return to the situation described at the beginning of the chapter, a star core with more than three times the mass of the Sun collapses forever, the distortions of spacetime very close to it can become truly mind-boggling.

Key Concepts and Summary

By considering the consequences of the equivalence principle, Einstein concluded that we live in a curved spacetime. The distribution of matter determines the curvature of spacetime; other objects (and even light) entering a region of spacetime must follow its curvature. Light must change its path near a massive object not because light is bent by gravity, but because spacetime is.

Glossary

spacetime

system of one time and three space coordinates, with respect to which the time and place of an event can be specified

Chapter 24 Section 24.3: Tests of General Relativity

24.3 Tests of General Relativity

Learning Objectives

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

- Describe unusual motion of Mercury around the Sun and explain how general relativity explains the observed behavior
- Provide examples of evidence for light rays being bent by massive objects, as predicted by general relativity's theory about the warping of spacetime

What Einstein proposed was nothing less than a major revolution in our understanding of space and time. It was a new theory of gravity, in which mass determines the curvature of spacetime and that curvature, in turn, controls how objects move. Like all new ideas in science, no matter who advances them, Einstein's theory had to be tested by comparing its predictions against the experimental evidence. This was quite a challenge because the effects of the new theory were apparent only when the mass was quite large. (For smaller masses, it required measuring techniques that would not become available until decades later.)

When the distorting mass is small, the predictions of general relativity must agree with those resulting from Newton's law of universal gravitation, which, after all, has served us admirably in our technology and in guiding space probes to the other planets. In familiar territory, therefore, the differences between the predictions of the two models are subtle and difficult to detect. Nevertheless, Einstein was able to demonstrate one proof of his theory that could be found in existing data and to suggest another one that would be tested just a few years later.

The Motion of Mercury

Of the planets in our solar system, Mercury orbits closest to the Sun and is thus most affected by the distortion of spacetime produced by the Sun's mass. Einstein wondered if the distortion might produce a noticeable difference in the motion of Mercury that was not predicted by Newton's law. It turned out that the difference was subtle, but it was definitely there. Most importantly, it had already been measured.

Mercury has a highly elliptical orbit, so that it is only about twothirds as far from the Sun at perihelion as it is at aphelion. (These terms were defined in the chapter on Orbits and Gravity.) The gravitational effects (perturbations) of the other planets on Mercury produce a calculable advance of Mercury's perihelion. What this means is that each successive perihelion occurs in a slightly different direction as seen from the Sun (Figure 1). Mercury's Wobble.



Figure 1. The major axis of the orbit of a planet, such as Mercury, rotates in space slightly because of various perturbations. In Mercury's case, the amount of rotation (or orbital precession) is a bit larger than can be accounted for by the gravitational forces exerted by other planets; this difference is precisely explained by the general theory of relativity. Mercury, being the planet closest to the Sun, has its orbit most affected by the warping of spacetime near the Sun. The change from orbit to orbit has been significantly exaggerated on this diagram.

According to Newtonian gravitation, the gravitational forces exerted by the planets will cause Mercury's perihelion to advance by about 531 seconds of arc (arcsec) per century. In the nineteenth century, however, it was observed that the actual advance is 574 arcsec per century. The discrepancy was first pointed out in 1859 by Urbain Le Verrier, the codiscoverer of Neptune. Just as discrepancies in the motion of Uranus allowed astronomers to discover the presence of Neptune, so it was thought that the discrepancy in the motion of Mercury could mean the presence of an undiscovered inner planet. Astronomers searched for this planet near the Sun, even giving it a name: Vulcan, after the Roman god of fire. (The name would later be used for the home planet of a famous character on a popular television show about future space travel.)

But no planet has ever been found nearer to the Sun than Mercury, and the discrepancy was still bothering astronomers when Einstein was doing his calculations. General relativity, however, predicts that due to the curvature of spacetime around the Sun, the perihelion of Mercury should advance slightly more than is predicted by Newtonian gravity. The result is to make the major axis of Mercury's orbit rotate slowly in space because of the Sun's gravity alone. The prediction of general relativity is that the direction of perihelion should change by an additional 43 arcsec per century. This is remarkably close to the observed discrepancy, and it gave Einstein a lot of confidence as he advanced his theory. The relativistic advance of perihelion was later also observed in the orbits of several asteroids that come close to the Sun.

Deflection of Starlight

Einstein's second test was something that had not been observed before and would thus provide an excellent confirmation of his theory. Since spacetime is more curved in regions where the gravitational field is strong, we would expect light passing very near the Sun to appear to follow a curved path (Figure 2), just like that of the ant in our analogy. Einstein calculated from general relativity theory that starlight just grazing the Sun's surface should be deflected by an angle of 1.75 arcsec. Could such a deflection be observed?

Curvature of Light Paths near the Sun.



Figure 2. Starlight passing near the Sun is deflected slightly by the "warping" of spacetime. (This deflection of starlight is one small example of a phenomenon called gravitational lensing, which we'll

discuss in more detail in The Evolution and Distribution of Galaxies.) Before passing by the Sun, the light from the star was traveling parallel to the bottom edge of the figure. When it passed near the Sun, the path was altered slightly. When we see the light,

we assume the light beam has been traveling in a straight path throughout its journey, and so we measure the position of the star to be slightly different from its true position. If we were to observe the star at another time, when the Sun is not in the way, we would measure its true position.

We encounter a small "technical problem" when we try to photograph starlight coming very close to the Sun: the Sun is an outrageously bright source of starlight itself. But during a total solar eclipse, much of the Sun's light is blocked out, allowing the stars near the Sun to be photographed. In a paper published during World War I, Einstein (writing in a German journal) suggested that photographic observations during an eclipse could reveal the deflection of light passing near the Sun.

The technique involves taking a photograph of the stars six months prior to the eclipse and measuring the position of all the stars accurately. Then the same stars are photographed during the eclipse. This is when the starlight has to travel to us by skirting the Sun and moving through measurably warped spacetime. As seen from Earth, the stars closest to the Sun will seem to be "out of place"-slightly away from their regular positions as measured when the Sun is not nearby.

A single copy of that paper, passed through neutral Holland, reached the British astronomer Arthur S. Eddington, who noted that the next suitable eclipse was on May 29, 1919. The British organized two expeditions to observe it: one on the island of Príncipe, off the coast of West Africa, and the other in Sobral, in northern Brazil. Despite some problems with the weather, both expeditions obtained successful photographs. The stars seen near the Sun were indeed displaced, and to the accuracy of the measurements, which was about 20%, the shifts were consistent with the predictions of general relativity. More modern experiments with radio waves traveling close to the Sun have confirmed that the actual displacements are within 1% of what general relativity predicts.

The confirmation of the theory by the eclipse expeditions in 1919 was a triumph that made Einstein a world celebrity.

Key Concepts and Summary

In weak gravitational fields, the predictions of general relativity agree with the predictions of Newton's law of gravity. However, in the stronger gravity of the Sun, general relativity makes predictions that differ from Newtonian physics and can be tested. For example, general relativity predicts that light or radio waves will be deflected when they pass near the Sun, and that the position where Mercury is at perihelion would change by 43 arcsec per century even if there were no other planets in the solar system to perturb its orbit. These predictions have been verified by observation.

Chapter 24 Section 24.4: Time in General Relativity

24.4 Time in General Relativity

Learning Objectives

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

- Describe how Einsteinian gravity slows clocks and can decrease a light wave's frequency of oscillation
- Recognize that the gravitational decrease in a light wave's frequency is compensated by an increase in the light wave's wavelength—the so-called gravitational redshift—so that the light continues to travel at constant speed

General relativity theory makes various predictions about the behavior of space and time. One of these predictions, put in everyday terms, is that the stronger the gravity, the slower the pace of time. Such a statement goes very much counter to our intuitive sense of time as a flow that we all share. Time has always seemed the most democratic of concepts: all of us, regardless of wealth or status, appear to move together from the cradle to the grave in the great current of time.

But Einstein argued that it only seems this way to us because all humans so far have lived and died in the gravitational environment of Earth. We have had no chance to test the idea that the pace of time might depend on the strength of gravity, because we have not experienced radically different gravities. Moreover, the differences in the flow of time are extremely small until truly large masses are involved. Nevertheless, Einstein's prediction has now been tested, both on Earth and in space.

The Tests of Time

An ingenious experiment in 1959 used the most accurate atomic clock known to compare time measurements on the ground floor and the top floor of the physics building at Harvard University. For a clock, the experimenters used the frequency (the number of cycles per second) of gamma rays emitted by radioactive cobalt. Einstein's theory predicts that such a cobalt clock on the ground floor, being a bit closer to Earth's center of gravity, should run very slightly slower than the same clock on the top floor. This is precisely what the experiments observed. Later, atomic clocks were taken up in highflying aircraft and even on one of the Gemini space flights. In each case, the clocks farther from Earth ran a bit faster. While in 1959 it didn't matter much if the clock at the top of the building ran faster than the clock in the basement, today that effect is highly relevant. Every smartphone or device that synchronizes with a GPS must correct for this (as we will see in the next section) since the clocks on satellites will run faster than clocks on Earth.

The effect is more pronounced if the gravity involved is the Sun's and not Earth's. If stronger gravity slows the pace of time, then it will take longer for a light or radio wave that passes very near the edge of the Sun to reach Earth than we would expect on the basis of Newton's law of gravity. (It takes longer because spacetime is curved in the vicinity of the Sun.) The smaller the distance between the ray of light and the edge of the Sun at closest approach, the longer will be the delay in the arrival time.

In November 1976, when the two Viking spacecraft were operating on the surface of Mars, the planet went behind the Sun as seen from Earth (Figure 1). Scientists had preprogrammed Viking to send a radio wave toward Earth that would go extremely close to the outer regions of the Sun. According to general relativity, there would be a delay because the radio wave would be passing through a region where time ran more slowly. The experiment was able to confirm Einstein's theory to within 0.1%.

Time Delays for Radio Waves near the Sun.



Figure 1. Radio signals from the Viking lander on Mars were delayed when they passed near the Sun, where spacetime is curved relatively strongly. In this picture, spacetime is pictured as a twodimensional rubber sheet.

Gravitational Redshift

What does it mean to say that time runs more slowly? When light emerges from a region of strong gravity where time slows down,

754 | Chapter 24 Section 24.4: Time in General Relativity

the light experiences a change in its frequency and wavelength. To understand what happens, let's recall that a wave of light is a repeating phenomenon-crest follows crest with great regularity. In this sense, each light wave is a little clock, keeping time with its wave cycle. If stronger gravity slows down the pace of time (relative to an outside observer), then the rate at which crest follows crest must be correspondingly slower-that is, the waves become less *frequent*.

To maintain constant light speed (the key postulate in Einstein's theories of special and general relativity), the lower frequency must be compensated by a longer wavelength. This kind of increase in wavelength (when caused by the motion of the source) is what we called a *redshift* in Radiation and Spectra. Here, because it is gravity and not motion that produces the longer wavelengths, we call the effect a gravitational redshift.

The advent of space-age technology made it possible to measure gravitational redshift with very high accuracy. In the mid-1970s, a hydrogen *maser*, a device akin to a laser that produces a microwave radio signal at a particular wavelength, was carried by a rocket to an altitude of 10,000 kilometers. Instruments on the ground were used to compare the frequency of the signal emitted by the rocket-borne maser with that from a similar maser on Earth. The experiment showed that the stronger gravitational field at Earth's surface really did slow the flow of time relative to that measured by the maser in the rocket. The observed effect matched the predictions of general relativity to within a few parts in 100,000.

These are only a few examples of tests that have confirmed the predictions of general relativity. Today, general relativity is accepted as our best description of gravity and is used by astronomers and physicists to understand the behavior of the centers of galaxies, the beginning of the universe, and the subject with which we began this chapter—the death of truly massive stars.

Relativity: A Practical Application

By now you may be asking: why should I be bothered with relativity? Can't I live my life perfectly well without it? The answer is you can't. Every time a pilot lands an airplane or you use a GPS to determine where you are on a drive or hike in the back country, you (or at least your GPS-enabled device) must take the effects of both general and special relativity into account.

