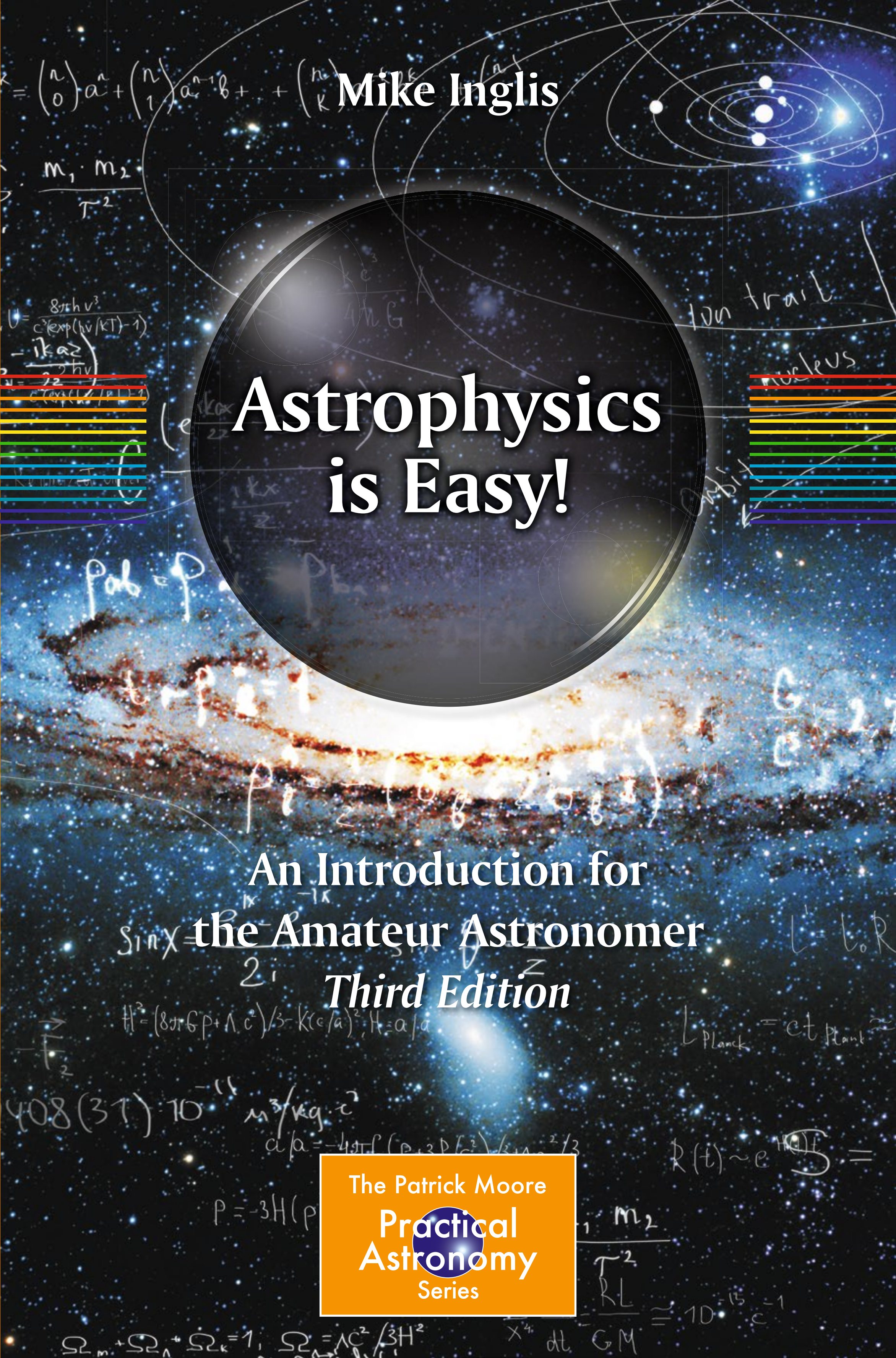


Mike Inglis

Astrophysics is Easy!

An Introduction for the Amateur Astronomer

Third Edition

The Patrick Moore
Practical Astronomy
Series

The Patrick Moore Practical Astronomy Series

Series Editor
Gerald R. Hubbell
Mark Slade Remote Observatory
Locust Grove, VA, USA

The Patrick Moore Practical Astronomy Series is a treasure trove of how-to guides for the amateur astronomer. The books in this series are written for hobbyists at all levels, from the enthusiastic newcomer to the veteran observer. They thus go far beyond more general, popular-level books in both scope and depth, exploring in detail the latest trends, techniques, and equipment being used by amateur astronomers around the world.

You will find herein a diverse list of books on constellations, astronomy catalogues, astrophotography, eclipse chasing, telescope equipment, software, and so much more. All books in the series boast full-color images as well as practical sections for putting your newfound knowledge to use, including star charts and target objects, glossaries, hands-on DIY projects, troubleshooting walk-throughs, and a plethora of other helpful features.

Overall, this series bridges the gap between the many introductory books available and more specialized technical publications, providing digestible, hands-on guides for those wishing to expand their knowledge of the night skies.

Astrophysics Is Easy!

An Introduction for the Amateur Astronomer

Michael Inglis

Third Edition

Springer

Michael Inglis
Long Island, NY, USA

ISSN 1431-9756 ISSN 2197-6562 (electronic)
The Patrick Moore Practical Astronomy Series
ISBN 978-3-031-16804-8 ISBN 978-3-031-16805-5 (eBook)
https://doi.org/10.1007/978-3-031-16805-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2007, 2015, 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to the memory of my great friends.

Michael Alan Mercer 1954–2020

Nigel Brian Pointer 1955–2020

Iain Nicolson 1945–2020

Donald Tinkler 1947–2022

And to my wonderful mother

Myra Doreen Inglis 1924–2021

Preface

When Springer asked me to consider a third edition of the book, I was delighted. It would allow me to correct any errors, expand certain chapters, and even add new chapters. Furthermore, as well as updating the book with the latest research results, I could add something I had been thinking about for quite some time—Thought Questions—embedded in the main text. These simple questions have no mathematics content, but rather are a way of testing yourself about the many disparate aspects of astrophysics.

In addition, at the end of the chapters are a few simple mathematically based questions that one could attempt. Note that all the answers to the questions and problems are provided at the end of the book. Finally, for the first time, I have included a few photos in the book. These are just to illustrate some of the objects under discussion. They were taken by amateur astronomers with simple equipment from suburban locations. There are many books that have far more photos, as well as the internet where there are literally thousands, but that was not the purpose of the book.

So, once again, I took paper to pen for this third edition, and began a journey to explain the mysterious, beautiful, and sometimes astounding complexities of stars, galaxies, the material that lies between, and the universe itself. It was a journey that took many roads with numerous side turnings as I often spent many long, lonely hours worrying whether I was being too obtuse, or at times patronizing, as it is a fact that many amateur astronomers are very knowledgeable of the subject that they pursue with a passion. However, the new edition eventually came into sight, and this, for me a mammoth task, was completed.

Throughout the entire process of writing the third edition, I was fortunate enough to have the support of the Hannah Kauffman, astronomy editor at Springer Publishing, who knows only too well that astronomy authors are a breed apart and need to be pampered and dealt with using extreme patience. Thank you, Hannah, dinner is on me! I must also thank my great friend John Watson, also associated with Springer, who gave the initial thumbs up when I first outlined expanding the original book with a third edition. John is an amateur astronomer himself, so he knows exactly what should go into a book, and perhaps even more importantly, what should be left out! John, I owe you a pint. I would also like to thank Janny Jonkers, a wonderful friend from the Netherlands, who took on the onerous task of proofreading the entire text.

I was fortunate to have been taught astronomy by some of the world's leading experts, and it was, and still is, a privilege to have known them. In my humble opinion, not only are they superb astronomers, whether theoretical or observational, but also wonderful educators. They are Chris Kitchin, Alan McCall, the late, great Iain Nicolson, Robert Forrest, and the late Lou Marsh. They were the best teachers I ever had.

It is important to acknowledge the pioneering work that is being done in amateur astronomical spectroscopy, and to that end I would like to thank the following spectroscopists for allowing me to use their work in the book. They are Tom Field, for designing a simple and affordable but superb piece of spectroscopic equipment, and for spearheading the revolution, along with Hansen Torsen, Ken Wright, William Wiethoff, and David Strange.

During the time spent writing both the first, second, and third editions, usually alone, usually at night, usually tired, I had the company of some wonderful musicians whose music is truly sublime. They are Steve Roach, David Sylvian, John Martyn, and the Blue Nile. And let's not forget the input of Arfur the Cat, who spent a lot of time lying on the keyboard as I was trying to type.

Many friends have helped raise my spirits during those times when not all was going right, according to the Inglis Master Plan. They listened to me complain, laughed at my jokes, and helped me remain sane—for the most part. So, I want to say thank you to my great friends—Professor Peter Harris and Dr. William Worthington. It is nice to know that beer is the universal lubricant of friendship, whether it is McMullen's or Harvey's.

Astronomy is a very important part of my life, but not as important as my family; my brother Bob is a great friend and a strong source of support, especially during the formative years as a young astronomer. My mother Myra was amazing, full of energy, spirit, and laughter, and had been supportive of my dream to be an astronomer since I was knee-high to a tripod. She was truly an example to us all. And of course, Karen, I am not exaggerating when I say this book would not have originally seen the light of day without her help. “*Diolch Cariad.*”

For making my life worthwhile and fun, cheers!

Long Island, NY, USA Michael Inglis

Rationale for the Book

To most normal people, astrophysics—the science of stars, galaxies, and the universe we live in—would seem to be a topic suited to a university-level textbook, and so the idea of a guide to astrophysics for the amateur astronomer may not, on first appearance, make any sense. However, let me assure you that anyone can understand how a star is born, lives its life, and dies, how galaxies are thought to evolve and what their shape can tell us about their origins and age, and even how the universe began and how it may end. It can even tell you how and why the planets move. In fact, very little mathematics is needed, and when it is used, it is only a matter of multiplication, division, subtraction, and addition[1]!

What's more, there are many wonderful objects that can be observed in the night sky that will illustrate even the most obtuse astrophysics concepts. All one needs is a willingness to learn and a dark night sky.

Learning about, say, the processes that give rise to star formation, or what happens to a very large star as it dies, what keeps the Moon orbiting Earth, or even why some galaxies are spiral in shape whereas others are elliptical can add another level of enjoyment and wonder to an observing session. For instance, many amateur astronomers are familiar with the star Rigel, in the constellation Orion, but how many of you know that it is a giant star, with a mass more than 40 times that of our Sun, and it is nearly half a million times more luminous than the Sun! Or that our closest large galaxy, M31 in

---

[1] OK, we do use powers of ten occasionally, and numbers multiplied by themselves from time to time. But nothing else … honest!

Andromeda, has a supermassive black hole lurking at its center with a mass of over 50 million times that of the Sun. Or that the Orion Nebula, regarded by many as the premier nebula in the sky, is in fact an enormous stellar nursery where stars are being born as you read this book. Knowing details such as these can add another level of enjoyment to your observing sessions.

Each section of this book addresses a specific aspect of astrophysics. The first part focuses on the concepts needed for a complete understanding of the remainder of the book, and as such will be divided into specific topics, such as the brightness, color, and distance of stars. Then we look at what is probably the most basic, yet important, tool of an astronomer, namely spectroscopy. It is true to say that nearly all of what we know about stars and galaxies was and is determined from this important technique, and there has been a revolution in amateur astronomical spectroscopy in the past few years.

We then spend a fair amount of time looking at something called the Hertszprung-Russell diagram; if ever a single concept or diagram could epitomize a star's life (and even a star cluster's life), the HR diagram, as it is known, is the one to do it. It is perhaps the most important and useful concept in all of stellar evolution, and it is fair to say that once you understand the HR diagram, you understand how a star evolves.

Moving on to the objects themselves, we then cover a topic that many may think strange to find in a book devoted to astrophysics, namely the Solar System. But as you will see, there is a surpassingly large amount of what could be described of as introductory astrophysics when discussing certain aspects of our Solar System, especially the dynamics of the planets (and indeed asteroids and comets). A small amount of history will also be covered dealing with the main antagonists in the story and how their ideas led, more or less, to the picture we have today of our Solar System.

Following this chapter, we look at the formation of stars from dust and gas clouds and conclude with the final aspect of a star's life, which can end in the spectacular event known as a supernova, resulting in the formation of a neutron star and even perhaps a black hole!

Such is the interest in Black Holes that they have a chapter all to themselves. But before this we diverge slightly to a short discussion on Einstein's Theories of Relativity.

Another chapter is the inclusion of a topic that only 15 years ago was a fledgling, and somewhat obscure field of study, but is now at the forefront of discovery, namely the detection of exoplanets[2]!

---

[2] As well as exomoons, exocomets, and exoasteroids!

On a grander scale, we delve into galaxies, their shapes (or morphology, as it is called), distribution in space, and origins.

We follow this with a chapter dealing with those galaxies that seem to have a lot more going on within them than one usually sees, or expects—active galaxies and their nuclei, or AGN as they are more properly referred to.

Our penultimate chapter discusses maybe the biggest subject of all—cosmology. Not a topic that is often discussed from an amateur astronomy point of view, but even here, there are a few surprising aspects that can be observed, including one deceptively simple, yet stunningly deep, question that can be asked at star parties, along with its surprising answer!

Finally, and new to this edition, is a chapter on those very esoteric parts of astrophysics that seem more fantasy than fact but are nevertheless discussed by many astrophysicists.

The topics covered are chosen specifically so that examples of objects under discussion can be observed; thus, at every point in our journey, an observing section will describe the objects that best demonstrate the topics discussed. Many of the objects, whether they are stars, nebulae, or galaxies, will be visible with modest optical instruments, and many with the naked eye. In a few exceptional cases, a medium-aperture telescope may be needed. Of course, not all observable objects will be presented, but just a representative few (usually the brightest examples). These examples will allow you to learn about stars, nebulae, and galaxies at your own pace, and they will provide a detailed panorama of the amazing objects that most of us observe on a clear night.

For those of you who have a mathematical mind, some mathematics will be provided in the specially labeled areas. But take heart and fear not—you do not have to understand any mathematics to be able to read and understand this book; it is only there to highlight and further describe the mechanisms and principles of astrophysics. However, if you are comfortable with the mathematics, then I recommend that you read these sections, as they will further your understanding of the various concepts and equip you to determine such parameters as a star's age and lifetime, distance, mass, and brightness. All the mathematics presented will be simple, of a level comparable to that of a high-school student. In fact, to make the mathematics simpler, we will use rough (but perfectly acceptable) approximations and perform back-of-the-envelope calculations, which, surprisingly, produce rather accurate answers! To further your expertise in the simple mathematics, several very easy problems are included at the end of each chapter. A completely new aspect of the book is the inclusion, throughout the text, of "Thought Questions." These seek to test how well you have understood the sections. However, with both the end of chapter problems and thought questions, you do not need to attempt any of them to enjoy the book, rather they are just there for fun.

An astute reader will notice immediately that there are *no* star maps in the book! The reason for this is simple. In previous books that I have written, star maps were included, but their size generated some criticism. Some readers believed that the maps were too small, and I tended to agree. To be able to offer large and detailed star maps of every object mentioned in this book would entail a doubling of its size, and probably a tripling of cost. With the plethora of star-map software that is available these days, it is far easier for readers to make their own maps than to present any here.

A final point I wish to emphasize here is that the book can be read in several ways. Certainly, you can start at the beginning and read through to the end. But if you are particularly interested in, say, supernovae and the final stages of a star's life, or in galaxy clusters, there is no reason that you shouldn't go straight to that section. Some of the nomenclature might be unfamiliar, but I've attempted to write the book with enough description that this shouldn't be a problem. Also, many of you will undoubtedly go straight to the observing lists. Read the book in the way that is most enjoyable to you.

Without further ado, let us begin on your voyage of discovery.

Acknowledgements

I would like to thank the following people and organizations for their help, and permission to quote their work and for use of the data they provided:

The European Space Organization, for permission to use the *Hipparcos* and *Tycho* catalogs.

My colleagues at *Suffolk County Community College* in the United States, for their support and encouragement.

The astronomers at *Princeton University* in the United States, for many helpful discussions on the whole process of star formation.

The astronomers at the *University of Hertfordshire* in the United Kingdom for inspirational lectures and discussions.

Gary Walker, of the *American Association of Variable Star Observers*, for information on the many types of variable stars.

Cheryl Gundy, of the *Space Telescope Science Institute* in the United States, for supplying astrophysical data on many of the objects discussed.

Dr. Stuart Young, formerly of the *University of Hertfordshire* in the United Kingdom, and *Rochester Institute of Technology* in the United States, for discussions and information relating to star formation and the Hertzsprung Russell diagram, and impromptu tutorials on many aspects of astronomy.

Dr. Chris Packham, formerly of the *University of Florida*, and currently at the *University of Texas*, both in the United States, for his help on pointing out several mistakes I have made over the years, and for his input regarding AGNs.

Karen Milstein for the superb and professional work that she did reading through the initial proofs of the book, when there seemed to be more errors than facts!

The *Smithsonian Astrophysical Observatory* in the United States, for providing data on many of the stars and star clusters.

Robert Forrest, formerly of the *University of Hertfordshire Observatory* in the United Kingdom, for use of his observing notes.

Michael Hurrell and the late Donald Tinkler of the *South Bayfordbury Astronomical Society* in the United Kingdom, for use of their observing notes.

The plethora of people who bought the first edition of the book and kindly pointed out the typos. I hope I caught them all.

In developing a book of this type, which presents a considerable amount of detail, it is nearly impossible to avoid error. If any arise, I apologize for the oversight, and mistakes are due to me and me alone.

Contents

About the Author

Mike Inglis was born in Wales in the United Kingdom but lives and works in the United States, where he is Professor of Astrophysics at the State University of New York. His qualifications include a BSc Physics, MSc in Astronomy and Astronautics, and a PhD in Astrophysics. He is a Fellow of the Royal Astronomical Society, a NASA Solar System Ambassador, a member of the European Astronomical Society, and a member of the International Astronomical Union. He is the author of many books and papers, including *Field Guide to Deep Sky Objects* (Springer), *An Observer's Guide to Stellar Evolution* (Springer), *Astronomy of the Milky Way*, Vols I & II (Springer), and *Observer's Guide to Star Clusters*. He is also the Series Editor of Springer's Astronomy Observing Guides, the Springer Lecture Notes in Undergraduate Physics and Astrophysics.

Chapter 1

Tools of the Trade

1.1. Angular Measurement

Let us begin our journey with the simple, but also very important, topic of angular measurement, as we will be using the concept discussed here throughout the book.

Although most of the objects described in the text are only seen telescopically, we will, when discussing a few objects, and especially the Solar System, refer to angular distances that can be estimated by eye alone.

Thus, from the horizon to the point directly above your head—the zenith—is 90°. If you look due south and scan the horizon going from south to west, continuing to the north, then east and back to south, you will have traversed 360°. In addition, 1° is quite a large size, so it can be subdivided into 60 arc minutes (′). Furthermore, an arc minute can be further subdivided into 60 arc seconds (″).

The angular diameter of the Moon and also of the Sun is 0.5° (or 30 arc minutes). Other distances that may be useful are:

δ–ε Orionis	1.25°
α–γ Aquilae	2°
δ–ζ Orionis	3°
α–β Canis Majoris	4°
α–β Ursae Majoris	5°
α–δ Ursae Majoris	10°

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_1

Further approximate distances are:

The width of the nail of your index finger at arm's length	1°.
The width of your clenched fist held at arm's length	8°.
The span of your open handheld at arm's length	18°.

Although this book deals primarily with those objects that lie beyond the Solar System, we nevertheless will initially be dealing with the dynamics of the Solar System in Chap. 2, and use terms that are often frequently quoted but rarely defined. So this section will deal with these ideas, described collectively as the configuration of the planets.

It is only in the past 450 years that astronomers have been able to use telescopes to observe, and indeed study, the sky.[1] Before this time, astronomy was limited to the naked eye, and when discussing the planets of the Solar System, this was limited again to those that could be seen visually with the unaided eye. For reasons that will be discussed in the chapter on the Solar System, the planets can be split into two groups—the inferior planets (Mercury and Venus) and the superior planets (Mars, Jupiter, and Saturn).[2]

The configuration of planets deals with their positions with respect to Earth's position.[3] The names of the terms origins lie in the deep past. However, these definitions are very useful to the amateur astronomer, as they can be used to determine the optimum planetary observing times throughout the year(s), and the times during the night when certain planets will be visible.

There are three diagrams that define these terms, Figs. 1.1, 1.2, and 1.3, and they will be very useful as an aid to understanding the following definitions.

At first glance, it may appear that these definitions are confusing and cumbersome, but in reality, they are very useful to the amateur astronomer, as it will allow one to know in advance the positions of the planets prior to planning an observing session. In addition, it now becomes crystal clear as to why the best observing can be at a planet's opposition, whereas at conjunction, the best that can be said is "…don't even bother[4]!"

---

[1] This has its upside. No light pollution whatsoever, except for the occasional burning field of hay, and burning city, e.g., the fire of London.

[2] It is possible to see Uranus visually with no optical aid. However, because of its faintness and very slow motion across the sky, it was, and continues to be, mistaken for a star.

[3] The planets all lie more or less on a flat plane, called the ecliptic. Pluto, now classified as a dwarf planet, lies at a large inclination the ecliptic, which provided some of the evidence that it was "different" from the other planets.

[4] There are a few rare occasions when one can observe an inferior planet at inferior conjunction, and this is when the planet *transits*. That is, it moves across the face of the Sun. The next transits of Mercury are 2016 and 2019. Venus, alas, has no transits until 2117. Bad luck.

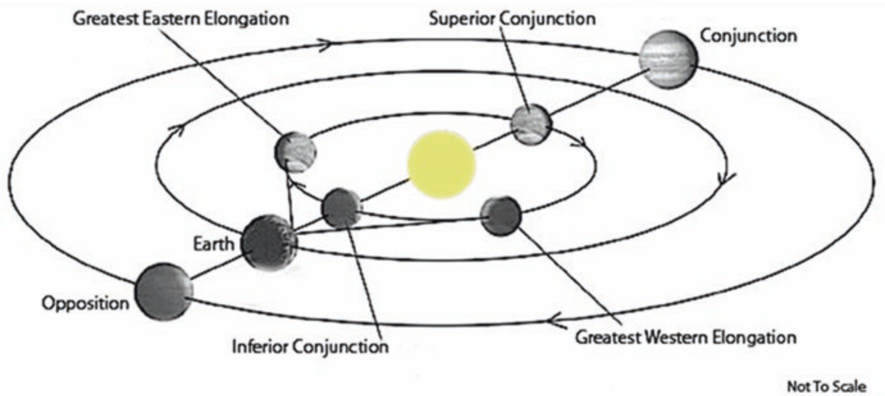


Fig. 1.1. Planetary configurations
CONJUNCTION: When a planet lies along the line of sight to the Sun [i.e., lies in the same direction of the Sun]
INFERIOR CONJUNCTION: The planet lies between Earth and the Sun (inferior planets only)
SUPERIOR CONJUNCTION: The planet lies beyond the Sun (both superior and inferior planets)
OPPOSITION: Earth lies *between* the planet and the Sun (superior planets only). The planet rises as the Sun is setting
GREATEST ELONGATION: When an inferior planet reaches its greatest angle away from the Sun as viewed from Earth

1.2. Distances in Astronomy

The most familiar unit of astronomical distance is the *light year*. This is simply the distance that electromagnetic radiation travels in a vacuum in 1 year. As light travels at a speed of 300,000 km per second (km s^{-1}), the distance it travels in 1 year is 9,460,000,000,000 km, which is close enough to call it 10 trillion km! This is often abbreviated to *l. y.*

The next commonly used distance unit is the *parsec*. This is the distance at which a star would have an annual parallax of 1 second of arc, hence the term *par*allax *sec*ond. The section that follows will discuss how the parallax is determined.