GPS relies on an array of 24 satellites orbiting the Earth, and at least 4 of them are visible from any spot on Earth. Each satellite carries a precise atomic clock. Your GPS receiver detects the signals from those satellites that are overhead and calculates your position based on the time that it has taken those signals to reach you. Suppose you want to know where you are within 50 feet (GPS devices can actually do much better than this). Since it takes only 50 billionths of a second for light to travel 50 feet, the clocks on the satellites must be synchronized to at least this accuracy–and relativistic effects must therefore be taken into account.

The clocks on the satellites are orbiting Earth at a speed of 14,000 kilometers per hour and are moving much faster than clocks on the surface of Earth. According to Einstein's theory of relativity, the clocks on the satellites are ticking more slowly than Earth-based clocks by about 7 millionths of a second per day. (We have not discussed the *special* theory of relativity, which deals with changes when objects move very fast, so you'll have to take our word for this part.)

The orbits of the satellites are 20,000 kilometers above Earth, where gravity is about four times weaker than at Earth's surface. General relativity says that the orbiting clocks should tick about 45 millionths of a second faster than they would on Earth. The net effect is that the time on a satellite clock advances by about 38 microseconds per day. If these relativistic effects were not taken into account, navigational errors would start to add up and positions would be off by about 7 miles in only a single day.

Key Concepts and Summary

General relativity predicts that the stronger the gravity, the more slowly time must run. Experiments on Earth and with spacecraft have confirmed this prediction with remarkable accuracy. When light or other radiation emerges from a compact smaller remnant, such as a white dwarf or neutron star, it shows a gravitational redshift due to the slowing of time.

Glossary

gravitational redshift

an increase in wavelength of an electromagnetic wave (light) when propagating from or near a massive object

Chapter 24 Section 24.5: Black Holes

24.5 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 1).

Formation of a Black Hole.



Figure 1. 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 2). 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."

John Wheeler (1911–2008).



Figure 2. 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 3). If the Sun could then shrink just a little more, even that one remaining light beam would no longer be able to escape.

Light Paths near a Massive Object.



Figure 3. 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 4). 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.



Karl Schwarzschild (1873–1916).

Figure 4. 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 10¹² 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.

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:

 $R_s = \frac{2GM}{c^2} = \frac{2(6.67 \times 10^{-11} N \times m^2/kg^2)(4 \times 10^6)(1.99 \times 10^{30} kg)}{(3.00 \times 10^8 m/s)^2} = 1.18 \times 10^{10} m$

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.

Check Your Learning

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:

Substituting the data into our equation gives $R_s = \frac{2GM}{c^2} = \frac{2(6.67 \times 10^{-11}N \times m^2/kg^2)(3000kg)}{(3.00 \times 10^8 m/s)^2} = 1.33 \times 10^{-23}m$ 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.

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.

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 www.astrosociety.org/scifi.

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.

From the previous discussion, you will probably agree that jumping into a black hole is definitely a once-in-alifetime 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.

Key Concepts and Summary

Theory suggests that stars with stellar cores more massive than three times the mass of the Sun at the time they exhaust their nuclear fuel will collapse to become black holes. The surface surrounding a black hole, where the escape velocity equals the speed of light, is called the event horizon, and the radius of the surface is called the Schwarzschild radius. Nothing, not even light, can escape through the event horizon from the black hole. At its center, each black hole is thought to have a singularity, a point of infinite density and zero volume. Matter falling into a black hole appears, as viewed by an outside observer, to freeze in position at the event horizon. However, if we were riding on the infalling matter, we would pass through the event horizon. As we approach the singularity, the tidal forces would tear our bodies apart even before we reach the singularity.

Glossary

black hole

a region in spacetime where gravity is so strong that nothing-not even light-can escape

event horizon

a boundary in spacetime such that events inside the boundary can have no effect on the world outside it—that is, the boundary of the region around a black hole where the curvature of spacetime no longer provides any way out

singularity

the point of zero volume and infinite density to which any object that becomes a black hole must collapse, according to the theory of general relativity

Chapter 24 Section 24.6: Evidence for Black Holes

24.6 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 [link]).

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 1).

Binary Black Hole.



Figure 1. 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 [link]). 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 lists the characteristics of some of the best examples.

Some Black Hole Candidates in Binary Star Systems

	Companion	Orbital	Black Hole
Name/Catalog Designation ¹	Star Spectral	Period	Mass Estimates
	Туре	(days)	(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
GRO J1655-40 (V1033 Sco)	F subgiant	2.6	7
GRS 1915+105	K giant	33.5	14
GS202+1338 (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.

Key Concepts and Summary

The best evidence of stellar-mass black holes comes from binary star systems in which (1) one star of the pair is not visible, (2) the flickering X-ray emission is characteristic of an accretion disk around a compact object, and (3) the orbit and characteristics of the visible star indicate that the mass of its invisible companion is greater than $3 M_{Sun}$. A number of systems with these characteristics have been found. Black holes with masses of millions to billions of solar masses are found in the centers of large galaxies.

Footnotes

• 1 As 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.

Glossary

accretion disk

the disk of gas and dust found orbiting newborn stars, as well as compact stellar remnants such as white dwarfs, neutron stars, and black holes when they are in binary systems and are sufficiently close to their binary companions to draw off material

Chapter 24 Section 24.7: Gravitational Wave Astronomy

24.7 Gravitational Wave Astronomy

Learning Objectives

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

- Describe what a gravitational wave is, what can produce it, and how fast it propagates
- Understand the basic mechanisms used to detect gravitational waves

Another part of Einstein's ideas about gravity can be tested as a way of checking the theory that underlies black holes. According to general relativity, the geometry of spacetime depends on where matter is located. Any rearrangement of matter–say, from a sphere to a sausage shape–creates a disturbance in spacetime. This disturbance is called a gravitational wave, and relativity predicts that it should spread outward at the speed of light. The big problem with trying to study such waves is that they are tremendously

786 | Chapter 24 Section 24.7: Gravitational Wave Astronomy weaker than electromagnetic waves and correspondingly difficult to detect.

Proof from a Pulsar

We've had indirect evidence for some time that gravitational waves exist. In 1974, astronomers Joseph Taylor and Russell Hulse discovered a pulsar (with the designation PSR1913+16) orbiting another neutron star. Pulled by the powerful gravity of its companion, the pulsar is moving at about one-tenth the speed of light in its orbit.

According to general relativity, this system of stellar corpses should be radiating energy in the form of gravitational waves at a high enough rate to cause the pulsar and its companion to spiral closer together. If this is correct, then the orbital period should decrease (according to Kepler's third law) by one ten-millionth of a second per orbit. Continuing observations showed that the period is decreasing by precisely this amount. Such a loss of energy in the system can be due only to the radiation of gravitational waves, thus confirming their existence. Taylor and Hulse shared the 1993 Nobel Prize in physics for this work.

Direct Observations

Although such an indirect proof convinced physicists that gravitational waves exist, it is even more satisfying to detect the waves directly. What we need are phenomena that are powerful enough to produce gravitational waves with amplitudes large enough that we can measure them. Theoretical calculations suggest some of the most likely events that would give a burst of gravitational waves strong enough that our equipment on Earth could measure it:

- the coalescence of two neutron stars in a binary system that spiral together until they merge
- the swallowing of a neutron star by a black hole
- the coalescence (merger) of two black holes
- the implosion of a really massive star to form a neutron star or a black hole
- the first "shudder" when space and time came into existence and the universe began

For the last four decades, scientists have been developing an audacious experiment to try to detect gravitational waves from a source on this list. The US experiment, which was built with collaborators from the UK, Germany, Australia and other countries, named LIGO (Laser Interferometer Gravitational-Wave is Observatory). LIGO currently has two observing stations, one in Louisiana and the other in the state of Washington. The effects of gravitational waves are so small that confirmation of their detection will require simultaneous measurements by two widely separated facilities. Local events that might cause small motions within the observing stations and mimic gravitational waves-such as small earthquakes, ocean tides, and even traffic-should affect the two sites differently.

Each of the LIGO stations consists of two 4-kilometer-long, 1.2-meter-diameter vacuum pipes arranged in an L-shape. A test mass with a mirror on it is suspended by wire at each of the four ends of the pipes. Ultra-stable laser light is reflected from the mirrors and travels back and forth along the vacuum pipes (Figure 1). If gravitational waves pass through the LIGO instrument, then, according to Einstein's theory, the waves will affect local spacetime-they will alternately stretch and shrink the distance the laser light must travel between the mirrors ever so slightly. When one arm of the instrument gets longer, the other will get shorter, and vice versa.

Gravitational Wave Telescope.



Figure 1. An aerial view of the LIGO facility at Livingston, Louisiana. Extending to the upper left and far right of the image are the 4-kilometer-long detectors. (credit: modification of work by

Caltech/MIT/LIGO Laboratory)

The challenge of this experiment lies in that phrase "ever so slightly." In fact, to detect a gravitational wave, the change in the distance to the mirror must be measured with an accuracy of *one ten-thousandth the diameter of a proton*. In 1972, Rainer Weiss of MIT wrote a paper suggesting how this seemingly impossible task might be accomplished.

A great deal of new technology had to be developed, and work on the laboratory, with funding from the National Science Foundation, began in 1979. A full-scale prototype to demonstrate the technology was built and operated from 2002 to 2010, but the prototype was not expected to have the sensitivity required to actually detect gravitational waves from an astronomical source. Advanced LIGO, built to be more precise with the improved technology developed in the prototype, went into operation in 2015–and almost immediately detected gravitational waves.

What LIGO found was gravitational waves produced in the final fraction of a second of the merger of two black holes (Figure 2). The black holes had masses of 20 and 36 times the mass of the Sun, and the merger took place 1.3 billion years ago—the gravitational waves occurred so far away that it has taken that long for them, traveling at the speed of light, to reach us.

In the cataclysm of the merger, about three times the mass of the Sun was converted to energy (recall $E = mc^2$). During the tiny fraction of a second for the merger to take place, this event produced power about 10 times the power produced by all the stars in the entire visible universe–but the power was all in the form of gravitational waves and hence was invisible to our instruments, except to LIGO. The event was recorded in Louisiana about 7 milliseconds before the detection in Washington–just the right distance given the speed at which gravitational waves travel–and indicates that the source was located somewhere in the southern hemisphere sky. Unfortunately, the merger of two black holes is not expected to produce any light, so this is the only observation we have of the event.

Signal Produced by a Gravitational Wave.





(b)

Figure 2. (a) The top panel shows the signal measured at Hanford, Washington; the middle panel shows the signal measured at Livingston, Louisiana. The smoother thin curve in each panel shows the predicted signal, based on Einstein's general theory of relativity, produced by the merger of two black holes. The bottom panel shows a superposition of the waves detected at the two LIGO observatories. Note the remarkable agreement of the two independent observations and of the observations with theory. (b) The painting shows an artist's impression of two massive black holes spiraling inward toward an eventual merger. (credit a, b: modification of work by SXS)

This detection by LIGO (and another one of a different black hole merger a few months later) opens a whole new window on the universe. One of the experimenters compared the beginning of gravitational wave astronomy to the era when silent films were replaced by movies with sound (comparing the vibration of spacetime during the passing of a gravitational wave to the vibrations that sound makes). By the end of 2017, LIGO had detected four more mergers of black holes. Two of these, like the initial discovery, involved mergers of black holes with a range of masses that have been observed only by gravitational waves. In one merger, black holes with masses of 31 and 25 times the mass of the Sun merged to form a spinning black hole with a mass of about 53 times the mass the Sun. This event was detected not only by the two LIGO detectors, but also by a newly operational European gravitational wave observatory, Virgo. The other event was caused by the merger of 20- and 30-solarmass black holes, and resulted in a 49-solar-mass black hole. Astronomers are not yet sure just how black holes in this mass range form.