Another unit of distance sometimes used is the a*stronomical unit* (AU), which is the mean distance between Earth and the Sun, and is 149,597,870 km. Note that 1 light year is nearly 63,200 au.

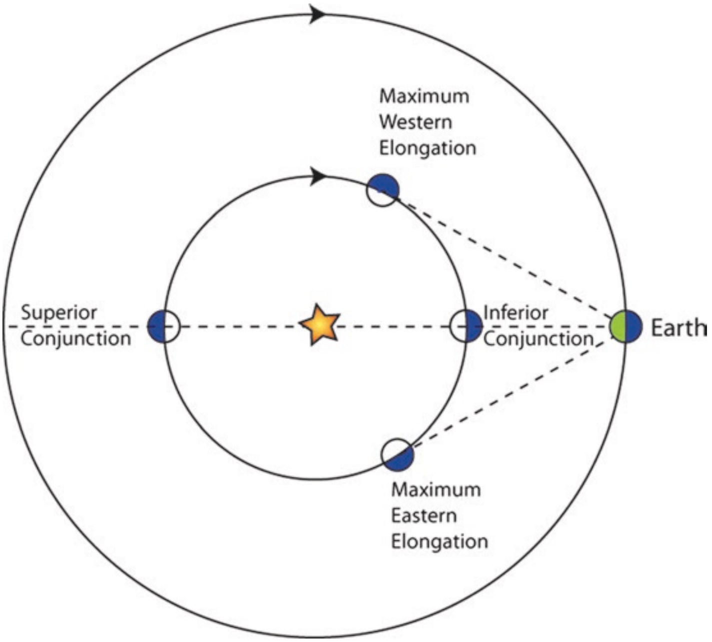


Fig. 1.2. Planetary configurations—elongations and conjunctions
MAXIMUM EASTERN ELONGATION: Planet is at its furthest *east* of the Sun as seen from Earth (28° for Mercury, 47° for Venus). Rises and sets *after* the Sun ("Evening Star")
MAXIMUM WESTERN ELONGATION: Planet is at its furthest *west* of the Sun as seen from Earth. Rises and sets *before* the Sun ("Morning Star")
EASTERN QUADRATURE: Planet at right angles to the Earth-Sun line. Planet rises at noon, sets at midnight (superior planets only)
WESTERN QUADRATURE: Planet at right angles to the Earth-Sun line. Planet rises at midnight, sets at noon (superior planets only)

In order to determine many of the basic parameters of any object in the sky, it is first necessary to determine its proximity to us. We shall see later that this is vitally important because a star's bright appearance in the night sky could signify that it is close to us OR that it may be an inherently bright star. Conversely, some stars may appear faint because they are at an immense distance from us or because they are very faint stars in their own right. We need to be able to decide which is the correct explanation.

Determining distances in astronomy has always been, and continues to be, fraught with difficulty and error. There is still no general consensus as to the best method, at least for distances to other galaxies and to the farthest edges of our own galaxy—the Milky Way. The oldest method, still used today, is probably the most accurate, especially for determining the distances to stars.

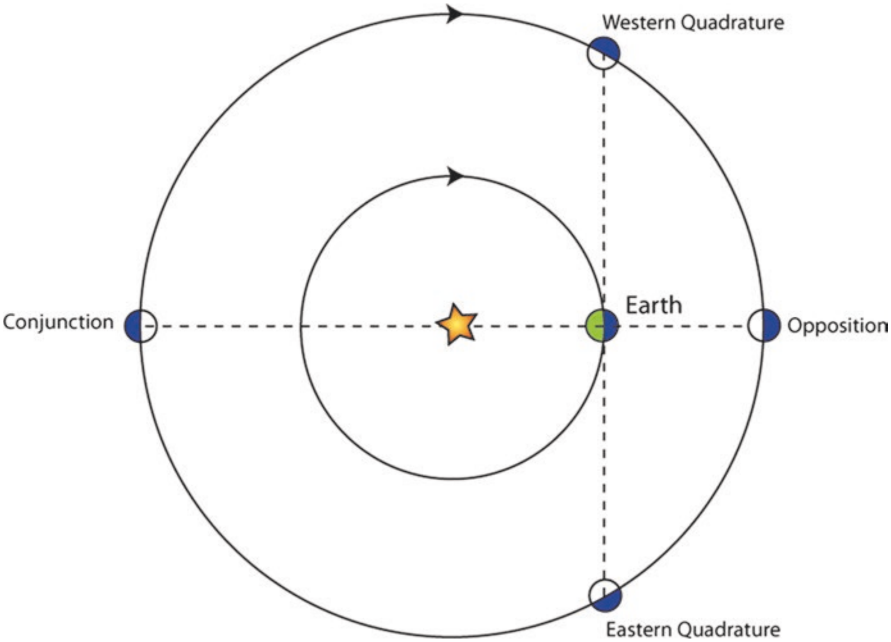


Fig. 1.3. Planetary configurations—opposition, conjunction, and quadrature

This simple technique is called *stellar parallax*, and it is basically the angular measurement when the star is observed from two different locations in Earth's orbit. These two positions are generally 6 months apart, so the star will appear to shift its position with respect to the more distant background stars. The parallax (p) of the star observed is equal to half the angle through which its apparent position appears to shift. The larger the parallax, p, the smaller the distance, d, to the star. Figure 1.4 illustrates this concept.

If a star has a measured parallax of 1 arc second (1/3600 of a degree) and the baseline is 1 au, which is the average distance from Earth to the Sun, then the star's distance is 1 parsec (pc)—"the distance of an object that has a **par**allax of one **sec**ond of arc." This is the origin of the term and the unit of distance used most frequently in astronomy.[5]

The distance, d, of a star in parsecs is given by the reciprocal of p and is usually expressed as thus:

$$d = \frac{1}{p}$$

Thus, using the above equation, a star with a measured parallax of 0.1 arc seconds is at a distance of 10 pc, and another with a parallax of 0.05 arc seconds is 20 pc distant. See Math Box 1.1 for further examples.

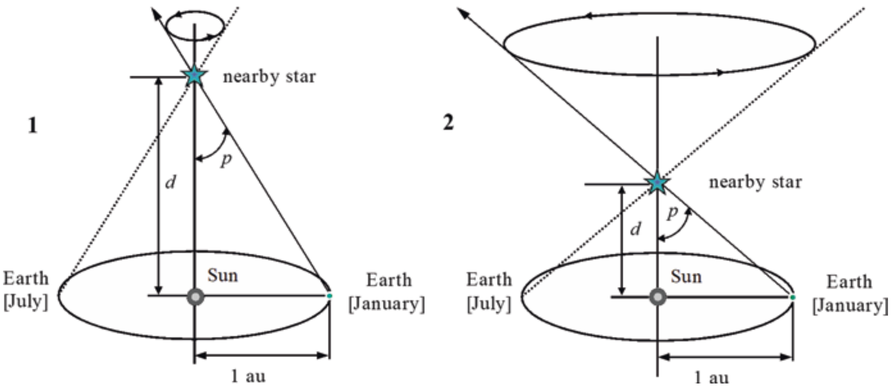


Fig. 1.4. Stellar parallax. (*1*) Earth orbits the Sun, and a nearby star will shift its position with respect to the background stars. The parallax, *p*, of the star is the angular measurement of Earth's orbit as seen from the star. (*2*) The closer the star, the greater the parallax angle

Math Box 1.1: Relationship Between Parallax and Distance

$$d = \frac{1}{p}$$

d = the distance to a star measured in parsecs.
p = the parallax angle of the measured star, in arc seconds.

This simple relationship is a significant reason that most astronomical distances are expressed in parsecs, rather than light years. The brightest star in the night sky is Sirius [α Canis Majoris], which has a parallax of 0.379 arc seconds. Thus, its distance from us is:

$$d = \frac{1}{p} = \frac{1}{0.379} = 2.63 \text{ pc}$$

Note that 1 parsec is equivalent to 3.26 light years, so this distance can also be expressed as:

$$d = 2.63 \times \frac{3.26 \text{ ly}}{1 \text{ pc}} = 8.6 \text{ light years}$$

Surprisingly, all known stars have a parallax angle smaller than 1 arc second, and angles smaller than about 0.01 arc seconds are very difficult to measure from Earth due to the effects of the atmosphere; this limits the distance measured to about 100 pc [1/0.01]. The satellite Hipparcos, however, launched in 1989, was able to measure parallax angles to an accuracy of 0.001 arc seconds, which allowed distances to be determined to about 1000 pc.[6]

However, even this great advance in distance determination is only useful for relatively close stars. Most of the stars in the galaxy are too far for parallax measurements to be taken. Another method must be used.

Many stars actually alter in brightness (these are the variable stars), and several of them play an important role in distance determination. Although we will discuss their properties in far greater detail later, it is instructive to mention them here.

Two types of variable stars in particular are useful in determining distances. These are the *Cepheid* variable stars and *RR Lyrae* variable stars.[7] Both are classified as *pulsating variables*, which are stars that actually change their diameter over a period of time. The importance of these stars lies in the fact that their average brightnesses, or luminosities,[8] and their periods of variability, are linked: the longer the time taken for the star to vary in brightness [the period], the greater the luminosity. This is the justifiably famous *period-luminosity relationship*.[9] It is relatively easy to measure the period of a star, and this is something that many amateur astronomers still do. Once this has been measured, you can determine the star's luminosity. By comparing the luminosity, which is a measure of the intrinsic brightness of the star, with the brightness it appears to have in the sky, its distance can be calculated.[10] Using Cepheids, distances out to around 60 million light years have been determined.

A similar approach is taken with the RR Lyrae stars, which are less luminous than Cepheids and have periods of less than a day. These stars allow distances to about two million light years to be determined.

---

[5] One parsec is equal to 3.26 light years, 3.09 × 1013 km, or 206,265 au. 1 au is 149,597,870 km.

[6] Nearly 200 previously unobserved stars were discovered, the nearest about 18 light years away. In addition, several hundred stars originally believed to be within 75 light years were in fact found to be much farther away.

[7] The most famous Cepheid variable star is Polaris, the North Star. It varies its visual brightness by about 10% in just under 4 days. Recent data show that the variability is decreasing, and the star may, at some time in the future, cease to pulsate. We discuss Polaris and other important variable stars in detail in a later section.

[8] We will discuss the meaning of the term luminosity later. For the time being, think of it as the star's brightness.

[9] Henrietta Leavitt discovered the period-luminosity relationship in 1908 while working at the Harvard College Observatory. She studied photographs of the Magellanic Clouds and found more than 1700 variable stars.

[10] The relationship between the apparent brightness of a star and its intrinsic brightness will be discussed in the next section.

Another method of distance determination is that of spectroscopic parallax, whereby determining a star's spectral classification can lead to a measure of its intrinsic luminosity, which can then be compared with its apparent brightness to determine its distance.

Our remaining distance determination methods are used for the objects farthest from us—galaxies. These methods are the Tully Fisher method and the very famous Hubble law.

Again, all of these methods—Cepheid variable, Tully Fisher, and the Hubble law—will be addressed in greater detail later in the book.

A final note on distance determination is in order. Do not be fooled into thinking that these various methods produce exact measurements. They do not. A small amount of error is inevitable. Sometimes this error is about 10% or 25%, but an error of 50% is not uncommon. Remember that a 25% error for a star estimated to be at a distance of 4000 l.y. means it could be anywhere from 3000 to 5000 l.y. away. Table 1.1, presented below, lists the 20 nearest stars.