Two other mergers detected by LIGO involved black holes with stellar masses comparable to those of black holes in X-ray binary systems. In one case, the merging black holes had masses of 14 and 8 times the mass of the Sun. The other event, again detected by both LIGO and Virgo, was produced by a merger of black holes with masses of 7 and 12 times the mass of the Sun. None of the mergers of black holes was detected in any other way besides gravitational waves. It is quite likely that the merger of black holes does not produce any electromagnetic radiation.

In late 2017, data from all three gravitational wave observatories was used to locate the position in the sky of a fifth event, which was produced by the merger of objects with masses of 1.1 to 1.6 times the mass of the Sun. This is the mass range for neutron stars (see The Milky Way Galaxy), so in this case, what was observed was the spiraling together of two neutron stars. Data obtained from all three observatories enabled scientists to narrow down the area in the sky where the event occurred. The Fermi satellite offered a fourth set of observational data, detecting a flash of gamma rays at the same time, which confirms the long-standing hypothesis that mergers of neutron stars are progenitors of short gamma-ray bursts (see The Mystery of Gamma-Ray Bursts). The *Swift* satellite also detected a flash of ultraviolet light at the same time, and in the same part of the

sky. This was the first time that a gravitational wave event had been detected with any kind of electromagnetic wave.

combined observations The from LIGO, Virgo, Fermi, and Swift showed that this source was located in NGC 4993, a galaxy at a distance of about 130 million light-years in the direction of the constellation Hydra. With a well-defined position, ground-based observatories could point their telescopes directly at the source and obtain its spectrum. These observations showed that the merger ejected material with a mass of about 6 percent of the mass of the Sun, and a speed of one-tenth the speed of light. This material is rich in heavy elements, just as the theory of kilonovas (see Short-Duration Gamma-Ray Bursts: Colliding Stellar Corpses) predicted. First estimates suggest that the merger produced about 200 Earth masses of gold, and around 500 Earth masses of platinum. This makes clear that neutron star mergers are a significant source of heavy elements. As additional detections of such events improve theoretical estimates of the frequency at which neutron star mergers occur, it may well turn out that the vast majority of heavy elements have been created in such cataclysms.

The detection of gravitational waves opens a whole new window to the universe. One of the experimenters compared the beginning of gravitational wave astronomy to the era when silent films were replaced by movies with sound (comparing the vibration of spacetime during the passing of a gravitational wave to the vibrations that sound makes). We can now learn about events, such as the merger of black holes, that can be studied in no other way.

Observing the merger of black holes via gravitational waves also means that we can now test Einstein's general theory of relativity where its effects are very strong-close to black holes-and not weak, as they are near Earth. One remarkable result from these detections is that the signals measured so closely match the theoretical predictions made using Einstein's theory. Once again, Einstein's revolutionary idea is found to be the correct description of nature.

Because of the scientific significance of the observations of

gravitational waves, three of the LIGO project leaders–Rainer Weiss of MIT, and Kip Thorne and Barry Barish of Caltech–were awarded the Nobel Prize in 2017.

Several facilities similar to LIGO and Virgo are under construction in other countries to contribute to gravitational wave astronomy and help us pinpoint more precisely pinpoint the location of signals we detect in the sky. The European Space Agency (ESA) is exploring the possibility of building an even larger detector for gravitational waves in space. The goal is to launch a facility called eLISA sometime in the mid 2030s. The design calls for three detector arms, each a million kilometers in length, for the laser light to travel in space. This facility could detect the merger of distant supermassive black holes, which might have occurred when the first generation of stars formed only a few hundred million years after the Big Bang.

In December 2015, ESA launched LISA Pathfinder and successfully tested the technology required to hold two gold-platinum cubes in a state of weightless, perfect rest, relative to one another. While LISA Pathfinder cannot detect gravitational waves, such stability is required if eLISA is to be able to detect the small changes in path length produced by passing gravitational waves.

We should end by acknowledging that the ideas discussed in this chapter may seem strange and overwhelming, especially the first time you read them. The consequences of the general theory of relatively take some getting used to. But they make the universe more bizarre–and interesting–than you probably thought before you took this course.

Key Concepts and Summary

General relativity predicts that the rearrangement of matter in space should produce gravitational waves. The existence of such waves was first confirmed in observations of a pulsar in orbit around another neutron star whose orbits were spiraling closer and losing energy in the form of gravitational waves. In 2015, LIGO found gravitational waves directly by detecting the signal produced by the merger of two stellar-mass black holes, opening a new window on the universe.

For Further Exploration

Articles

Black Holes

Charles, P. & Wagner, R. "Black Holes in Binary Stars: Weighing the Evidence." Sky & Telescope (May 1996): 38. Excellent review of how we find stellar-mass black holes.

Gezari, S. "Star-Shredding Black Holes." Sky & Telescope (June 2013): 16. When black holes and stars collide.

Jayawardhana, R. "Beyond Black." Astronomy (June 2002): 28. On finding evidence of the existence of event horizons and thus black holes.

Nadis, S. "Black Holes: Seeing the Unseeable." Astronomy (April 2007): 26. A brief history of the black hole idea and an introduction to potential new ways to observe them.

Psallis, D. & Sheperd, D. "The Black Hole Test." Scientific American (September 2015): 74–79. The Event Horizon Telescope (a network of radio telescopes) will test some of the stranger predictions of general relativity for the regions near black holes. The September 2015 issue of *Scientific American* was devoted to a celebration of the 100th anniversary of the general theory of relativity.

Rees, M. "To the Edge of Space and Time." Astronomy (July 1998): 48. Good, quick overview.

Talcott, R. "Black Holes in our Backyard." Astronomy (September

2012): 44. Discussion of different kinds of black holes in the Milky Way and the 19 objects known to be black holes.

Gravitational Waves

Bartusiak, M. "Catch a Gravity Wave." Astronomy (October 2000): 54.

Gibbs, W. "Ripples in Spacetime." Scientific American (April 2002): 62.

Haynes, K., & Betz, E. "A Wrinkle in Spacetime Confirms Einstein's Gravitation." Astronomy (May 2016): 22. On the direct detection of gravity waves.

Sanders, G., and Beckett, D. "LIGO: An Antenna Tuned to the Songs of Gravity." Sky & Telescope (October 2000): 41.

Websites

Black Holes

Black Hole Encyclopedia: http://blackholes.stardate.org. From StarDate at the University of Texas McDonald Observatory.

Black Holes: http://science.nasa.gov/astrophysics/focus-areas/ black-holes. NASA overview of black holes, along with links to the most recent news and discoveries.

Black Holes FAQ: http://cfpa.berkeley.edu/Education/ BHfaq.html. Frequently asked questions about black holes, answered by Ted Bunn of UC–Berkeley's Center for Particle Astrophysics.

Black Holes: Gravity's Relentless Pull: http://hubblesite.org/ explore_astronomy/black_holes/home.html. The Hubble Space Telescope's Journey to a Black Hole and Black Hole Encyclopedia (a good introduction for beginners).

Introduction to Black Holes: http://www.damtp.cam.ac.uk/

research/gr/public/bh_intro.html. The Cambridge University Relativity Group's pages on black holes and related calculations.

March 1918: Testing Einstein: http://www.nature.com/nature/ podcast/index-pastcast-2014-03-20.html. Nature Podcast about the 1919 eclipse expedition that proved Einstein's General Theory of Relativity.

MoviesfromtheEdgeofSpacetime: http://archive.ncsa.illinois.edu/Cyberia/NumRel/MoviesEdge.html. Physicists simulate the behavior of various blackholes.

Virtual Trips into Black Holes and Neutron Stars: http://antwrp.gsfc.nasa.gov/htmltest/rjn_bht.html. By Robert Nemiroff at Michigan Technological University.

Gravitational Waves

Advanced LIGO: https://www.advancedligo.mit.edu. The full story on this gravitational wave observatory.

eLISA: https://www.elisascience.org.

Gravitational Waves Detected, Confirming Einstein's Theory: http://www.nytimes.com/2016/02/12/science/ligo-gravitational-waves-black-holes-einstein.html. New York Times article and videos on the discovery of gravitational waves.

Gravitational Waves Discovered from Colliding Black Holes: http://www.scientificamerican.com/article/gravitationalwaves-discovered-from-colliding-black-holes1. *Scientific American* coverage of the discovery of gravitational waves (note the additional materials available in the menu at the right).

LIGO Caltech: https://www.ligo.caltech.edu.

Videos

Black Holes

Black Holes: The End of Time or a New Beginning?: https://www.youtube.com/watch?v=mgtJRsdKe6Q. 2012 Silicon Valley Astronomy Lecture by Roger Blandford (1:29:52).

Death by Black Hole: http://www.openculture.com/2009/02/ death_by_black_hole_and_its_kind_of_funny.htm. Neil deGrasse Tyson explains spaghettification with only his hands (5:34).

HeartsofDarkness:BlackHolesinSpace: https://www.youtube.com/watch?v=4tiAOldypLk.2010Silicon Valley Astronomy Lecture by Alex Filippenko (1:56:11).

Gravitational Waves

Journey of a Gravitational Wave: https://www.youtube.com/ watch?v=FlDtXIBrAYE. Introduction from LIGO Caltech (2:55).

LIGO's First Detection of Gravitational Waves: https://www.youtube.com/watch?v=gw-i_VKd6Wo. Explanation and animations from PBS Digital Studio (9:31).

Two Black Holes Merge into One: https://www.youtube.com/ watch?v=I_88S8DWbcU. Simulation from LIGO Caltech (0:35).

What the Discovery of Gravitational Waves Means: https://www.youtube.com/watch?v=jMVAgCPYYHY. TED Talk by Allan Adams (10:58).

Collaborative Group Activities

A. A computer science major takes an astronomy course like the one you are taking and becomes fascinated with black holes.

Later in life, he founds his own internet company and becomes very wealthy when it goes public. He sets up a foundation to support the search for black holes in our Galaxy. Your group is the allocation committee of this foundation. How would you distribute money each year to increase the chances that more black holes will be found?

- B. Suppose for a minute that stars evolve *without* losing any mass at any stage of their lives. Your group is given a list of binary star systems. Each binary contains one main-sequence star and one invisible companion. The spectral types of the mainsequence stars range from spectral type O to M. Your job is to determine whether any of the invisible companions might be black holes. Which ones are worth observing? Why? (Hint: Remember that in a binary star system, the two stars form at the same time, but the pace of their evolution depends on the mass of each star.)
- C. You live in the far future, and the members of your group have been convicted (falsely) of high treason. The method of execution is to send everyone into a black hole, but you get to pick which one. Since you are doomed to die, you would at least like to see what the inside of a black hole is like–even if you can't tell anyone outside about it. Would you choose a black hole with a mass equal to that of Jupiter or one with a mass equal to that of an entire galaxy? Why? What would happen to you as you approached the event horizon in each case? (Hint: Consider the difference in force on your feet and your head as you cross over the event horizon.)
- D. General relativity is one of the areas of modern astrophysics where we can clearly see the frontiers of human knowledge. We have begun to learn about black holes and warped spacetime recently and are humbled by how much we still don't know. Research in this field is supported mostly by grants from government agencies. Have your group discuss what reasons there are for our tax dollars to support such "far out" (seemingly impractical) work. Can you make a list of "far out"

areas of research in past centuries that later led to practical applications? What if general relativity does not have many practical applications? Do you think a small part of society's funds should still go to exploring theories about the nature of space and time?

- E. Once you all have read this chapter, work with your group to come up with a plot for a science fiction story that uses the properties of black holes.
- F. Black holes seem to be fascinating not just to astronomers but to the public, and they have become part of popular culture. Searching online, have group members research examples of black holes in music, advertising, cartoons, and the movies, and then make a presentation to share the examples you found with the whole class.
- G. As mentioned in the Gravity and Time Machines feature box in this chapter, the film *Interstellar* has a lot of black hole science in its plot and scenery. That's because astrophysicist Kip Thorne at Caltech had a big hand in writing the initial treatment for the movie, and later producing it. Get your group members together (be sure you have popcorn) for a viewing of the movie and then try to use your knowledge of black holes from this chapter to explain the plot. (Note that the film also uses the concept of a *wormhole*, which we don't discuss in this chapter. A wormhole is a theoretically possible way to use a large, spinning black hole to find a way to travel from one place in the universe to another without having to go through regular spacetime to get there.)

Review Questions

How does the equivalence principle lead us to suspect that spacetime might be curved?

If general relativity offers the best description of what happens in the presence of gravity, why do physicists still make use of Newton's equations in describing gravitational forces on Earth (when building a bridge, for example)?

Einstein's general theory of relativity made or allowed us to make predictions about the outcome of several experiments that had not yet been carried out at the time the theory was first published. Describe three experiments that verified the predictions of the theory after Einstein proposed it.

If a black hole itself emits no radiation, what evidence do astronomers and physicists today have that the theory of black holes is correct?

What characteristics must a binary star have to be a good candidate for a black hole? Why is each of these characteristics important?

A student becomes so excited by the whole idea of black holes that he decides to jump into one. It has a mass 10 times the mass of our Sun. What is the trip like for him? What is it like for the rest of the class, watching from afar?

What is an event horizon? Does our Sun have an event horizon around it?

What is a gravitational wave and why was it so hard to detect?

What are some strong sources of gravitational waves that astronomers hope to detect in the future?

Suppose the amount of mass in a black hole doubles. Does the event horizon change? If so, how does it change?

Thought Questions

Imagine that you have built a large room around the people in [link] and that this room is falling at exactly the same rate as they are. Galileo showed that if there is no air friction, light and heavy objects that are dropping due to gravity will fall at the same rate. Suppose that this were not true and that instead heavy objects fall faster. Also suppose that the man in [link] is twice as massive as the woman. What would happen? Would this violate the equivalence principle?

A monkey hanging from a tree branch sees a hunter aiming a rifle directly at him. The monkey then sees a flash and knows that the rifle has been fired. Reacting quickly, the monkey lets go of the branch and drops so that the bullet can pass harmlessly over his head. Does this act save the monkey's life? Why or why not? (Hint: Consider the similarities between this situation and that of Exercise.)

Why would we not expect to detect X-rays from a disk of matter about an ordinary star?

Look elsewhere in this book for necessary data, and indicate what the final stage of evolution–white dwarf, neutron star, or black hole–will be for each of these kinds of stars.

- A. Spectral type-O main-sequence star
- B. Spectral type-B main-sequence star
- C. Spectral type-A main-sequence star
- D. Spectral type-G main-sequence star
- E. Spectral type-M main-sequence star

Which is likely to be more common in our Galaxy: white dwarfs or black holes? Why?

If the Sun could suddenly collapse to a black hole, how would the period of Earth's revolution about it differ from what it is now?

Suppose the people in [link] are in an elevator moving upward with an acceleration equal to *g*, but in the opposite direction. The woman throws the ball to the man with a horizontal force. What happens to the ball?

You arrange to meet a friend at 5:00 p.m. on Valentine's Day on the observation deck of the Empire State Building in New York City. You arrive right on time, but your friend is not there. She arrives 5 minutes late and says the reason is that time runs faster at the top of a tall building, so she is on time but you were early. Is your friend right? Does time run slower or faster at the top of a building, as compared with its base? Is this a reasonable excuse for your friend arriving 5 minutes late?

You are standing on a scale in an elevator when the cable snaps, sending the elevator car into free fall. Before the automatic brakes stop your fall, you glance at the scale reading. Does the scale show your real weight? An apparent weight? Something else?

Figuring for Yourself

Look up G, c, and the mass of the Sun in Appendix E and calculate the radius of a black hole that has the same mass as the Sun. (Note that this is only a theoretical calculation. The Sun does not have enough mass to become a black hole.)

Suppose you wanted to know the size of black holes with masses that are larger or smaller than the Sun. You could go through all the steps in Exercise, wrestling with a lot of large numbers with large exponents. You could be clever, however, and evaluate all the constants in the equation once and then simply vary the mass. You could even express the mass in terms of the Sun's mass and make future calculations really easy. Show that the event horizon equation is equivalent to saying that the radius of the event horizon is equal to 3 km times the mass of the black hole in units of the Sun's mass.

Use the result from Exercise to calculate the radius of a black hole with a mass equal to: the Earth, a B0-type main-sequence star, a globular cluster, and the Milky Way Galaxy. Look elsewhere in this text and the appendixes for tables that provide data on the mass of these four objects.

Since the force of gravity a significant distance away from the event horizon of a black hole is the same as that of an ordinary object of the same mass, Kepler's third law is valid. Suppose that Earth collapsed to the size of a golf ball. What would be the period of revolution of the Moon, orbiting at its current distance of 400,000 km? Use Kepler's third law to calculate the period of revolution of a spacecraft orbiting at a distance of 6000 km.

Glossary

gravitational wave

a disturbance in the curvature of spacetime caused by changes in how matter is distributed; gravitational waves propagate at (or near) the speed of light.

Chapter 25 The Milky Way Galaxy Section 25.1: The Architecture of the Galaxy

Milky Way Galaxy.



Figure 1. The Milky Way rises over Square Tower, an ancestral pueblo building at Hovenweep National Monument in Utah. Many stars and dark clouds of dust combine to make a spectacular celestial sight of our home Galaxy. The location has been designated an International Dark Sky Park by the International Dark Sky Association.

Today, we know that our Sun is just one of the many billions of stars that make up the huge cosmic island we call the Milky Way Galaxy. How can we "weigh" such an enormous system of stars and measure its total mass?

One of the most striking features you can see in a truly dark sky-one without light pollution-is the band of faint white light called the Milky Way, which stretches from one horizon to the other. The name comes from an ancient Greek legend that compared its faint white splash of light to a stream of spilled milk. But folktales differ from culture to culture: one East African tribe thought of the hazy band as the smoke of ancient campfires, several Native American stories tell of a path across the sky traveled by sacred animals, and in Siberia, the diffuse arc was known as the seam of the tent of the sky.

In 1610, Galileo made the first telescopic survey of the Milky Way and discovered that it is composed of a multitude of individual stars. Today, we know that the Milky Way comprises our view inward of the huge cosmic pinwheel that we call the Milky Way Galaxy and that is our home. Moreover, our Galaxy is now recognized as just one galaxy among many billions of other galaxies in the cosmos.

25.1 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 1) 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 2).

William Herschel (1738–1822) and Caroline Herschel (1750–1848).



Figure 1. 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

808 | Chapter 25 The Milky Way Galaxy Section 25.1: The Architecture of the Galaxy

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.

Herschel's Diagram of the Milky Way.



Figure 2. 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.

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 3).

Harlow Shapley and His Diagram of the Milky Way.



Figure 3. (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.

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 4 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.
Schematic Representation of the Galaxy.



Figure 4. 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 4 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 5 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.

Unbarred and Barred Spiral Galaxies.



Figure 5. (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 galaxyNGC 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 3000 light-years above and below the midplane of the thin disk and contains only about 5% as much mass as the thin disk.

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 6).

Inner Part of the Milky Way Galaxy.



Figure 6. 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

Chapter 25 The Milky Way Galaxy Section 25.1: The Architecture of the Galaxy \mid 817

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, with an illustration in Figure 7. 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.

Property	Thin Disk	Thick Disk	Stellar Halo (Excludes Dark Matter)
Stellar mass	$^{4 imes}_{10^{10}}$ M _{Sun}	A few percent of the thin disk mass	$10^{10}\ M_{Sun}$
Luminosity	$3\times 10^{10} \ L_{Sun}$	A few percent of the thin disk luminosity	$8\times 10^8 \ L_{Sun}$
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

Characteristics of the Milky Way Galaxy

Major Parts of the Milky Way Galaxy.



Figure 7. This schematic shows the major components of our Galaxy.

Establishing this overall picture of the Galaxy from our dustshrouded 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, spiralshaped 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.

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 8).

The dynastic Chinese regarded the Milky Way as a "silver river" that was made to separate two starcrossed 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 8). 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.

The Milky Way in Myth.



Figure 8. (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

822 | Chapter 25 The Milky Way Galaxy Section 25.1: The Architecture of the Galaxy

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.

Key Concepts and Summary

The Milky Way Galaxy consists of a thin disk containing dust, gas, and young and old stars; a spherical halo containing populations of very old stars, including RR Lyrae variable stars and globular star clusters; a thick, more diffuse disk with stars that have properties intermediate between those in the thin disk and the halo; a peanutshaped nuclear bulge of mostly old stars around the center; and a supermassive black hole at the very center. The Sun is located roughly halfway out of the Milky Way, about 26,000 light-years from the center.

Glossary

dark matter halo

the mass in the Milky Way that extends well beyond the boundary of the luminous stars to a distance of at least 200,000 light-years from the center of the Galaxy; although we deduce its existence from its gravity, the composition of this matter remains a mystery

halo

the outermost extent of our Galaxy (or another galaxy), containing a sparse distribution of stars and globular clusters in a more or less spherical distribution

Milky Way Galaxy

the band of light encircling the sky, which is due to the many stars and diffuse nebulae lying near the plane of the Milky Way Galaxy

central bulge

(or nuclear bulge) the central (round) part of the Milky Way or a similar galaxy

For Further Exploration

Collaborative Group Activities

- A. You are captured by space aliens, who take you inside a complex cloud of interstellar gas, dust, and a few newly formed stars. To escape, you need to make a map of the cloud. Luckily, the aliens have a complete astronomical observatory with equipment for measuring all the bands of the electromagnetic spectrum. Using what you have learned in this chapter, have your group discuss what kinds of maps you would make of the cloud to plot your most effective escape route.
- B. The diagram that Herschel made of the Milky Way has a very

824 | Chapter 25 The Milky Way Galaxy Section 25.1: The Architecture of the Galaxy

irregular outer boundary (see [link]). Can your group think of a reason for this? How did Herschel construct his map?

- C. Suppose that for your final exam in this course, your group is assigned telescope time to observe a star selected for you by your professor. The professor tells you the position of the star in the sky (its right ascension and declination) but nothing else. You can make any observations you wish. How would you go about determining whether the star is a member of population I or population II?
- D. The existence of dark matter comes as a great surprise, and its nature remains a mystery today. Someday astronomers will know a lot more about it (you can learn more about current findings in The Evolution and Distribution of Galaxies). Can your group make a list of earlier astronomical observations that began as a surprise and mystery, but wound up (with more observations) as well-understood parts of introductory textbooks?
- E. Physicist Gregory Benford has written a series of science fiction novels that take place near the center of the Milky Way Galaxy in the far future. Suppose your group were writing such a story. Make a list of ways that the environment near the galactic center differs from the environment in the "galactic suburbs," where the Sun is located. Would life as we know it have an easier or harder time surviving on planets that orbit stars near the center (and why)?
- F. These days, in most urban areas, city lights completely swamp the faint light of the Milky Way in our skies. Have each member of your group survey 5 to 10 friends or relatives (you could spread out on campus to investigate or use social media or the phone), explaining what the Milky Way is and then asking if they have seen it. Also ask their age. Report back to your group and discuss your reactions to the survey. Is there any relationship between a person's age and whether they have seen the Milky Way? How important is it that many kids growing up on Earth today never (or rarely) get to see our

home Galaxy in the sky?

Chapter 26 Galaxies Section 26.1: The Discovery of Galaxies

Thinking Ahead

Spiral Galaxy.



Figure 1. NGC 6946 is a spiral galaxy also known as the "Fireworks galaxy." It is at a distance of about 18 million light-years, in the direction of the constellations Cepheus and Cygnus. It was discovered by William Herschel in 1798. This galaxy is about one-third the size of the Milky Way. Note on the left how the colors of the galaxy change from the yellowish light of old stars in the center

to the blue color of hot, young stars and the reddish glow of hydrogen clouds in the spiral arms. As the image shows, this galaxy is rich in dust and gas, and new stars are still being born here. In the right-hand image, the x-rays coming from this galaxy are shown in purple, which has been added to other colors showing visible light. (Credit left: modification of work by NASA, ESA, STScI,

R. Gendler, and the Subaru Telescope (NAOJ); credit right: modification of work by X-ray: NASA/CXC/MSSL/R.Soria et al, Optical: AURA/Gemini OBs)

In the last chapter, we explored our own Galaxy. But is it the only one? If there are others, are they like the Milky Way? How far away are they? Can we see them? As we shall learn, some galaxies turn out to be so far away that it has taken billions of years for their light to reach us. These remote galaxies can tell us what the universe was like when it was young.