Let us now look at some of the nearest stars in the night sky from an observational point of view. The list discussed here is by no means complete but rather includes those stars that are most easily seen. Many of the nearest

Table 1.1 The 20 nearest stars in the sky[a]

	Star	Distance (l. y.)	Constellation
1	Sun	–	–
2	Proxima Centauri	4.22	Centaurus
3	Alpha Centauri A[b]	4.39	Centaurus
4	Barnard's star	5.94	Ophiuchus
5	Wolf 359	7.8	Leo
6	Lalande 21185	8.31	Ursa Major
7	Sirius A[9]	8.60	Canis Major
8	UV Ceti A[9]	8.7	Cetus
9	Ross 154	9.69	Sagittarius
10	Ross 248	10.3	Andromeda
11	Epsilon Eridani	10.49	Eradinus
12	HD 217987	10.73	Piscis Austrinus
13	Ross 128	10.89	Virgo
14	L 789-6 A[9]	11.2	Aquarius
15	61 Cygni A	11.35	Cygnus
16	Procyon A[9]	11.42	Canis Minoris
17	61 Cygni B	11.43	Cygnus
18	HD 173740	11.47	Draco
19	HD 173739	11.64	Draco
20	GX Andromadae[9]	11.64	Andromeda

[a] Brown dwarf stars are not included in the list

[b] This signifies that the star is in fact part of a double star system, and the distance quoted is for components A and B

stars are very faint and thus present an observing challenge, so they are not included here.

Throughout the book, we will use the following nomenclature with regard to stars: first will be its common name, followed by its scientific designation. The next item will be its position in right ascension and declination. The final item will identify the months when the star is best positioned for observation.

The next line will present both standard data and information that is pertinent to the star under discussion—its apparent magnitude, followed by its absolute magnitude (both these terms are discussed in detail in following sections), specific data relating to the topic, and, finally, the constellation in which the star resides.

Here is the listing of the nearest stars.[11]

The Sun		January–December
−26.78 m	4.82 M	G2 V

The closest star to Earth and the object without which no life would have evolved on Earth. It is visible every day, throughout the year, unless you happen to live in the UK.

Proxima Centauri	V645 Cen	14h 29.7m	−62° 41′	April
11.01vm[12]	15.45 M	4.22 l. y.	0.772″	Centaurus

This is the second-closest star to Earth and the closest star to the Solar System. and thus it is included, albeit faint. It is a red dwarf star and also a flare star with frequent bursts, having maximum amplitude of around one magnitude. Recent data indicate that it is not, as previously thought, physically associated with α , but is in fact on a hyperbolic orbit around the star and just passing through the system.

Sirius A	α Canis Majoris	06h 45.8m	−16° 43′	January
−1.46 m/+8.44 m	1.42 M/11.34 M	8.6 l. y.	0.379″	Canis Major

Sirius, also known as the Dog Star, is a lovely star to observe and is the sixth-closest star and also the brightest star in the sky. It is famous among amateurs for the exotic range of colors it exhibits, due to the effects of the atmosphere. It also has a white dwarf star companion—the first to be discovered. A dazzling sight in any optical device.

---

[11] Most of the nearest stars are very faint, so only the brighter ones will be mentioned here. Exceptions to this will be made, however, if the object has an important role in astronomy. A companion book to this one—*Field Guide to the Deep Sky Objects*—provides much more information and detail regarding the nearest stars. Furthermore, the field guide addresses many techniques to enhance your observational skills, such as dark adaption, averted vision, etc. Note that brown dwarf stars are not listed, even though a few are very close to us.

[12] Denotes that the star, and thus the magnitude, is variable.

Procyon	α Canis Minoris	07h 39.3m	−56° 13′	January
0.38 m/10.7 m	2.66 M/12.98 M	11.41 l. y.	0.283″	Canis Minor

The fourteenth nearest star is a very easy object to observe as well as being the eighth-brightest star in the sky. It is notable for the fact that it has, like nearby Sirius, a companion star that is a white dwarf. However, unlike Sirius, the dwarf star is not easily visible in small amateur telescopes, having a magnitude of 10.8 and a mean separation of only 5 arc seconds.

Barnard's Star	HD21185	17h 57.8m	+4° 41′	May
9.53 m	13.24 M	5.94 l. y.	0.545″	Ophiuchus

The third-closest star is a red dwarf, but what makes this star so famous is that it has the largest proper motion of any star[13]—0.4 arc seconds per year. Barnard's Star, also known as Barnard's Runaway Star, has a velocity of 140 km per second; at this rate, it would take 150 years for the star to move the distance equivalent to the Moon's diameter across the sky. It is believed to be one of the oldest stars in the Milky Way, and in 1998 a stellar flare was believed to have occurred on the star. Due to the unpredictability of flares, this makes the star a perfect target for observers. It is also thought that the star belongs to the galaxy's halo population.

61 Cygni A/B	Bessel's Star	21h 06.9m	+38° 45′	August
5.21/6.03vm	7.49/8.31 M	11.35 l. y.	0.285″	Cygnus

This is a very nice double star, with a separation 30.3 arc seconds and a PA of 150°. Both stars are dwarfs and have a nice orange color. Bessel's is famous as the first star to have its distance measured successfully by F. W. Bessel in 1838 using stellar parallax.

GX And	Grb 34	00h 18.3m	+44° 01′	September
8.09vm	10.33 M	11.65 l. y.	0.279″	Andromeda

This is half of a noted red dwarf binary systems with the primary star itself a spectroscopic double star. Also known as Groombridge 34 A, it is located about 1/4° north of 26 Andromedae.

Lacille 9352	HD 217987[14]	23h 05.5m	−35° 52′	September
7.34 m	9.76 M	10.73 l. y.	0.305″	Pisces Austrinus

This is a red dwarf star, with the fourth-fastest proper motion of any known star traversing a distance of nearly 7 arc seconds a year, and thus would take about 1000 years to cover the angular distance of the full Moon, which is 1/2°. Lacille is in the extreme southeast of the constellation, about 1° SSE of π Pisces Austrinus.

---

[13] The proper motion of a star is its apparent motion across the sky.

[14] The HD signifies it is the 217,987th object in the Henry Draper Catalogue.

UV Ceti	Luyten 726-8 A	01h 38.8m	−17° 57′	October
12.57vm	15.42 M	8.7 l. y.	0.381″	Cetus

The seventh-closest star is a red dwarf system and is rather difficult, but not impossible, to observe. The UV prefix indicates that the two components are flare stars; the fainter star is referred to in older texts as "Luytens Flare Star," after its discoverer, W. J. Luyten, who first observed it in 1949.

Epsilon Eridani	HD 22049	03h 32.9m	−09° 27′	November
3.73 m	6.19 M	10.49 l. y.	0.311″	Eridanus

The tenth-closest star is a naked-eye object. It is the third closest individual star or star system visible to the unaided eye that some observers describe as having a yellow color, while others say it is more orange. The star was believed to the closest system that had a planet in orbit, and maybe even two, until the unconfirmed discovery of Alpha Centauri Bb. There is also evidence that Epilson Eridani has two asteroid belts made of rocky and metallic debris left over from the early stages of planetary formation, similar to our Solar System, and even a broad outer ring of icy objects similar to our Kuiper Belt. All in all a very interesting star!

1.3. Brightness and Luminosity of Astronomical Objects

There are an immense number of stars and galaxies in the sky, and, for the most part, all are powered by the same process that powers the Sun. But this does not mean that they are all alike—far from it. Stars differ in many respects, such as mass, size, etc. One of the most important characteristics is their *luminosity*, *L*. Luminosity is usually measured in *watts* (*W*), or as a multiple of the Sun's luminosity,[15] $L_\odot$. This is the amount of energy that the star emits each second. However, we cannot measure a star's luminosity directly because its brightness as seen from Earth depends on its distance as well as its true luminosity. For instance, α Centauri A, and the Sun have similar luminosities, but in the night sky, α Centauri A is a dim point of light because it is about 270,000 times farther from Earth than the Sun is.

To determine the true luminosity of a star, we need to know its *apparent brightness* and we define this as the amount of light reaching Earth per unit area.[16] As light moves away from the star, it will spread out over increasingly larger regions of space, obeying what is termed *inverse square law*. Let me illustrate this with the following examples.

---

[15] One watt is equal to 1 J per second. The Sun's luminosity is 3.86×10^{26} W. It is often designated by the symbol $L_\odot$.

[16] The scientific term for apparent brightness is *flux*.

If the Sun were viewed at a distance twice that of Earth's, then it would appear fainter by a factor of $2^2 = 4$.

Similarly, if we viewed it a distance three times that of Earth's, it would now be fainter by a factor of $3^2 = 9$.

If we now viewed it from a distance ten times that of Earth's, it would appear $10^2 = 100$ times fainter.

You can now probably get the idea of an inverse square relationship.

Thus, if we observed the Sun from the same location as α Centauri A, it would be dimmed by $270{,}000^2$, or about 70 billion times!

The inverse square law describes the amount of energy that enters, say, your eye or a detector. Try to imagine an enormous sphere of radius *d*, centered on a star. The amount of light that will pass through a square meter of the sphere's surface is the total luminosity, *L*, divided by the total surface area of the sphere. Now, as the surface area of a sphere is given by the simple formula $4\pi d^2$, you can see that as the area of the sphere increases, *d* increases, and so the amount of luminosity that reaches you will decrease. You can see why the amount of luminosity that arrives on Earth from a star is determined by the star's distance. See Math Box 1.2 for an example of the use of the formula.

Math Box 1.2: The Luminosity Distance Formula

The relationship among distance, brightness and luminosity is given as:

$$b = \frac{L}{4\pi \mathrm{d}^2}$$

where

b is the brightness of the star in W/m^2
L is the star's luminosity in W.
d is the distance to the star in meters.

Example:

Let us apply this formula to Sirius, which is at a distance of 8.6 light years and has a luminosity of 25.4 $L_\odot$. [Note: 1 light year is 9.46×10^{15} m, thus 8.6 light years is $8.6 \times 9.46 \times 10^{15} = 8.14 \times 10^{16}$ m]

$$b = \frac{25.4 \times 3.86 \times 10^{26}\,\mathrm{W}}{4\pi\left(8.14 \times 10^{16}\,\mathrm{m}\right)^2}$$

$$b \approx 1.18 \times 10^{-7}\,\mathrm{W/m^2}$$

This means that, say, a detector of a 1 m^2 area (possibly a very large Dobsonian reflecting telescope) will receive approximately one millionth of a watt!

This quantity, the amount of energy that arrives at your eye, is the apparent brightness mentioned earlier (sometimes just called the brightness of a star) and is measured in watts per square meter (W/m^2). See Math Box 1.3 for an example of the use of the formula.

Astronomers measure a star's brightness with light-sensitive detectors, and the procedure is called *photometry*.

Math Box 1.3: Luminosity, Distance and Brightness

To determine a star's luminosity, we need to know its distance and apparent brightness. We can achieve this quite easily by using the Sun as a reference. First, let's rearrange the formula thus:

$$L = 4\ d^2 b$$

$$\mathrm{L}_\odot = 4\pi\ d_\odot^2 \mathrm{b}_\odot$$

Now let's take the ratio of the two formulas:

$$\left(L = 4\pi d^2 b\right) / \left(\mathrm{L}_\odot = 4\pi\ \mathrm{d}_\odot^2 \mathrm{b}_\odot\right)$$

which gives us:

$$L / \mathrm{L}_\odot = \left(d / \mathrm{d}_\odot\right)^2 b / \mathrm{b}_\odot$$

Therefore, all we need to know to determine a star's luminosity is how far away it is compared with the Earth-Sun distance, given as $d/\mathrm{d}_\odot$, and how bright it is compared with that of the Sun, given as $b/\mathrm{b}_\odot$.

Example:

Let Star 1 be at half the distance of Star 2 and appear twice as bright. Compare the luminosities. First, $d_1/d_2 = 1/2$, also, $b_1/b_2 = 2$. Then:

$$\frac{L_1}{L_2} = \left(\frac{1}{2}\right)^2 \times 2 = 0.5$$

What this means is that Star 1 has only half the luminosity of Star 2, but it appears brighter because it is closer to us.