In this chapter, we start our exploration of the vast realm of galaxies. Like tourists from a small town making their first visit to the great cities of the world, we will be awed by the beauty and variety of the galaxies. And yet, we will recognize that much of what we see is not so different from our experiences at home, and we will be impressed by how much we can learn by looking at structures built long ago.

We begin our voyage with a guide to the properties of galaxies, much as a tourist begins with a guidebook to the main features of the cities on the itinerary. In later chapters, we will look more carefully at the past history of galaxies, how they have changed over time, and how they acquired their many different forms. First, we'll begin our voyage through the galaxies with the question: is our Galaxy the only one?

26.1 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-anddust 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 1). 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.

Andromeda Galaxy.



Figure 1. 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.

EDWIN HUBBLE: EXPANDING THE UNIVERSE

The son of a Missouri insurance agent, Edwin Hubble (Figure 2) 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.

Edwin Hubble (1889–1953).

Figure 2. 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 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).

Key Concepts and Summary

Faint star clusters, clouds of glowing gas, and galaxies all appeared as faint patches of light (or nebulae) in the telescopes available at the beginning of the twentieth century. It was only when Hubble measured the distance to the Andromeda galaxy using cepheid variables with the giant 2.5-meter reflector on Mount Wilson in 1924 that the existence of other galaxies similar to the Milky Way in size and content was established.

Chapter 26 Section 26.2: Types of Galaxies

26.2 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 [link]). 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 1).

Spiral Galaxies.



Figure 1. (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 2). Showing great originality, astronomers call these galaxies barred spirals.



Barred Spiral Galaxy.



Figure 2. 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 [link]). 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 3, along with elliptical galaxies for comparison.

Hubble Classification of Galaxies.



Figure 3. 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 4). 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.

Elliptical Galaxies.



Figure 4. (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 4) 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 5. 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.

Dwarf Elliptical Galaxy.



Figure 5. 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 6), 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-theworld 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.

4-Meter Telescope at Cerro Tololo Inter-American Observatory Silhouetted against the Southern Sky.



Figure 6. 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,

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 3 (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.

Key Concepts and Summary

The majority of bright galaxies are either spirals or ellipticals. Spiral galaxies contain both old and young stars, as well as interstellar matter, and have typical masses in the range of 10^9 to 10^{12} M_{Sun}. Our own Galaxy is a large spiral. Ellipticals are spheroidal or slightly elongated systems that consist almost entirely of old stars, with very little interstellar matter. Elliptical galaxies range in size from giants, more massive than any spiral, down to dwarfs, with masses of only about 10^6 M_{Sun}. Dwarf ellipticals are probably the most common type of galaxy in the nearby universe. A small percentage of galaxies may change their appearance over time due to collisions with other galaxies or by a change in the rate of star formation.

Glossary

elliptical galaxy

a galaxy whose shape is an ellipse and that contains no conspicuous interstellar material

irregular galaxy

a galaxy without any clear symmetry or pattern; neither a spiral nor an elliptical galaxy

spiral galaxy

a flattened, rotating galaxy with pinwheel-like arms of interstellar material and young stars, winding out from its central bulge

Chapter 26 Section 26.3: Properties of Galaxies

26.3 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-tolight 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 $4 \times 10^{11} M_{Sun}$ (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 $1.4 \times 10^{12} M_{Sun}$. The mass of the **Milky Way Galaxy** is estimated to be $8.5 \times 10^{11} M_{Sun}$, 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 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.

Characteristics of the Different Types of Galaxies

Characteristic	Spirals	Ellipticals
Mass (M _{Sun})	10 ⁹ to 10 ¹²	10 ⁵ to 10 ¹³
Diameter (thousands of light-years)	15 to 150	3 to >700
Luminosity (L _{Sun})	10 ⁸ to 10 ¹¹	10 ⁶ to 10 ¹¹
Populations of stars	Old and young	Old
Interstellar matter	Gas and dust	Almost no dust; little
Mass-to-light ratio in the visible part	2 to 10	10 to 20
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 [link]). 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-tolight ratios of 10 to 20.

But these figures refer only to the inner, conspicuous parts of galaxies (Figure 1). 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.

M101, the Pinwheel Galaxy.



Figure 1. 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.

Key Concepts and Summary

The masses of spiral galaxies are determined from measurements of their rates of rotation. The masses of elliptical galaxies are estimated from analyses of the motions of the stars within them. Galaxies can be characterized by their mass-to-light ratios. The luminous parts of galaxies with active star formation typically have mass-to-light ratios in the range of 1 to 10; the luminous parts of elliptical galaxies, which contain only old stars, typically have massto-light ratios of 10 to 20. The mass-to-light ratios of whole galaxies, including their outer regions, are as high as 100, indicating the presence of a great deal of dark matter.

Glossary

mass-to-light ratio

the ratio of the total mass of a galaxy to its total luminosity, usually expressed in units of solar mass and solar luminosity; the mass-to-light ratio gives a rough indication of the types of stars contained within a galaxy and whether or not substantial quantities of dark matter are present

Chapter 26 Section 26.4: The Extragalactic Distance Scale

26.4 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 1).

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.

Cepheid Variable Star.



Figure 1. 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 2).

Type Ia Supernova.



Figure 2. 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 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.

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

Some Methods for Estimating Distance to Galaxies

Key Concepts and Summary

Astronomers determine the distances to galaxies using a variety of methods, including the period-luminosity relationship for cepheid variables; objects such as type Ia supernovae, which appear to be standard bulbs; and the Tully-Fisher relation, which connects the line broadening of 21-cm radiation to the luminosity of spiral galaxies. Each method has limitations in terms of its precision, the kinds of galaxies with which it can be used, and the range of distances over which it can be applied.

Glossary

type Ia supernova

a supernova formed by the explosion of a white dwarf in a binary system and reach a luminosity of about $4.5 \times 10^9 L_{Sun}$; can be used to determine distances to galaxies on a large scale

Chapter 26 Section 26.5: The Expanding Universe

26.5 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 1), 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.

Vesto M. Slipher (1875–1969).



Figure 1. 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

866 | Chapter 26 Section 26.5: The Expanding Universe

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 2). 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.

Milton Humason (1891–1972).



Figure 2. 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 3), just as Lemaître had suggested.

Hubble's Law.



Figure 3. (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

$$V = H \times d$$

where v is the recession speed, d is the distance, and H is a number called the **Hubble constant**. This equation is now known as **Hubble's law**.

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 velocity 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, 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.

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

$$d = \frac{v}{H} = \frac{18,000 km/s}{\frac{22 km/s}{1 millionlight-years}} = \frac{18,000}{22} \times \frac{1 millionlight-years}{1} = 818 millionlight-years$$

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.

Check Your Learning

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:

 $v = d \times H = 500 million light - years \times \frac{22 km/s}{1 million light - years} = 11,000 km/s$

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 4). 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 4. 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 5. 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).

Expanding Raisin Bread.



Figure 5. 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.

Key Concepts and Summary

The universe is expanding. Observations show that the spectral lines of distant galaxies are redshifted, and that their recession velocities are proportional to their distances from us, a relationship known as Hubble's law. The rate of recession, called the Hubble constant, is approximately 22 kilometers per second per million light-years. We are not at the center of this expansion: an observer in any other galaxy would see the same pattern of expansion that we do. The expansion described by Hubble's law is best understood as a stretching of space.

For Further Exploration

Articles

Andrews, B. "What Are Galaxies Tell Trying to Us?" Astronomy (February 2011): 24. Introduction to our understanding of the shapes and evolution of different types of galaxies.

Bothun, G. "Beyond the Hubble Sequence." Sky & Telescope (May 2000): 36. History and updating of Hubble's classification scheme.

Christianson, G. "Mastering the Universe." Astronomy (February 1999): 60. Brief introduction to Hubble's life and work.

Dalcanton, J. "The Overlooked Galaxies." Sky & Telescope (April 1998): 28. On low-brightness galaxies, which have been easy to miss.

Freedman, W. "The Expansion Rate and Size of the Universe." *Scientific American* (November 1992): 76.

Hodge, P. "The Extragalactic Distance Scale: Agreement at Last?" Sky & Telescope (October 1993): 16.

Jones, B. "The Legacy of Edwin Hubble." Astronomy (December 1989): 38.

Kaufmann, G. and van den Bosch, F. "The Life Cycle of Galaxies." *Scientific American* (June 2002): 46. On galaxy evolution and how it leads to the different types of galaxies.

Martin, P. and Friedli, D. "At the Hearts of Barred Galaxies." Sky & Telescope (March 1999): 32. On barred spirals.

Osterbrock, D. "Edwin Hubble and the Expanding Universe." Scientific American (July 1993): 84.

Russell, D. "Island Universes from Wright to Hubble." Sky & Telescope (January 1999) 56. A history of our discovery of galaxies.

Smith, R. "The Great Debate Revisited." Sky & Telescope (January 1983): 28. On the Shapley-Curtis debate concerning the extent of the Milky Way and the existence of other galaxies.

ABC's of Distance: http://www.astro.ucla.edu/~wright/ distance.htm. A concise summary by astronomer Ned Wright of all the different methods we use to get distances in astronomy.

Cosmic Times 1929: http://cosmictimes.gsfc.nasa.gov/ online_edition/1929Cosmic/index.html. NASA project explaining Hubble's work and surrounding discoveries as if you were reading newspaper articles.

Edwin Hubble: The Man Behind the Name: https://www.spacetelescope.org/about/history/

the_man_behind_the_name/. Concise biography from the people at the Hubble Space Telescope.

Edwin Hubble: http://apod.nasa.gov/diamond_jubilee/d_1996/ sandage_hubble.html. An article on the life and work of Hubble by his student and successor, Allan Sandage. A bit technical in places, but giving a real picture of the man and the science.

NASA Science: Introduction to Galaxies: http://science.nasa.gov/ astrophysics/focus-areas/what-are-galaxies/. A brief overview with links to other pages, and recent Hubble Space Telescope discoveries.

National Optical Astronomy Observatories Gallery of Galaxies: https://www.noao.edu/image_gallery/galaxies.html. A collection of images and information about galaxies and galaxy groups of different types. Another impressive archive can be found at the European Southern Observatory site: https://www.eso.org/public/images/archive/category/galaxies/.

Sloan Digital Sky Survey: Introduction to Galaxies: http://skyserver.sdss.org/dr1/en/astro/galaxies/galaxies.asp. Another brief overview.

Universe Expansion: http://hubblesite.org/newscenter/archive/ releases/1999/19. The background material here provides a nice chronology of how we discovered and measured the expansion of the universe. Edwin Hubble (Hubblecast Episode 89): http://www.spacetelescope.org/videos/hubblecast89a/. (5:59).

Galaxies: https://www.youtube.com/watch?v=I82ADyJC7wE. An introduction.

Hubble's Views of the Deep Universe: https://www.youtube.com/ watch?v=argR2U15w-M. A 2015 public talk by Brandon Lawton of the Space Telescope Science Institute about galaxies and beyond (1:26:20).

Collaborative Group Activities

- A. Throughout much of the last century, the 100-inch telescope on Mt. Wilson (completed in 1917) and the 200-inch telescope on Palomar Mountain (completed in 1948) were the only ones large enough to obtain spectra of faint galaxies. Only a handful of astronomers (all male–since, until the 1960s, women were not given time on these two telescopes) were allowed to use these facilities, and in general the observers did not compete with each other but worked on different problems. Now there are many other telescopes, and several different groups do often work on the same problem. For example, two different groups have independently developed the techniques for using supernovae to determine the distances to galaxies at high redshifts. Which approach do you think is better for the field of astronomy? Which is more cost effective? Why?
- B. A distant relative, whom you invite to dinner so you can share all the exciting things you have learned in your astronomy class, says he does not believe that other galaxies are made up of stars. You come back to your group and ask them to help you respond. What kinds of measurements would you make to

show that other galaxies are composed of stars?