1.4. Magnitudes

Probably the first thing anyone notices when they glance up at the night sky is that the stars differ in brightness. There are a handful of bright stars, a few more are fairly bright and the majority are faint. This characteristic, the brightness of a star, is called the *magnitude*, of a star.[17]

---

[17] Actually, apparent magnitude can refer to any astronomical object and is not limited to just stars.

1.4.1 Apparent Magnitude

Magnitude is one of the oldest scientific classifications used today, devised by the Greek astronomer Hipparchus. Hipparchus classified the brightest stars as first-magnitude stars; those that were about half as bright as first-magnitude stars were called second-magnitude stars, and so on, down to sixth-magnitude, which were the faintest he could see. Today, we can see much fainter stars, and so the magnitude range is even greater, down to thirtieth-magnitude. Because the scale relates to how bright the stars appear to an observer on Earth, the term is more correctly called *apparent magnitude*,[18] and is denoted by m.

You have probably noticed by now that this is a confusing measurement because the brighter objects have smaller numerical values [e.g., a star of apparent magnitude +4 (fourth-magnitude) is fainter than a star of apparent magnitude +3 (third-magnitude)]. Despite its potential for causing confusion, apparent magnitude is used universally today; astronomers are happy with, but the rest of the world dislikes it intensely.

A further point is that the classification has undergone a revision since Hipparchus's day, and an attempt was made to put the scale on a scientific footing. In the nineteenth century, astronomers accurately measured the light from stars and were able to determine that a first-magnitude star is about 100 times brighter than a sixth-magnitude star, as observed from Earth. Or, to put it another way, it would take 100 sixth-magnitude stars to emit the light of one first-magnitude star.

The magnitude scale is very important and as we shall be using the magnitude system from this point on for every single object we discuss in the book, it's worthwhile looking at it in just a little greater detail.

A difference between two objects of one magnitude means that the object is about 2.512 times brighter (or fainter) than the other. Thus a first-magnitude object (magnitude $m = 1$) is 2.512 times brighter than a second-magnitude object ($m = 2$). This definition means that a first-magnitude star is brighter than a sixth-magnitude star by the factor of 2.512 raised to the power of 5. That is a 100-fold difference in brightness, and so a definition for the magnitude scale can be stated to be thus: a difference of five magnitudes corresponds exactly to a factor of 100 in brightness (see Table 1.2), thus

$$2.512 \times 2.512 \times 2.512 \times 2.512 \times 2.512 = (2.512)^5 = 100$$

The naked-eye limit of what you can see is about magnitude 5, in urban or suburban skies.[19] See Math Box 1.4 for an example of the use of the formula.

---

[18] The apparent part means this is how bright a star apparently looks, irrespective of its actual energy output, or whether it is close by or distant.

[19] Observers have reported that under excellent conditions, and with very dark skies, objects down to magnitude 8 can be seen with the naked eye.

Table 1.2 Magnitude and brightness ratio difference

Magnitude difference	Brightness ratio
0.0	1.0
0.1	1.1
0.2	1.2
0.3	1.3
0.4	1.45
0.5	1.6
0.7	1.9
1	2.5
2	6.3
3	16
4	40
5	100
7	630
10	10,000
15	1000,000
20	100,000,000

Table 1.3 The 20 brightest stars in the sky

	Star	Apparent magnitude	Constellation
1	Sirius	-1.44_{v}[a]	Canis Major
2	Canopus	-0.62_{v}	Carina
3	Alpha C.	−0.28	Centaurus
4	Arcturus	-0.05_{v}	Boötes
5	Vega	0.03_{v}	Lyra
6	Capella	0.08_{v}	Auriga
7	Rigel	0.18_{v}	Orion
8	Procyon	0.40	Canis Minor
9	Achernar	0.45_{v}	Eradinus
10	Betelgeuse	0.45_{v}	Orion
11	Hadar	0.61_{v}	Centaurus
12	Altair	0.76_{v}	Aquila
13	Acrux	0.77	Crux
14	Aldebaran	0.87	Taurus
15	Spica	0.98_{v}	Virgo
16	Antares	1.05_{v}	Scorpius
17	Pollux	1.16	Gemini
18	Fomalhaut	1.16	Piscis Austrinus
19	Becrux	1.25_{v}	Crux
20	Deneb	1.25	Cygnus

[a] Many stars are variable, so the value for their apparent magnitude will change. The suffix v indicates a variable star, and the value given is the mean value

Using this modern scale, several objects now have negative magnitude values. *Sirius*, the brightest star in the sky, has a value of −1.44 m, *Venus* (at brightest) is −4.4 m, the full *Moon* is −12.6 m, and the *Sun* is −26.7 m. Table 1.3 shows the 20 brightest stars.

Math Box 1.4: Apparent Magnitude and Brightness Ratio

Consider two stars, s_1 and s_2, which have apparent magnitudes m_1 and m_2 and brightness's b_1 and b_2, respectively. The relationship between them can be written as:

$$m_1 - m_2 = -2.5\log\left(\frac{b_1}{b_2}\right)$$

What this means is that the ratio of their apparent brightness's (b_1/b_2) corresponds to the *difference* in their apparent magnitudes ($m_1 - m_2$).

Example:

Sirius A has a magnitude of −1.44, while the Sun has a magnitude of −26.8. The ratio of their brightness's is thus:

$$m_1 - m_2 = -2.5\log\left(\frac{b_1}{b_2}\right)$$

$$-1.44-(-26.8) = -2.5\log\left(\frac{b_{sirius}}{b_{sun}}\right)$$

$$-10.21 = \log\left(\frac{b_{sirius}}{b_{sun}}\right)$$

$$\left(\frac{b_{sirius}}{b_{sun}}\right) \sim 10^{-10.1} = 7.9\times10^{-11} = 1/1.32\times10^{10}$$

Thus, Sirius appears 13,200,000,000 times fainter than the Sun, even though it is actually more luminous; remember, it is also more distant.

1.4.2 Absolute Magnitude

However, no matter how useful the apparent magnitude is scale is, it doesn't actually tell us whether a star is bright because it is close to us or faint because it is small or distant; all that this classification tells us is the apparent brightness of the star—that is, the star's brightness as observed visually, with the naked eye or telescope.

A more precise definition is the *absolute magnitude*, *M*, of an object, defined as the brightness an object would have at a distance of 10 parsecs. This is an arbitrary distance, derived from stellar parallax, the technique mentioned earlier; nevertheless, it does quantify the brightness of stars in a more rigorous way. See Math Box 1.5 for an example of the use of the formula.

As an example, Deneb, a lovely star of the summer sky, in the constellation Cygnus, has an absolute magnitude, *M*, of −8.73, making it one of the intrinsically brightest stars, while Van Biesbroeck's star has a value of *M* of +18.6, making it one of the intrinsically faintest stars known.

Naturally, the preceding discussion of magnitudes assumes that one is looking at objects in the visible part of the spectrum. It won't come as any surprise to know that there are several further definitions of magnitude that rely on the brightness of an object when observed at a different wavelength, or waveband, the most common being the U, B and V wavebands, corresponding to the wavelengths 350, 410 and 550 nm, respectively.

Furthermore, there is also a magnitude system based on photographic plates: the photographic magnitude, m_{pg}, and the photovisual magnitude, m_{pv}. Finally, there is the bolometric magnitude, $m_{\mathrm{BOL},}$ which is the measure of all the radiation emitted from the object.

From this point forward in the book, wherever we refer to the "magnitude" of an object its apparent magnitude is meant, unless stated otherwise.

Math Box 1.5: Relationship Between Apparent Magnitude and Absolute Magnitude

The apparent magnitude and absolute magnitude of a star can be used to determine its distance, the formula for which is:

$$m - M = 5 \log d - 5$$

where

m = the star's apparent magnitude.
M = the star's absolute magnitude.
d = the distance to the star (in parsecs).

The term $m - M$ is referred to as the *distance modulus*.

Example:

Sirius is at a distance of 2.63 parsecs and has an apparent magnitude of −1.44. Its absolute magnitude can be calculated thus:

$$m - M = 5 \log d - 5$$

$$\mathrm{M} = \mathrm{m} - 5 \log d + 5$$

$$-1.44 - 5 \log(2.63) + 5$$

$$\mathrm{M} \sim 1.46$$

Thought Question 1.1

Sirius emits 23 times more energy than the Sun, but why does the Sun appear much brighter in the daytime sky?

1.5. The Visually Brightest Stars

Below is a list of some of the brightest stars in the sky. It is of course by no means complete. For those interested in observing additional bright stars, check out the sister volume to this book—*Field Guide To Deep Sky Objects*.

Several of the brightest stars will have already been mentioned earlier. For the sake of clarity and space, they will not be repeated here, but there is one caveat. There are several disparate lists of the brightest stars that can be found on the Internet and in various books. With new measuring techniques and observations, the lists are always in a constant state of change, and stars are being added or removed. This list is as accurate as can be for summer 2014. It will change!

Pollux	β Gem	07h 45.3m	+28° 02′	January
1.15 m	1.09 M	33.72 l.y.		Gemini

This is the brighter of the two famous stars in Gemini, the other of course being Castor. It is also, however, the less interesting. It has a ruddier color than its brother and is the bigger star.

Becrux	β Crucis	12h 47.7m	−59° 41′	April
1.30[20]vm	−3.92 M	352.1 l.y.		Crux

This star lies in the same field as the glorious Jewel Box star cluster. It is a pulsating variable star with a very small change in brightness. It does however lie too far south for northern hemisphere observers.

Spica	α Virginis	13h 25.2m	−11° 10′	April
1.04vm	−3.55 M	262 l.y.		Virgo

The fifteenth-brightest star is a large spectroscopic binary with the companion star lying very close to it and thus eclipsing it slightly. Spica is also a pulsating variable star, though the variability and pulsations are not visible with amateur equipment.

Hadar	β Centauri	14h 03.8m	−60° 22′	April
0.6vm	−5.45 M	525 l.y.		Centaurus

---

[20] Denotes that the star, and thus the magnitude, is variable.

This is the eleventh-brightest star in the sky, and it is unknown to northern observers because of its low latitude (lying as it does only 4.5° from Alpha [α] Centauri). It has a luminosity that is an astonishing 10,000 times that of the Sun. A white star, it has a companion of magnitude 4.1, but it is a difficult double to split, as the companion is only 1.28 arc seconds from the primary.

Arcturus	α Boötis	14$^{\mathrm{H}}$ 15.6$^{\mathrm{M}}$	+19° 11′	April
−0.04vm	−0.10 M	36.7 l.y.		Boötes

The fourth-brightest star in the sky, and the brightest star north of the celestial equator, having a lovely orange color. Notable for its peculiar motion through space, Arcturus, unlike most stars, is not traveling in the plane of the Milky Way, but is instead circling the galactic center in a highly inclined orbit. Calculations predict that it will swoop past the Solar System in several thousand years' time, moving towards the constellation Virgo. Some astronomers believe that in as little as half a million years Arcturus will have disappeared from naked-eye visibility. At present, it is about 100 times more luminous than the Sun.

Rigil Kentaurus	α Centauri	14$^{\mathrm{H}}$ 39.6$^{\mathrm{M}}$	−60° 50′	April
−0.1 m	4.07 M	4.39 l.y.		Centaurus

The third-brightest star in the sky, this is in fact part of a triple system, with the two brightest components contributing most of the light. The system contains the closest star to the Sun, Proxima Centauri. The group also has a very large proper motion (its apparent motion in relation to the background). Alas, it is too far south to be seen by any northern observer. However observers have claimed that the star is visible in the daylight with any aperture. Note that the magnitude value of −0.1 is the value for the combined magnitudes of the double star system.

Antares	α Scorpii	16h 29.4m	−26° 26′	May
1.09vm	−5.28 M	604 l.y.		Scorpius

This is a red giant star, the sixteenth brightest in the sky, with a luminosity 6000 times that of the Sun, and a diameter hundreds of times larger than the Sun's. But what makes this star especially worthy of observation is the vivid color contrast that is seen between it and its companion star, often described as vivid green when seen with the red of Antares. The companion has a magnitude of 5.4, with a PA of 273°, lying 2.6″ away.