- C. Look at [link] with your group. What does the difference in color between the spiral arms and the bulge of Andromeda tell you about the difference in the types of stars that populate these two regions of the galaxy? Which side of the galaxy is closer to us? Why?
- D. What is your reaction to reading about the discovery of the expanding universe? Discuss how the members of the group feel about a universe "in motion." Einstein was not comfortable with the notion of a universe that had some overall movement to it, instead of being at rest. He put a kind of "fudge factor" into his equations of general relativity for the universe as a whole to keep it from moving (although later, hearing about Hubble and Humason's work, he called it "the greatest blunder" he ever made). Do you share Einstein's original sense that this is not the kind of universe you feel comfortable with? What do you think could have caused space to be expanding?
- E. In science fiction, characters sometimes talk about visiting other galaxies. Discuss with your group how realistic this idea is. Even if we had fast spaceships (traveling close to the speed of light, the speed limit of the universe) how likely are we to be able to reach another galaxy? Why?
- F. Despite his son's fascination with astronomy in college, Edwin Hubble's father did not want him to go into astronomy as a profession. He really wanted his son to be a lawyer and pushed him hard to learn the law when he won a fellowship to study abroad. Hubble eventually defied his father and went into astronomy, becoming, as you learned in this chapter, one of the most important astronomers of all time. His dad didn't live to see his son's remarkable achievements. Do you think he would have reconciled himself to his son's career choice if he had? Do you or does anyone in your group or among your friends have to face a choice between the passion in your heart and what others want you to do? Discuss how people in college today are dealing with such choices.

Review Questions

Describe the main distinguishing features of spiral, elliptical, and irregular galaxies.

Why did it take so long for the existence of other galaxies to be established?

Explain what the mass-to-light ratio is and why it is smaller in spiral galaxies with regions of star formation than in elliptical galaxies.

If we now realize dwarf ellipticals are the most common type of galaxy, why did they escape our notice for so long?

What are the two best ways to measure the distance to a nearby spiral galaxy, and how would it be measured?

What are the two best ways to measure the distance to a distant, isolated spiral galaxy, and how would it be measured?

Why is Hubble's law considered one of the most important discoveries in the history of astronomy?

What does it mean to say that the universe is expanding? What is expanding? For example, is your astronomy classroom expanding? Is the solar system? Why or why not?

Was Hubble's original estimate of the distance to the Andromeda galaxy correct? Explain.

Does an elliptical galaxy rotate like a spiral galaxy? Explain.

Why does the disk of a spiral galaxy appear dark when viewed edge on?

What causes the largest mass-to-light ratio: gas and dust, dark matter, or stars that have burnt out?

What is the most useful standard bulb method for determining distances to galaxies?

When comparing two isolated spiral galaxies that have the same apparent brightness, but rotate at different rates, what can you say about their relative luminosity?

If all distant galaxies are expanding away from us, does this mean we're at the center of the universe?

Is the Hubble constant actually constant?

Thought Questions

Where might the gas and dust (if any) in an elliptical galaxy come from?

Why can we not determine distances to galaxies by the same method used to measure the parallaxes of stars?

Which is redder-a spiral galaxy or an elliptical galaxy?

Suppose the stars in an elliptical galaxy all formed within a few million years shortly after the universe began. Suppose these stars have a range of masses, just as the stars in our own galaxy do. How would the color of the elliptical change over the next several billion years? How would its luminosity change? Why?

Starting with the determination of the size of Earth, outline a sequence of steps necessary to obtain the distance to a remote cluster of galaxies. (Hint: Review the chapter on Celestial Distances.)

Suppose the Milky Way Galaxy were truly isolated and that no other galaxies existed within 100 million light-years. Suppose that galaxies were observed in larger numbers at distances greater than 100 million light-years. Why would it be more difficult to determine accurate distances to those galaxies than if there were also galaxies relatively close by?

Suppose you were Hubble and Humason, working on the distances and Doppler shifts of the galaxies. What sorts of things would you have to do to convince yourself (and others) that the relationship you were seeing between the two quantities was a real feature of the behavior of the universe? (For example, would data from two galaxies be enough to demonstrate Hubble's law? Would data from just the nearest galaxies—in what astronomers call "the Local Group"—suffice?)

What does it mean if one elliptical galaxy has broader spectrum lines than another elliptical galaxy?

Based on your analysis of galaxies in [link], is there a correlation between the population of stars and the quantity of gas or dust? Explain why this might be.

Can a higher mass-to-light ratio mean that there is gas and dust present in the system that is being analyzed?

Figuring for Yourself

According to Hubble's law, what is the recessional velocity of a galaxy that is 10^8 light-years away from us? (Assume a Hubble constant of 22 km/s per million light-years.)

A cluster of galaxies is observed to have a recessional velocity of 60,000 km/s. Find the distance to the cluster. (Assume a Hubble constant of 22 km/s per million light-years.)

Suppose we could measure the distance to a galaxy using one of the distance techniques listed in [link] and it turns out to be 200 million

886 | Chapter 26 Section 26.5: The Expanding Universe
light-years. The galaxy's redshift tells us its recessional velocity is 5000 km/s. What is the Hubble constant?

Calculate the mass-to-light ratio for a globular cluster with a luminosity of $10^{6} L_{Sun}$ and 10^{5} stars. (Assume that the average mass of a star in such a cluster is $1 M_{Sun}$.)

Calculate the mass-to-light ratio for a luminous star of 100 M_{Sun} having the luminosity of 10⁶ L_{Sun} .

Glossary

Hubble constant

a constant of proportionality in the law relating the velocities of remote galaxies to their distances

Hubble's law

a rule that the radial velocities of remove galaxies are proportional to their distances from us

redshift

when lines in the spectra are displaced toward longer wavelengths (toward the red end of the visible spectrum)

Chapter 29 The Big Bang Section 29.3: The Beginning of the Universe

Space Telescope of the Future.



Figure 1. This drawing shows the James Webb Space Telescope, which is currently planned for launch in 2018. The silver sunshade shadows the primary mirror and science instruments. The primary mirror is 6.5 meters (21 feet) in diameter. Before and during launch,

the mirror will be folded up. After the telescope is placed in its

orbit, ground controllers will command it to unfold the mirror petals. To see distant galaxies whose light has been shifted to long wavelengths, the telescope will carry several instruments for taking infrared images and spectra. (credit: modification of work by NASA)

In previous chapters, we explored the contents of the universe-planets, stars, and galaxies-and learned about how these objects change with time. But what about the universe as a whole? How old is it? What did it look like in the beginning? How has it changed since then? What will be its fate?

Cosmology is the study of the universe as a whole and is the

subject of this chapter. The story of observational cosmology really begins in 1929 when Edwin Hubble published observations of redshifts and distances for a small sample of galaxies and showed the then-revolutionary result that we live in an expanding universe–one which in the past was denser, hotter, and smoother. From this early discovery, astronomers developed many predictions about the origin and evolution of the universe and then tested those predictions with observations. In this chapter, we will describe what we already know about the history of our dynamic universe and highlight some of the mysteries that remain.

29.3 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

Chapter 29 The Big Bang Section 29.3: The Beginning of the Universe | 889

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 1). 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."

Abbé Georges Lemaître (1894–1966).



Figure 1. 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. 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 2) 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.

George Gamow and Collaborators.



Figure 2. 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 3 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.

Temperature of the Universe.



Figure 3. 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 $E = mc^2$ (see the chapter on The Sun: A Nuclear Powerhouse).

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 3). 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 4). 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.



Particle Interactions in the Early Universe.

Chapter 29 The Big Bang Section 29.3: The Beginning of the Universe | 897

Figure 4. (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 4). 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.

These nice animations that explain the way in which different elements formed in the history of the universe are from the University of Chicago's Origins of the Elements site.

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.

900 | Chapter 29 The Big Bang Section 29.3: The Beginning 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/m³. 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.

Key Concepts and Summary

Lemaître, Alpher, and Gamow first worked out the ideas that are today called the Big Bang theory. The universe cools as it expands. The energy of photons is determined by their temperature, and calculations show that in the hot, early universe, photons had so much energy that when they collided with one another, they could produce material particles. As the universe expanded and cooled, protons and neutrons formed first, then came electrons and positrons. Next, fusion reactions produced deuterium, helium, and lithium nuclei. Measurements of the deuterium abundance in today's universe show that the total amount of ordinary matter in the universe is only about 5% of the critical density.

Glossary

deuterium

a form of hydrogen in which the nucleus of each atom consists of one proton and one neutron

fusion

the building of heavier atomic nuclei from lighter ones

lithium

the third element in the periodic table; lithium nuclei with three protons and four neutrons were manufactured during the first few minutes of the expansion of the universe

Chapter 29 Section 29.4: The Cosmic Microwave Background

29.4 The Cosmic Microwave Background



The description of the first few minutes of the universe is based

Chapter 29 Section 29.4: The Cosmic Microwave Background | 903 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 ([link]). 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 1).

Cosmic Microwave Background and Clouds Compared.



Figure 1. (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 Expansion and Redshift image [link]). 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 2) 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.

Robert Wilson (left) and Arno Penzias (right).



Figure 2. These two scientists are standing in front of the hornshaped 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)**¹ 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.

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 3). 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.



Cosmic Background Radiation.

Figure 3. 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-thanaverage density would have attracted additional matter and eventually grown into the galaxies and clusters that we see today. It turned 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 4).

CMB Observations.



Figure 4. 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 5).

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.

Picturing Space Curvature for the Entire Universe.



Figure 5. 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 6, 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.

Comparison of CMB Observations with Possible Models of the Universe.



Figure 6. 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
- Fraction of universe's content that is "dark energy": 68.5% ± 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.

Key Concepts and Summary

When the universe became cool enough to form neutral hydrogen atoms, the universe became transparent to radiation. Scientists have detected the cosmic microwave background (CMB) radiation from this time during the hot, early universe. Measurements with the COBE satellite show that the CMB acts like a blackbody with a temperature of 2.73 K. Tiny fluctuations in the CMB show us the seeds of large-scale structures in the universe. Detailed measurements of these fluctuations show that we live in a criticaldensity universe and that the critical density is composed of 31% matter, including dark matter, and 69% dark energy. Ordinary matter-the kinds of elementary particles we find on Earth-make up only about 5% of the critical density. CMB measurements also indicate that the universe is 13.8 billion years old.

Footnotes

• 1 Recall that microwaves are in the radio region of the electromagnetic spectrum.

Glossary

cosmic microwave background (CMB)

microwave radiation coming from all directions that is the redshifted afterglow of the Big Bang

flat universe

a model of the universe that has a critical density and in which the geometry of the universe is flat, like a sheet of paper

photon decoupling time

when radiation began to stream freely through the universe without interacting with matter

For Further Exploration

Collaborative Group Activities

- A. This chapter deals with some pretty big questions and ideas. Some belief systems teach us that there are questions to which "we were not meant to know" the answers. Other people feel that if our minds and instruments are capable of exploring a question, then it becomes part of our birthright as thinking human beings. Have your group discuss your personal reactions to discussing questions like the beginning of time and space, and the ultimate fate of the universe. Does it make you nervous to hear about scientists discussing these issues? Or is it exciting to know that we can now gather scientific evidence about the origin and fate of the cosmos? (In discussing this, you may find that members of your group strongly disagree; try to be respectful of others' points of view.)
- B. A popular model of the universe in the 1950s and 1960s was the so-called steady-state cosmology. In this model, the universe was not only the same everywhere and in all directions (homogeneous and isotropic), but also the same *at all times*. We know the universe is expanding and the galaxies are thinning out, and so this model hypothesized that new matter was continually coming into existence to fill in the space between galaxies as they moved farther apart. If so, the infinite universe did not have to have a sudden beginning, but could simply exist forever in a steady state. Have your group discuss your reaction to this model. Do you find it more appealing philosophically than the Big Bang model? Can you cite some evidence that indicates that the universe was not the same billions of years ago as it is now-that it is not in a steady state?
- C. One of the lucky accidents that characterizes our universe is the fact that the time scale for the development of intelligent

life on Earth and the lifetime of the Sun are comparable. Have your group discuss what would happen if the two time scales were very different. Suppose, for example, that the time for intelligent life to evolve was 10 times greater than the mainsequence lifetime of the Sun. Would our civilization have ever developed? Now suppose the time for intelligent life to evolve is ten times shorter than the main-sequence lifetime of the Sun. Would we be around? (This latter discussion requires considerable thought, including such ideas as what the early stages in the Sun's life were like and how much the early Earth was bombarded by asteroids and comets.)