Vega	α Lyrae	18h 36.9m	+38° 47′	July
0.03vm	0.58 M	25.3 l.y.		Lyra

The fifth-brightest star, very familiar to northern observers, located high in the summer sky. Although similar to Sirius in composition and size, Vega is three times as distant, and thus appears fainter. Often described as having a steely-blue color, it was one of the first stars observed to have a disc of dust surrounding it—a possible proto-solar system in formation. Vega is one of those stars that has been the object of much research and has in fact been called “arguably the next most important star in the sky after the Sun.” It was the first star other than the Sun to be photographed, the first to have its spectrum recorded, and one of the first stars whose distance was estimated through parallax measurements. Additionally, it served as the baseline for calibrating the photometric brightness scale,[21] as well as one of the stars used to define the mean values for the UBV photometric system. Vega was the Pole Star some 12,000 years ago and will be again in another 12,000 years.

Altair	α Aquilae	19h 50.8m	+08° 52′	July
0.77vm	2.20 M	16.77 l.y.		Aquila

The twelfth-brightest star, Altair has the honor of being the fastest-spinning of the bright stars, completing one revolution in approximately 6½ h. Such a high speed deforms the star into what is called a flattened ellipsoid, and it is believed that because of this amazing property, the star may have an equatorial diameter twice that of its polar diameter. The star’s color has been reported as completely white, although some observers see a hint of yellow.

Deneb	α Cygni	20h 41.4m	+45° 17′	August
1.25vm	−8.73 M		2600 l.y.	Cygnus

The nineteenth-brightest star is very familiar to observers in the northern hemisphere. This pale-blue supergiant has recently been recognized as the prototype of a class of non-radially pulsating variable stars. Although the magnitude change is very small, the time scale is from days to weeks. It is believed that the luminosity of Deneb is some 60,000 times that of the Sun, with a diameter 60 times greater. Its distance is the subject of much debate, as previous estimates were widely inaccurate. Nevertheless, it is the brightest and most distant of the stars that have an apparent magnitude brighter than 1.5, and the most distant (by a factor of almost 2) of the 30 brightest stars.

Fomalhaut	α Pisces Austrini	22h 57.6m	−29° 37′	September
1.16 m	1.74 M	25.07 l.y.		Pisces Austrinus

---

[21] This is no longer the case, as the apparent magnitude zero point is now commonly defined in terms of a particular numerically specified energy output. A far more convenient approach for astronomers, as Vega is not always available for calibration.

The eighteenth-brightest star is a white one, which often appears reddish to northern observers due to the effect of the atmosphere. It lies in a barren area of the sky and is remarkable only in that a star close to it, which is not bound gravitationally, yet lies at the same distance from Earth, is moving through space in a manner and direction similar to Fomalhaut's. It has been suggested that the two stars are remnants of a star cluster or star association that has long since dispersed. The companion (?) star is an orange 6.5-magnitude object about 2° south of Fomalhaut.

Achernar	α Eridani	01h 37.7m	−57° 14′	October
0.50vm	−2.77 M	144 l.y.		Eridanus

The ninth-brightest star in the sky lies too far south for northern observers, at the southernmost end of the constellation. Among the bright stars, it is one of the very few that have the designation "p" in its stellar classification, indicating that it is a "peculiar" star.

Aldebaran	α Tauri	04h 35.9m	+16° 31′	November
0.85 m	−0.63 M	65.11 l.y.		Taurus

The fourteenth-brightest star appears to be located in the star cluster Hyades; however, it is not physically in the cluster at all, lying as it does twice as close as the cluster members. This pale-orange star is approximately 120 times more luminous than the Sun. It is also a double star but a very difficult one to separate due to the extreme faintness of the companion. The companion star, a red dwarf, magnitude 13.4, lies at a PA of 34° at a distance of 121.7″.

Rigel	β Orionis	05h 14.5m	−08° 12′	December
−0.12vm	−6.69 M	773 l.y.		Orion

The seventh-brightest star in the sky, Rigel is in fact brighter than Alpha (α) Orionis. This supergiant star is one of the most luminous stars in our part of the galaxy, almost 560,000 times more luminous than our Sun but at a greater distance than any other nearby bright star. Often described as a bluish star, it is truly tremendous, with about 50 times the mass of the Sun and about 50 times the diameter. It has a close bluish companion at a PA of 202° at an apparent magnitude of 6.8 and a distance of 9 arc seconds, which should be visible with a 15 cm telescope, or one even smaller under excellent observing conditions.

Capella	α Aurigae	05h 16.7m	+46° 00′	December
0.08vm	−0.48 M	42 l.y.		Auriga

High in the sky in winter, it has a definite yellow color, reminiscent of the Sun's own hue. It is in fact a spectroscopic double and is thus not split in a telescope; however, it has a fainter tenth-magnitude star about 12 arc seconds to the southeast, at a PA of 137°. This is a red dwarf star, which in turn is itself a double (only visible in larger telescopes). Thus, Capella is in fact a quadruple system.

Betelgeuse	α Orionis	05h 55.2m	+07° 24′	December
0.58vm	−5.14 M	427 l.y.		Orion

The tenth-brightest star in the sky, and a favorite among observers, this orange-red star is a giant variable, with an irregular period. Recent observations by the Hubble Space Telescope have shown that it has features on its surface that are similar to sunspots but much larger, covering perhaps a tenth of the surface. It also has a companion star, which may be responsible for the non-spherical shape it exhibits. Although a giant star, it has a very low density and a mass only 20 times greater than the Sun's, which together mean that the density is in fact about 0.000000005 that of the Sun! A lovely sight in a telescope of any aperture, subtle color changes have been reported as the star goes through its variability cycle.

Thought Question 1.2

Will stars that appear brighter in the sky have larger or smaller magnitudes than fainter stars?

1.6. The Colour of Stars

When we look up into the sky, we see many stars, all of the same general color, usually white. There are, of course, a few that exhibit a distinct color—Betelgeuse (α Orionis) is most definitely red, as is Antares (α Scorpi). Capella (α Aurigae) is yellow, and Vega (α Lyrae) is steely blue. But for the most part, there does not seem to be any great variation in color. Look through binoculars or a telescope, however, and the situation changes dramatically.[22] Variations in color and hue abound![23]

---

[22] The eye does not recognize color at low light levels. This is why at night, with the naked eye, we see only shades of gray, white and black.

[23] The most important factor determining the color of a star you see is you—the observer! It is purely a matter of physiological and psychological influences. What one observer describes as a blue star another may describe as a white star; or one may see an orange star, while another observes the same star as yellow. You might even observe a star to have different color when using different telescopes or magnifications, and atmospheric conditions will certainly have a role to play.

The color of a star is determined by its surface temperature. A red star has a lower temperature than a yellow star, which in turn has a lower temperature than a blue star. This is an example of what is called *Wien's law* (See Math Box 1.6). The law shows that low-temperature stars emit most of their energy in the red to infrared part of the spectrum, while much hotter stars emit in the blue to ultraviolet part of the spectrum. Some very hot stars emit most of their energy in the ultraviolet, so in fact we only see a fraction of their light. Furthermore, many stars emit nearly all of their light in the infrared, so we do not see them at all. Surprisingly, these low-mass (discussed later), low-temperature stars make up about 70% of the stars in our galaxy, but you would never know this by going out and observing on a clear night; we just cannot see them with our eyes alone.

An important point to notice here is how hotter objects emit more energy at *all* wavelengths due to the higher average energy of all the photons. This is illustrated in Fig. 1.5. The graphs show how the light from three different stars is distributed, depending on the stars' temperature. The colored block represents the visible part of the spectrum. The first plot shows the light that would be measured from a colored star of about 3000 K. Note that the curved line peaks at about 1100 nm, which would make the star appear red. The second plot shows a star at about 5500 K (similar to the Sun's temperature) and peaks in the middle of the visible spectrum, thus looking yellowish. The final plot illustrates a very hot star, at 25,000 K; it peaks at about 400 nm, so this star will appear blue. Thus, a star's color, from an astronomical viewpoint, depends on where the peak of the curve is; short wavelengths (to the left part of the plot) indicate a hot, blue-white star, while longer wavelengths (the right part of the plot) indicate a cool, reddish-orange star. The Sun actually peaks in the green part of the spectrum, but because there is a mixture of light from all the other parts of the visible spectrum—the blues, reds, and yellows—we actually observe the Sun as being yellowish-white.

An interesting observation is that a few stars are so hot, possibly in the millions of degrees, they emit their energy at very short wavelengths. In fact, they radiate X-rays. These are neutron stars!

Note that when we speak of a star's temperature, we are referring to its surface temperature. The internal temperature cannot be measured directly, and it is usually determined from theoretical temperatures. So when you read that a star's temperature is "25,000 K," this refers to the surface temperature.[24]

---

[24] From here on, when we mention temperature, we are referring to the surface temperature, unless indicated otherwise.

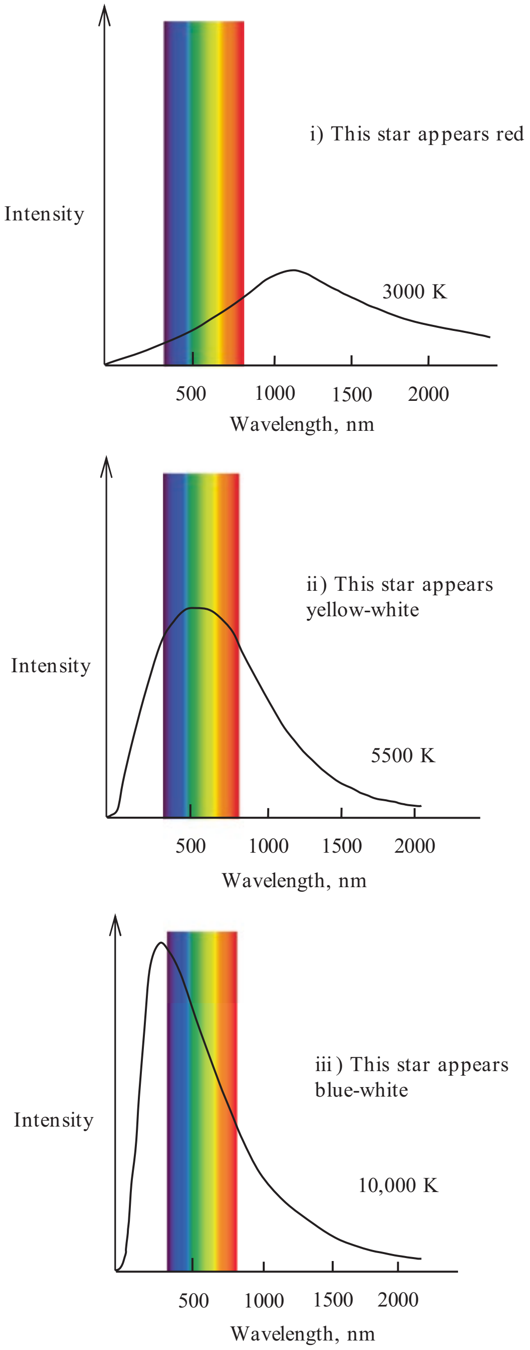


Fig. 1.5. Color and temperature

Math Box 1.6: Wien's Law

This can be stated as:

$$\lambda_{\text{max}} = \frac{2{,}900{,}000}{\text{T (Kelvin)}} \text{ nm}$$

Example:

Two stars, α Canis Majoris and o Ceti, have a temperature of 9200 K and 1900 K, respectively. What are their peak wavelengths?

$\lambda_{\text{max}} = \frac{2{,}900{,}000}{9200\text{(Kelvin)}}\text{nm} = 315 \text{ nm}$ (i.e., in the ultraviolet[25])

And

$\lambda_{\text{max}} = \frac{2{,}900{,}000}{1900 \text{ (Kelvin)}} \text{ nm} = 1526 \text{ nm}$ (i.e., in the infrared[26])

Thus, α Canis Majoris emits a LOT of light in the ultraviolet, even though it shines brightly white.

$\lambda_{\text{max}} = \frac{2{,}900{,}000}{1900 \text{ (Kelvin)}} \text{ nm} = 1526 \text{ nm}$ (i.e., in the infrared[27])

Thus, α Canis Majoris emits a LOT of light in the ultraviolet, even though it shines brightly white.

A scientific description of a star's color is based on the stellar classification, which in turn is dependent upon the star's chemical composition and temperature. A term commonly used by astronomers is the *color index*. This is determined by observing a star through two filters, the B and V filters, which correspond to wavelengths of 440 nm and 550 nm, respectively, and by measuring its brightness. Subtracting the two values obtained, B−V, produces the color index. Usually, a blue star will have a negative color index (e.g., −0.3); orange-red stars could have a value greater than 0.0 and upwards to about 3.00 and greater for very red stars (Magniture 6[28] and greater).

Having discussed the colors of stars, let's now look at some examples. Following is a representative selection of bright stars. There are, of course,

---

[25] This star is the brightest in the night sky. It is, of course, Sirius.

[26] This is the most famous irregular variable star, Mira.

[27] This is the most famous irregular variable star, Mira.

[28] The spectral classification of a star is covered in a later chapter.

literally thousands of other colored stars that are visible. Also, the stars listed earlier, the brightest stars, offer many examples of stars exhibiting distinct colors. In addition, many double stars (not mentioned here) show very distinctly colored hues and tints. The nomenclature used here is the same as that used previously, with the addition of the star's temperature and color.[29]

Bellatrix	γ Ori	05h 25.7m	+06° 21′	Nov-Dec-Jan
1.64 m	−2.72 M	21,450 K	Blue	Orion

Also known as the Amazon Star, Bellatrix is a very steely-blue color. Some observers report a faint nebulosity associated with the star, but this may just be part of the general nebulosity that envelops much of Orion.

Merope	23 Tau	03h 46.3m	+23° 57′	Oct-Nov-Dec
4.14 m	−1.07 M	10,600 K	Blue	Taurus

Located within the Pleiades star cluster this is a breathtaking and spectacular view when seen through binoculars, and the cluster is a highlight of the night sky. Almost any of the stars in this cluster are worth looking at, as they are all a lovely steely-blue color.

Regulus	α Leo	10h 08.3m	+11° 58′	Jan-Feb-Mar
1.36 m	−0.52 M	12,000 K	Blue-white	Leo

Alpha (α) Leonis is the handle of the Lion's sickle. It is an easy double star, the companion, an eighth magnitude, orange-red color, about 3′ away.

Acrux	α Crucis	12h 26.6m	−63° 06′	Feb-Mar-Apr
0.77 $^{\text{v}}$ m	−4.19 M	28,000/26,000 K	White	Crux

Acrux is a multiple star system, but only its brightest members are seen as a double star, with components about 4″ apart. Both stars are about the same magnitude, 1.4 for α^1 and 1.9 for α^2. The colors of the stars are white and blue-white, respectively. An interesting aside is that in 2008, the Cassini-Huygens spacecraft managed to resolved three of the components (A, B and C) of the system as it was occulted by Saturn's disk.

Zubeneschamali	β Lib	15h 17.0m	−09° 23′	Apr-May-Jun
2.61 m	−0.84 M	11,000 K	Green!	Libra

A mysterious star for two reasons. Historical records state that it was much brighter than it appears today, and observers of the past 100 years

---

[29] Remember that a star's color is observer-dependent! What one person sees as yellow, another sees as white. Do not be surprised if you see a different color from that mentioned.

have declared that it is greenish or pale emerald in color. Thus it is one of the rare green-colored stars!

The Sun				Jan-Dec
−26.78 m	4.82 M	5800 K	Yellow	The Zodiac

Our closest star, and the object without which no life would have evolved on Earth. Visible every day throughout the year, unless you happen to live in the UK. DO NOT OBSERVE THE SUN THROUGH ANY KIND OF OPTICAL EQUIPMENT.

Hind's Crimson Star	R Leporis	04h 59.6m	−14° 48′	Nov-Dec-Jan
7.71vm	1.08 M	3000 K[30]	Red	Lepus

This star, a classic long-period variable, period about 432 days, varies in brightness between 5.5 and 11.7 m. At maximum brightness, it displays the famous ruddy color that gives it its name. Discovered in 1845 by J. R. Hind with a color described as "intense smoky red." This may be the reddest star in the sky, but that may only be due to the presence of carbon in its outer atmosphere that has the effect of filtering out the blue light. It is also an AGB star (see the later section on star death).

1.7. The Sizes of Stars

Stars lie at an immense distance from us, so no matter how much we magnify a star's image, it will, in all but a handful of cases,[31] remain just a point of light. So how do we determine the size of a star? The answer is quite simple. By measuring both the star's luminosity (derived from its distance and brightness) and its surface temperature (determined from its spectral type), it is just a matter of manipulating the numbers with a few formulas. Using this technique, astronomers have discovered that there are a plethora of stars much smaller than the Sun, while many are thousands of times larger.

To accurately determine a star's size, a physical law, called the *Stefan-Boltzmann law*, is used. We won't bother looking at how this law came about, but rather just quote it and show how it is used (see Math Box 1.7). What the law tells us is that the amount of energy that a star radiates per

---

[30] The real temperature of the star is still undetermined.

[31] A few stars, such as Betelgeuse, have had their radii determined by the technique known as interferometry. For the vast majority of stars, the technique is not applicable, either due to distance or faintness.

[32] To be accurate, the law refers to a blackbody, which is something that emits thermal radiation. Thus, thermal radiation is blackbody radiation. It can be applied to a star because, to all intents and purposes, a star's surface behaves like a blackbody.

second, from a square meter of its surface,[32] is proportional to the fourth power of the temperature, T, of the star's surface. Don't let the complexity of this statement distract you. It really just tells us that the *energy flux* (F) is related to the temperature, which makes sense when you think about it. A cool object has lower thermal energy than a hot object.

Now think back and recall that we discussed how the luminosity of a star is a measure of the energy emitted from the surface every second. This luminosity is in fact the flux, F, multiplied by the number of square meters there are on the star's surface. If we now assume that most stars are spherical (which is not as silly as it sounds because a few stars are not spherical!), then the quantity highlighted in the previous sentence is in fact the surface area of the star. This is given by a very simple formula, which most of us know: $4\pi R^2$, where R is the radius of the star (taken as the distance from the center of the star to its surface[33]); see Math Box 1.8.

Math Box 1.7: Flux, Luminosity and Radius

The flux from a star is given by the Stefan-Boltzmann law:

$$F = \sigma T^4$$

The relationship between the flux, F, luminosity, L, and radius, R, of a star is:

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

where

L is the star's luminosity in watts (W).
R is the star's radius in meters (m).
σ is the Stefan-Boltzmann constant; 5.67×10^{-8} W m^{-2} K^{-4}.
T is the star's temperature in Kelvin (K).

What the above equations tell us is that a coolish star (that is, one that has a low surface temperature, T) will have a low flux, but it may be quite luminous because it could have a very large radius, and thus a large surface area. In a similar vein, a hot star (with a high temperature) can have a low luminosity if it has a small radius, which would mean a low surface area. Now you can see that knowing a star's temperature alone does not indicate how luminous it will be—we also need its radius!

Although we can now determine such parameters as the radius, temperature, luminosity and brightness of a star, it is often more useful to relate these values to that of the Sun. It is easier to have a mental picture of a star if we say it is about 10 times hotter than the Sun than if we say it is 54,000 K. The same applies for L and R.

---

[33] No doubt some of you are already asking, "Where is the surface of a star? A star is made of gas!" Fear not…all will be revealed in later chapters.

Math Box 1.8: Even More About Flux, Luminosity and Radius

We regard the Sun as a typical star, so we can relate most of another star's characteristics to that of the Sun. For instance:

$$\mathrm{L}_{\odot} = 4\pi\ \mathrm{R}_{\odot}^{2}\sigma\ \mathrm{T}_{\odot}^{4}$$

where

$\mathrm{L}_{\odot}$ is the Sun's luminosity.
$\mathrm{R}_{\odot}$ is the Sun's radius.
$\mathrm{T}_{\odot}$ is the Sun's temperature.

If we now divide the luminosity equation for a star by that for the Sun, we get:

$$\mathrm{L} / L_{\odot} = \left(R / \mathrm{R}_{\odot}\right)^{2} \left(T / \mathrm{T}_{\odot}\right)^{2}$$

The constants, σ and the 4π, have now gone, and we can also rearrange the formula to read:

$$R / \mathrm{R}_{\odot} = \left(\mathrm{T}_{\odot} / T\right)^{2} \left(L / \mathrm{L}_{\odot}\right)^{1/2}$$

where the 1/2 factor indicates a square root.

Now, $R/\mathrm{R}_{\odot}$ is the ratio of the star's radius to that of the Sun's.
$\mathrm{T}_{\odot}/T$ is the ratio of the Sun's temperature to that of the star's.
$L/\mathrm{L}_{\odot}$ is the ration of the star's luminosity to that of the Sun's.

Example:
Sirius has a temperature of about 9200 K and has a luminosity of about 23 $\mathrm{L}_{\odot}$.
To determine its radius ratio:

$$R / \mathrm{R}_{\odot} = \left(\frac{5800}{9200}\right)^{2} \times \sqrt{23} \sim 2$$

Thus, its radius is about twice that of the Sun.

Let's now look at some examples of giant stars, particularly those that can be seen with the naked eye. Note that $\mathrm{R}_{\odot}$ is the radius of the Sun.

α HERCULIS	ADS 10418	17H 14.6M	+14° 23′	MAY-JUN-JUL
2.9V, 5.4VM	−2.30 M	387$\mathrm{R}_{\odot}$		HERCULES

A lovely color-contrast double: orange and bluish-green the star lies at a distance of about 400 light years and is a semi-regular, super-giant variable star. The primary star is itself variable, while the secondary is an unresolvable double.

ψ^1 Aurigae	HD 44537	06h 24.9$^{\mathrm{M}}$	+49° 17′	Nov-Dec-Jan
4.91vm	−5.43 M	637$R_{\odot}$		Auriga

This star has an incredible luminosity of over 63,600 $L_{\odot}$. It is an irregular variable star, the diameter of which is still not accurately known. It is believed to be about 4300 light years distant.

VV Cephei	HD 208816	21h 56.6m	+63° 37′	Sep-Oct-Nov
4.91 m	−6.0 M	1050–1900$R_{\odot}$		Cepheus

This star has a luminosity of 275,000–575,000 $L_{\odot}$, and it lies at a distance of 2000 l.y. It is one of the famous *eclipsing binary*-type variable stars, with a period of just over 20 years. The system consists of an O-type dwarf and an M-type supergiant and if placed at the center of our Solar System, this giant star would extend to the orbit of Saturn!

Garnet Star	μ Cep	21h 43.5m	+58° 47′	Jul-Aug-Sep
4.08vm	−7.3 M	650–1420$R_{\odot}$		Cepheus

Located on the northeastern edge of the nebulosity IC1396, the Garnet Star, named by William Herschel, is one of the reddest stars in the sky, having a deep orange or red color seen against a backdrop of faint white stars. It is a pulsating red giant star, with a period of about 730 days, varying from 3.4 to 5.1 m. This is a runaway star with a velocity of 80.7 km/s. The distance to Mu Cephei is not very well known, but latest research places it 6000 light years away. Since 1943, the stars' spectrum has served as one of the stable anchor points by which other stars are classified.