- D. The grand ideas discussed in this chapter have a powerful effect on the human imagination, not just for scientists, but also for artists, composers, dramatists, and writers. Here we list just a few of these responses to cosmology. Each member of your group can select one of these, learn more about it, and then report back, either to the group or to the whole class.
 - The California poet Robinson Jeffers was the brother of an astronomer who worked at the Lick Observatory. His poem "Margrave" is a meditation on cosmology and on the kidnap and murder of a child: http://www.poemhunter.com/best-poems/robinson
 - jeffers/margrave/.
 - In the science fiction story "The Gravity Mine" by Stephen Baxter, the energy of evaporating supermassive black holes is the last hope of living beings in the far future in an ever-expanding universe. The story has poetic description of the ultimate fate of matter and life and is available online at: http://www.infinityplus.co.uk/stories/ gravitymine.htm.
 - The musical piece YLEM by Karlheinz Stockhausen takes its title from the ancient Greek term for primeval material revived by George Gamow. It tries to portray the oscillating universe in musical terms. Players actually

expand through the concert hall, just as the universe does, and then return and expand again. See: http://www.karlheinzstockhausen.org/ylem_english.htm.

- The musical piece Supernova Sonata http://www.astro.uvic.ca/~alexhp/new/ supernova_sonata.html by Alex Parker and Melissa Graham is based on the characteristics of 241 type Ia supernova explosions, the ones that have helped astronomers discover the acceleration of the expanding universe.
- Gregory Benford's short story "The Final Now" envisions the end of an accelerating open universe, and blends religious and scientific imagery in a very poetic way. Available free online at: http://www.tor.com/stories/ 2010/03/the-final-now.
- E. When Einstein learned about Hubble's work showing that the universe of galaxies is expanding, he called his introduction of the cosmological constant into his general theory of relativity his "biggest blunder." Can your group think of other "big blunders" from the history of astronomy, where the thinking of astronomers was too conservative and the universe turned out to be more complicated or required more "outside-the-box" thinking?
Chapter 30 Life in the Universe Section 30.1: The Cosmic Context for Life

Thinking Ahead

Astrobiology: The Road to Life in the Universe.



Figure 1. In this fanciful montage produced by a NASA artist, we see one roadmap for discovering life in the universe. Learning more about the origin, evolution, and properties of life on Earth aids us in searching for evidence of life beyond our planet. Our neighbor world, Mars, had warmer, wetter conditions billions of years ago that might have helped life there begin. Farther out, Jupiter's moon

Europa represents the icy moons of the outer solar system. Beneath their shells of solid ice may lie vast oceans of liquid water that could support biology. Beyond our solar system are stars that host their own planets, some of which might be similar to Earth in the ability to support liquid water—and a thriving biosphere—at the planet's surface. Research is pushing actively in all these directions with the goal of proving a scientific answer to the question, "Are we alone?" (credit: modification of work by NASA)

As we have learned more about the universe, we have naturally wondered whether there might be other forms of life out there. The ancient question, "Are we alone in the universe?" connects us to generations of humans before us. While in the past, this question was in the realm of philosophy or science fiction, today we have the means to seek an answer through scientific inquiry. In this chapter, we will consider how life began on Earth, whether the same processes could have led to life on other worlds, and how we might seek evidence of life elsewhere. This is the science of astrobiology.

The search for life on other planets is not the same as the search for *intelligent* life, which (if it exists) is surely much rarer. Learning more about the origin, evolution, and properties of life on Earth aids us in searching for evidence of all kinds of life beyond that on our planet.

30.1 The Cosmic Context for Life



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 1). The third planet from the Sun, as it cooled, eventually allowed the formation of large quantities of liquid water on its surface.

Comet Hyakutake.



Figure 1. 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 2). 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.

Young Human.



Chapter 30 Life in the Universe Section 30.1: The Cosmic Context for Life | 927

Figure 2. 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, smaller planets are not unique.

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. 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?"

Key Concepts and Summary

Life on Earth is based on the presence of a key unit known as an organic molecule, a molecule that contains carbon, especially complex hydrocarbons. Our solar system formed about 5 billion years ago from a cloud of gas and dust enriched by several generations of heavier element production in stars. Life is made up of chemical combinations of these elements made by stars. The Copernican principle, which suggests that there is nothing special about our place in the universe, implies that if life could develop on Earth, it should be able to develop in other places as well. The Fermi paradox asks why, if life is common, more advanced life-forms have not contacted us.

Glossary

organic molecule

a combination of carbon and other atoms-primarily hydrogen, oxygen, nitrogen, phosphorus, and sulfur-some of which serve as the basis for our biochemistry

Chapter 30 Section 30.2: Astrobiology

30.2 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

932 | Chapter 30 Section 30.2: Astrobiology 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 1).

Cloud of Gas and Dust.



Figure 1. 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.

Hydrothermal vents are beginning to seem more likely as early contributors to the organic compounds found on Earth. Read about hydrothermal vents, watch videos and slideshows on these and other deep-sea wonders, and try an interactive simulation of hydrothermal circulation at the Woods Hole Oceanographic Institution 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 2). 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.

Stromatolites Preserve the Earliest Physical Representation of Life on Earth.



(b)

Figure 2. 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.

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 3). Understanding what processes control the accessibility of elements at all scales is thus a critical part of identifying habitable environments.

Chlorophyll Abundance.



Figure 3. 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.

Grand Prismatic Spring in Yellowstone National Park.



Figure 4. 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 5). 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 5); 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 temperatures than this. And yet, there are organisms whose biochemistry remains intact and operates just fine at temperatures 20 degrees higher.

Hydrothermal Vent on the Sea Floor.



Figure 5. 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 6). 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.

Spain's Rio Tinto.



Figure 6. 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 7).

Salt Ponds.



Figure 7. 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.

Key Concepts and Summary

The study of life in the universe, including its origin on Earth, is called astrobiology. Life as we know it requires water, certain elemental raw materials (carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur), energy, and an environment in which the complex chemistry of life is stable. Carbon-based (or organic) molecules are abundant in space and may also have been produced by processes on Earth. Life appears to have spread around our planet within 400 million years after the end of heavy bombardment, if not sooner. The actual origin of life-the processes leading from chemistry to biology-is not completely understood. Once life took hold, it evolved to use many energy sources, including first a range of different chemistries and later light, and diversified across a range of environmental conditions that humans consider "extreme." This proliferation of life into so many environmental niches, so relatively soon after our planet became habitable, has served to make many scientists optimistic about the chances that life could exist elsewhere.

Glossary

amino acids

organic compounds that are the molecular building blocks of proteins

astrobiology

the multidisciplinary study of life in the universe: its origin, evolution, distribution, and fate; similar terms are *exobiology* and *bioastronomy*

DNA (deoxyribonucleic acid)

a molecule that stores information about how to replicate a cell and its chemical and structural components

extremophile

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

gene

the basic functional unit that carries the genetic (hereditary) material contained in a cell

habitable environment

an environment capable of hosting life

organic compound

a compound containing carbon, especially a complex carbon compound; not necessarily produced by life

photosynthesis

a 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

protein

a key biological molecule that provides the structure and function of the body's tissues and organs, and essentially carries out the chemical work of the cell

RNA (ribonucleic acid)

a molecule that aids in the flow of genetic information from DNA to proteins

stromatolites

solid, layered rock formations that are thought to be the fossils of oxygen-producing photosynthetic bacteria in rocks that are 3.5 billion years old

thermophile

an organism that can tolerate high temperatures

Chapter 30 Section 30.3: Searching for Life beyond Earth

30.3 Searching for Life beyond Earth



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 NASA's *Curiosity* rover, which landed on Mars' surface in 2012.

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 1). The mudstone can tell us a great deal about the wet environments in which they formed.

Mudstone.



Figure 1. Shown are the first holes drilled by

NASA's *Curiosity* Mars rover into a mudstone, with "fresh" drillpilings 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 2). 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 3). 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.

Jupiter's Moons.



Figure 2. 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 oxygencontaining 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.

Jupiter's Moon Europa, as Imaged by NASA's Galileo Mission.



Figure 3. 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 4), 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.

Image of Saturn's Moon Enceladus from NASA's Cassini Mission.



Figure 4. 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 5). 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.

Image of Saturn's Moon Titan from NASA's Cassini Mission.





Figure 5. (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) 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?

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.

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 6). What types of observations might constitute good evidence for life?

Earth, as Seen by NASA's Voyager 1.



Figure 6. 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 7).

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.



Spectrum of Light Transmitted through Earth's Atmosphere.

Figure 7. 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)

Key Concepts and Summary

The search for life beyond Earth offers several intriguing targets. Mars appears to have been more similar to Earth during its early history than it is now, with evidence for liquid water on its ancient surface and perhaps even now below ground. The accessibility of the martian surface to our spacecraft offers the exciting potential to directly examine ancient and modern samples for evidence of life. In the outer solar system, the moons Europa and Enceladus likely host vast sub-ice oceans that may directly contact the underlying rocks–a good start in providing habitable conditions–while Titan offers a fascinating laboratory for understanding the sorts of organic chemistry that might ultimately provide materials for life. And the last decade of research on exoplanets leads us to believe that there may be billions of habitable planets in the Milky Way Galaxy. Study of these worlds offers the potential to find biomarkers indicating the presence of life.

Footnotes

• 1 The 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 (2²), and when the orbital distance increases tenfold, the illumination decreases by 100 times (10²). 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.

Glossary

biomarker

evidence of the presence of life, especially a global indication of life on a planet that could be detected remotely (such as an unusual atmospheric composition)

habitable zone

the region around a star in which liquid water could exist on the surface of terrestrial-sized planets, hence the most probable place to look for life in a star's planetary system

Chapter 30 Section 30.4: The Search for Extraterrestrial Intelligence

30.4 The Search for Extraterrestrial Intelligence



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 1). 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.

Interstellar Messages.



Figure 1. (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)

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 longestwavelength 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 lowestfrequency (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). 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 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 defines the cosmic haystack problem.

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 2). 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.

Project Ozma and the Allen Telescope Array.



Figure 2. (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 life–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 include the giant Arecibo radio dish in Puerto Rico and the Green Bank Telescope in West Virginia, which is the largest steerable radio telescope in the world.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.

JILL TARTER: TRYING TO MAKE CONTACT

1997 was quite a year for Jill Cornell Tarter (Figure 3), 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 3. 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 extraterrestrial civilizations, leading Jill Tarter to receive 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, among others.

Watch the TED talk Jill Tarter gave on the fascination of the search for intelligence.

The Drake Equation

scientific the first devoted At meeting 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 4.)

Drake Equation.



Figure 4. 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,

 $N = R_{total} \times L$

To make this formula easier to use (and more interesting), however, Drake separated the rate of formation R_{total} into a series of probabilities:

$$R_{total} = R_{star} \times f_p \times f_e \times f_l \times f_i \times f_e$$

R_{star} is the rate of formation of stars like the Sun in our Galaxy, which is about 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

$N = R_{total} \times L = L$

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.

Check Your Learning

Suppose we assume that this stage in our history lasts only one century.

ANSWER:

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.

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 possibilities. 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 reradiated into space. That infrared radiation could be detected by our instruments, and searches for such infrared sources are also underway (Figure 5).

Wide-Field Infrared Survey Explorer (WISE).



Figure 5. 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

Chapter 30 Section 30.4: The Search for Extraterrestrial Intelligence | 991

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 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?

Key Concepts and Summary

Some astronomers are engaged in the search for extraterrestrial intelligent life (SETI). Because other planetary systems are so far away, traveling to the stars is either very slow or extremely expensive (in terms of energy required). Despite many UFO reports and tremendous media publicity, there is no evidence that any of these are related to extraterrestrial visits. Scientists have determined that the best way to communicate with any intelligent civilizations out there is by using electromagnetic waves, and radio waves seem best suited to the task. So far, they have only begun to comb the many different possible stars, frequencies, signal types, and other factors that make up what we call the cosmic haystack problem. Some astronomers are also undertaking searches for brief, bright pulses of visible light and infrared signatures of huge construction projects by advanced civilizations. If we do find a signal someday, deciding whether to answer and what to answer may be two of the greatest challenges humanity will face.

For Further Exploration

Articles

Astrobiology

Chyba, C. "The New Search for Life in the Universe." Astronomy (May 2010): 34. An overview of astrobiology and the search for life out there in general, with a brief discussion of the search for intelligence.

Dorminey, B. "A New Way to Search for Life in Space." Astronomy (June 2014): 44. Finding evidence of photosynthesis on other worlds.

McKay, C., & Garcia, V. "How to Search for Life on Mars." Scientific American (June 2014): 44–49. Experiments future probes could perform.

Reed, N. "Why We Haven't Found Another Earth Yet." Astronomy (February 2016): 25. On the search for smaller earthlike planets in their star's habitable zones, and where we stand.

Shapiro, R. "A Simpler Origin of Life." Scientific American (June

994 | Chapter 30 Section 30.4: The Search for Extraterrestrial Intelligence
2007): 46. New ideas about what kind of molecules formed first so life could begin.

Simpson, S. "Questioning the Oldest Signs of Life." *Scientific American* (April 2003): 70. On the difficulty of interpreting biosignatures in rocks and the implications for the search for life on other worlds.

SETI

Chandler, D. "The New Search for Alien Intelligence." *Astronomy* (September 2013): 28. Review of various ways of finding other civilizations out there, not just radio wave searches.

Crawford, I. "Where Are They?" Scientific American (July 2000): 38. On the Fermi paradox and its resolutions, and on galactic colonization models.

Folger, T. "Contact: The Day After." *Scientific American* (January 2011): 40–45. Journalist reports on efforts to prepare for ET signals; protocols and plans for interpreting messages; and discussions of active SETI.

Kuhn, J., et al. "How to Find ET with Infrared Light." Astronomy (June 2013): 30. On tracking alien civilizations by the heat they put out.

Lubick, N. "An Ear to the Stars." Scientific American (November 2002): 42. Profile of SETI researcher Jill Tarter.

Nadis, S. "How Many Civilizations Lurk in the Cosmos?" Astronomy (April 2010): 24. New estimates for the terms in the Drake equation.

Shostak, S. "Closing in on E.T." Sky & Telescope (November 2010): 22. Nice summary of current and proposed efforts to search for intelligent life out there.

Websites

Astrobiology

Astrobiology Web: http://astrobiology.com/. A news site with good information and lots of material.

Exploring Life's Origins: http://exploringorigins.org/index.html. A website for the Exploring Origins Project, part of the multimedia exhibit of the Boston Museum of Science. Explore the origin of life on Earth with an interactive timeline, gain a deeper knowledge of the role of RNA, "build" a cell, and explore links to learn more about astrobiology and other related information.

History of Astrobiology: https://astrobiology.nasa.gov/about/ history-of-astrobiology/. By Marc Kaufman, on the NASA Astrobiology site.

Life, Here and Beyond: https://astrobiology.nasa.gov/about/. By Marc Kaufman, on the NASA Astrobiology site.

SETI

Berkeley SETI Research Center: https://seti.berkeley.edu/. The University of California group recently received a \$100 million grant from a Russian billionaire to begin the Breakthrough: Listen project.

Fermi Paradox: http://www.seti.org/seti-institute/project/ details/fermi-paradox. Could we be alone in our part of the Galaxy or, more dramatic still, could we be the only technological society in the universe? A useful discussion.

Planetary Society: http://www.planetary.org/explore/projects/ seti/. This advocacy group for exploration has several pages devoted to the search for life.

SETI Institute: http://www.seti.org. A key organization in the search for life in the universe; the institute's website is full of information and videos about both astrobiology and SETI.

SETI: http://www.skyandtelescope.com/tag/seti/. Sky Telescope magazine offers good articles on this topic.

Videos

Astrobiology

Copernicus Complex: Are We Special in the Cosmos?: https://www.youtube.com/watch?v=ERp0AHYRm_Q. A video of a popular-level talk by Caleb Scharf of Columbia University (1:18:54).

Life at the Edge: Life in Extreme Environments on Earth and the Search for Life in the Universe: https://www.youtube.com/ watch?v=91JQmTn0SF0. A video of a 2009 nontechnical lecture by Lynn Rothschild of NASA Ames Research Center (1:31:21).

Saturn's Moon Titan: A World with Rivers, Lakes, and Possibly Even Life: https://www.youtube.com/watch?v=bbkTJeHoOKY. A video of a 2011 talk by Chris McKay of NASA Ames Research Center (1:23:33).

SETI

Allen Telescope Array: The Newest Pitchfork for Exploring the Cosmic Haystack: https://www.youtube.com/ watch?v=aqsI1HZCgUM. A 2013 popular-level lecture by Jill Tarter of the SETI Institute (1:45:55).

Confessions of an Alien Hunter: http://fora.tv/2009/03/31/ Seth_Shostak_Confessions_of_an_Alien_Hunter. 2009 interview with Seth Shostak on FORA TV (36:27).

Search for Extra-Terrestrial Intelligence: Necessarily a Long-Term Strategy: http://www.longnow.org/seminars/02004/jul/ 09/the-search-for-extra-terrestrial-intelligence-necessarily-along-term-strategy/. 2004 talk by Jill Tarter at the Long Now Foundation (1:21:13).

Search for Intelligent Life Among the Stars: New Strategies: https://www.youtube.com/watch?v=m9WxW2ktcKU. A 2010 nontechnical talk by Seth Shostak of the SETI Institute (1:29:58).

Collaborative Group Activities

- A. If one of the rocks from Mars examined by a future mission to the red planet does turn out to have unambiguous signs of ancient life that formed on Mars, what does your group think would be the implications of such a discovery for science and for our view of life elsewhere? Would such a discovery have any long-term effects on your own thinking?
- B. Suppose we receive a message from an intelligent civilization around another star. What does your group think the implications of this discovery would be? How would your own thinking or personal philsophy be affected by such a discovery?
- C. A radio message has been received from a civilization around a star 40 light-years away, which contains (in pictures) quite a bit of information about the beings that sent the message. The president of the United States has appointed your group a high-level commission to advise whether humanity should answer the message (which was not particularly directed at us, but comes from a beacon that, like a lighthouse, sweeps out a circle in space). How would you advise the president? Does your group agree on your answer or do you also have a minority view to present?
- D. If there is no evidence that UFOs are extraterrestrial visitors, why does your group think that television shows, newspapers, and movies spend so much time and effort publicizing the

point of view that UFOs are craft from other worlds? Make a list of reasons. Who stands to gain by exaggerating stories of unknown lights in the sky or simply fabricating stories that alien visitors are already here?

- E. Does your group think scientists should simply ignore all the media publicity about UFOs or should they try to respond? If so, how should they respond? Does everyone in the group agree?
- F. Suppose your group is the team planning to select the most important sights and sounds of Earth to record and put on board the next interstellar spacecraft. What pictures (or videos) and sounds would you include to represent our planet to another civilization?
- G. Let's suppose Earth civilization has decided to broadcast a message announcing our existence to other possible civilizations among the stars. Your group is part of a large task force of scientists, communications specialists, and people from the humanities charged with deciding the form and content of our message. What would you recommend? Make a list of ideas.
- H. Think of examples of contact with aliens you have seen in movies and on TV. Discuss with your group how realistic these have been, given what you have learned in this class. Was the contact in person (through traveling) or using messages? Why do you think Hollywood does so many shows and films that are not based on our scientific understanding of the universe?
- I. Go through the Drake equation with your group and decide on values for each factor in the estimate. (If you disagree on what a factor should be within the group, you can have a "minority report.") Based on the factors, how many intelligent, communicating civilizations do you estimate to be thriving in our Galaxy right now?

Review Questions

What is the Copernican principle? Make a list of scientific discoveries that confirm it.

Where in the solar system (and beyond) have scientists found evidence of organic molecules?

Give a short history of the atoms that are now in your little finger, going back to the beginning of the universe.

What is a biomarker? Give some possible examples of biomarkers we might look for beyond the solar system.

Why are Mars and Europa the top targets for the study of astrobiology?

Why is traveling between the stars (by creatures like us) difficult?

What are the advantages to using radio waves for communication between civilizations that live around different stars? List as many as you can.

What is the "cosmic haystack problem"? List as many of its components as you can think of.

What is a habitable zone?

Why is the simultaneous detection of methane and oxygen in an atmosphere a good indication of the existence of a biosphere on that planet?

What are two characteristic properties of life that distinguish it from nonliving things?

What are the three requirements that scientists believe an

environment needs to supply life with in order to be considered habitable?

Can you name five environmental conditions that, in their extremes, microbial life has been challenged by and has learned to survive on Earth?

Thought Questions

Would a human have been possible during the first generation of stars that formed right after the Big Bang? Why or why not?

If we do find life on Mars, what might be some ways to check whether it formed separately from Earth life, or whether exchanges of material between the two planets meant that the two forms of life have a common origin?

What kind of evidence do you think would convince astronomers that an extraterrestrial spacecraft has landed on Earth?

What are some reasons that more advanced civilizations might want to send out messages to other star systems?

What are some answers to the Fermi paradox? Can you think of some that are not discussed in this chapter?

Why is there so little evidence of Earth's earliest history and therefore the period when life first began on our planet?

Why was the development of photosynthesis a major milestone in the evolution of life?

Does all life on Earth require sunshine?

Why is life unlikely to be found on the surface of Mars today?

In this chapter, we identify these characteristic properties of life: life extracts energy from its environment, and has a means of encoding and replicating information in order to make faithful copies of itself. Does this definition fully capture what we think of as "life"? How might our definition be biased by our terrestrial environment?

Given that no sunlight can penetrate Europa's ice shell, what would be the type of energy that could make some form of europan life possible?

Why is Saturn's moon Enceladus such an exciting place to send a mission?

In addition to an atmosphere dominated by nitrogen, how else is Saturn's moon Titan similar to Earth?

How can a planet's atmosphere affect the width of the habitable zone in its planetary system?

Why are we limited to finding life on planets orbiting other stars to situations where the biosphere has created planet-scale changes?

Figuring for Yourself

Suppose astronomers discover a radio message from a civilization whose planet orbits a star 35 light-years away. Their message encourages us to send a radio answer, which we decide to do. Suppose our governing bodies take 2 years to decide whether and how to answer. When our answer arrives there, their governing bodies also take two of our years to frame an answer to us. How long after we get their first message can we hope to get their reply to ours? (A question for further thinking: Once communication gets going, should we continue to wait for a reply before we send the next message?)

1002 | Chapter 30 Section 30.4: The Search for Extraterrestrial Intelligence

The light a planet receives from the Sun (per square meter of planet surface) decreases with the square of the distance from the Sun. So a planet that is twice as far from the Sun as Earth receives $(1/2)^2 = 0.25$ times (25%) as much light and a planet that is three times as far from the Sun receives $(1/3)^2 = 0.11$ times (11%) as much light. How much light is received by the moons of Jupiter and Saturn (compared to Earth), worlds which orbit 5.2 and 9.5 times farther from the Sun than Earth?

Think of our Milky Way Galaxy as a flat disk of diameter 100,000 light-years. Suppose we are one of 1000 civilizations, randomly distributed through the disk, interested in communicating via radio waves. How far away would the nearest such civilization be from us (on average)?

Glossary

Drake equation

a formula for estimating the number of intelligent, technological civilizations in our Galaxy, first suggested by Frank Drake

SETI

the search for extraterrestrial intelligence; usually applied to searches for radio signals from other civilizations

This is where you can add appendices or other back matter.