VY Canis Majoris	HD 58061	07h 23$^{\mathrm{M}}$	−25° 46′	Dec-Jan-Feb
6.5–9.6 m	−5.25 M	1420$R_{\odot}$		Canis Major

Lying at a distance of about 3900 light years from us, this is a semi-regular variable star. However, this star is notorious for the controversy that surrounds it. One school of thought suggests that the star is a very large and very luminous red hypergiant,[34] but the many estimates of such a large size and luminosity currently lie outside the bounds of stellar theory, both beyond the maximum predicted size of any star and far cooler than a star of its luminosity can become. Another idea is that the star is a normal red supergiant with a radius around 600 $R_{\odot}$—falling comfortably inside models

---

[34] We will discuss these stars in a later chapter.

of stellar structure and evolution. And it doesn't stop there! The star illustrates the problem of defining the surface and radius of a star when its average density 100,000 times less dense than that of Earth's at sea level. Where does the star end and space begin? We may not have heard the last from this particular star.

Note that there are also other large stars—Arcturus, Alderbaran, Rigel, Antares, Betelgeuse—all visible with the naked eye. But none of these are as large as the above behemoths.

Thought Question 1.3

Polaris has the same surface temperature as the Sun, with a diameter around 100 times that of the Sun, thus it has a luminosity of 10,000 $L_\odot$. *It also has a magnitude ~2. What does all this information tell us about Polaris?*

1.8. The Constituents of Stars

Although we will be covering this topic in far greater detail later in the book, it is important that we at least discuss briefly here what stars are made of.

A star is an enormous sphere of hot gas. It is as simple, or as complex, as that, whichever way you wish to look at it. Of course, the processes involved in making and maintaining a star are, as to be expected, very, very complex!

The gases composing a star are primarily hydrogen (H, the most common element in the universe), helium (He), and then some other elements.[35] By and large, most stars are nearly all hydrogen, less helium, and very small amounts of everything else. The mix is usually about 75% hydrogen, 24% helium and the remainder metals. This figure changes, however, when we discuss either very, very old stars, which are nearly all hydrogen and helium with tiny amounts of metals, or recently born stars, which can contain as much as 2–3% metals.

The energy needed to create and then maintain a star is formed by nuclear fusion; hydrogen is converted to helium due to two immense forces at work, namely, a very high temperature and very strong gravitational force. Due to its very large mass and concomitant strong gravitational field, conditions in the center of the ball of gas are such that the temperature can be about 10 *million* K. At such extremes of pressure and heat, nuclear fusion can occur, and hydrogen is converted into helium in a self-sustaining reac-

---

[35] Astronomers call every element other than hydrogen and helium a metal. It's odd, but don't worry about it—just accept it.

tion. The outcome of this nuclear reaction is a tiny amount of energy, in the form of gamma rays. It may not seem like much, but when you consider that billions of these reactions take place every second, the amount of energy liberated is quite substantial…enough, in fact, to make a star shine!

As the star ages, it uses up more and more hydrogen in its core, in order to keep the nuclear reactions going. A byproduct of this reaction is helium. Thus, as time passes, the amount of hydrogen in the core decreases and the helium increases. If conditions are right (these include a higher temperature and a large mass), then the helium itself will start to undergo nuclear fusion at the star's core. After a certain amount of time, this, in turn, will produce the element carbon as a byproduct of the reaction; again, if conditions are suitable, this, too, will initiate nuclear fusion and produce more energy. An important point to emphasize is that each step requires a higher temperature to begin the nuclear reactions, and if a star does not have the conditions necessary to produce this temperature, further reactions will not occur. So, the "burning" of hydrogen and helium is the power source for nearly all the stars you can see, and a star's mass determines how the reaction will proceed. We will return to these concepts in the following chapters, as we cover the lives, and deaths, of stars.

1.9. Telescope Basics

You may be wondering why a book on astrophysics has a section on telescopes. Well, the answer is simple; a large part of the book is devoted to observing many of the objects under discussion, so a brief precise of some telescope basics would be appropriate.

I am not going to spend any time discussing the many disparate types of telescopes, or the discussion about which type is better suited to which observation. Rather, I will cover those topics that I think are the most relevant.

I assume that many of the readers of this book will have some background knowledge on a few aspects of amateur astronomy, even if only a fleeting acquaintance. Nevertheless, I still believe it is worthwhile mentioning a few important subjects that everyone who wishes to be an accomplished observer should be familiar with (even if some of them may appear trivial).

These are.

(i) Magnification
(ii) Resolution
(iii) Limiting Magnitude
(iv) Field of View

(v) Atmospheric Effects
  (a) Transparency
  (b) Seeing
  (c) Light Pollution
(vi) Dark Adaption and Averted Vision

1.9.1 Magnification

I am sure that most of us are familiar with a story like this. Someone decides to "take up astronomy" and spends a considerable amount of money on a telescope. It has superb optics, computer control, and comes with three eyepieces. Then a bright supernova is reported in the faint galaxy Messier 33 (NGC 598), in the constellation Triangulum, and the budding astronomer rushes out into the evening, sets up the telescope and using the highest magnification tells the computer to GOTO M33. Nothing is visible. Not a glimmer. The telescope ends up back in the shop, along with its irate owner.

Well, maybe that's just a *little* bit of an exaggeration. But who among us has not been so eager to try out some new piece of equipment, or view an exciting object, that we rush out and try to observe with an inappropriate magnification, only to be disappointed when the image doesn't live up to our expectations?[36] The purpose of this section is to explain how magnification works, and when best to use a certain magnification for a particular object, or when circumstances warrant it.

The topic of magnification can be a confusing one for newcomers to amateur astronomy. What is the best time to use high magnification, or low? Why doesn't higher magnification split close doubles? Why do some extended objects seem to get fainter at higher magnification, while others seem to get brighter with low magnification? This section will help clarify these and other points.

A full description of the physical process of magnification is beyond the scope of this book, but a few details are appropriate.[37]

The magnification of a telescope (or binoculars) is given by a simple formula: the focal length of the objective lens (or mirror), f_o, divided by the focal length of the eyepiece, f_e. This is usually written as.

$$magnification = \frac{f_o}{f_e}$$

---

[36] I plead guilty!

[37] A list of several books that deal with the subject in more detail can be found in Appendix A.

An example will be useful.

A telescope with a mirror of focal length 1000 mm, used with an eyepiece of focal length 25 mm, gives a magnification of 1000/25 = 40, thus the magnification of the telescope using this eyepiece is 40 times, sometimes written as 40X. In this way, a selection of eyepieces of differing focal lengths provides several different magnifications.

It may be surprising to learn that there is a minimum magnification that can be usefully used. This is sometimes quoted as $M \geq 1.7D$, where D is the diameter of the telescope in cm. What this means is that there is a minimum magnification at which all of the light from the telescope passes into your eye. If you use a lower magnification some of the light is wasted as it is spread out over an area larger than the pupil of your eye.

Here's another example.

If you have a telescope of diameter 20 cm, then the minimum magnification is $1.7 \times 20 = 34$. Thus, using the eyepiece from the example above will give a magnification of 40X, which is appropriate for this optical system.

It comes as less of a surprise that there is also a limit to the *highest* useful magnification that can be employed. In the past, advertisements for telescopes would quote ridiculously high values for magnifications, often several hundred times. While it is true to say that these magnifications are theoretically possible, in practice they are, in a word, useless.

Although there is no hard-and-fast rule for the limit of highest magnification, a good rule of thumb is that the highest power, on average, should be from about $10D$ to $20D$, where D is the diameter of the primary mirror (or lens) in centimetres. For a 20 cm telescope this would result in magnification from 200X to 400X. Such high magnifications are however, rarely used, as they suffer from the following drawbacks:

(i) A smaller field of view.
(ii) A decrease in brightness of extended or non-stellar objects.
(iii) An exaggeration of atmospheric defects.
(iv) An exaggeration of any tremors or defects in the mount or drive system.

Usually, most astronomers have a minimum of three eyepieces that provide a good range of magnifications. These are:

(i) Low power, $2D$ to $3D$—this shows the largest amount of sky.
(ii) Medium power, $5D$ to $8D$—used for more general observations.
(iii) High power, 10D to 20D—useful for double star work, and detailed planetary observation.

Most people who are not familiar with observing would expect an object like a nebula to be brighter when viewed through a telescope than when seen with the naked eye. It's usually a surprise that this is not the case.[38]

Basically, because of light losses and other effects, the brightness of a nebula—or any other extended deep-sky object—are fainter when seen through the telescope than if viewed by the naked eye! But what a telescope does is increase the apparent size of a nebula from an inappreciable to appreciable extent. And the background sky appears darker through a telescope than when seen with the naked eye. Too high a magnification, however, spreads the light out to such an extent as to make any detailed observation suspect.

Finally in this section on magnification, I should mention that there is a minimum magnification that is needed, if you are to *resolve* all the detail that your telescope is capable of achieving. Although the section on resolution comes later. I think it is appropriate to mention this aspect here.

It's probably easiest to discuss this point by using an example. Take for instance a close double star. You know from the telescope's handbook (or you may have calculated for yourself) that the resolution of your telescope has a certain value. You see in the section on double stars in this book that your telescope is capable of splitting these stars. However, when you observe them, instead of seeing two separate and distinct stars, you see instead just one star, or maybe an elongated blur. This may be because the magnification is too low; although the double star should be resolvable by the objective lens (or mirror), the two components will not be observed as individual stars unless a high magnification is used in order to bring them above the resolving threshold of the human eye (about 2 to 3 minutes of arc).

Ignoring for the moment atmospheric effects and other considerations which limit resolution, this magnification is given a value anywhere from 10 to 16 times the telescope's aperture in centimetres. A ballpark number is 13 times the aperture in centimetres. Thus, to split very close double stars and to resolve detail close to your resolution limit, you need not only superb optics, good weather and so on, but also, on occasion, a high magnification.

Remember, however, that you can never increase the resolution by increasing magnification ad infinitum. As you will see in the next section, the resolution of your telescope is constrained both by the size of the primary mirror or lens and by the physical nature of light itself.

---

[38] Note that this discussion does not apply to point sources, i.e. stars!

1.9.2 Resolution

The topic of resolution is extremely theoretical, and a full description of the theory would be better suited to an undergraduate textbook in astrophysics.[39] Not surprisingly, it is also confusing for many amateur astronomers (and even a few professional astronomers), as more often than not, there are few books specifically written for the amateur which describe the Rayleigh resolution, the Dawes limit, the resolving power, the Airy disc, and so on. With this in mind I will not bother to explain where the theory and formulas come from, but just write them down without explanation as to their derivation.

Let's begin our foray into the area of telescopic resolution by starting with some simple theory. You might expect that stars which appear as incredibly small points of light to the naked eye because of their immense distance from us would, when magnified, still appear as small points of light—but observation tells us otherwise. The image of a star, when at the focus of a telescope, appears as a finite—although very small—disc of light.[40] This is called the *Airy disc*. In fact, the Airy disc represents about 83% of the star's light. The remaining 17% can often be seen as faint diffraction rings around the star's image.

This is the first counter-intuitive result: no matter how big a telescope you have, how perfect the optics, or how high a magnification you use, not all of the star's light goes into making the central image. This is a consequence of the wave nature of light.

The normally accepted definition of the theoretical resolution of a telescope is given by what is called the *Rayleigh criterion*, denoted by α and given by the formula.

$$\alpha = 1.22\frac{\lambda}{D}\left(radians\right)$$

where λ is the wavelength of light and D is the diameter in metres of the lens or mirror. Note however, that the unit of measurement for this definition is the radian—one that strikes terror into most people!

However, if we assume that the wavelength of light, λ, is about 500 nanometres,[41] a perfectly acceptable value when using the telescope for optical observations, a more user-friendly formula giving an answer in arc seconds can be obtained.

$$r = \frac{0.122}{D}\left(arcsec\right)$$

---

[39] For those with a theoretical mind, a list of astrophysics texts is in the appendices.

[40] Note that this disc of light is not the actual surface of a star, but just a disc of light from the star.

[41] A nanometre is one billionth of a metre.

where r is the angular resolution in arc seconds[42] and D is the diameter in metres of the objective lens or mirror.

There's another definition of the highest resolution of a telescope to be found in the literature, and this is the *Dawes limit*. This one isn't derived from any theory, but is an empirical criterion, the result of a series of observations made with telescopes of various apertures. The resolution in arc seconds for the Dawes criterion is given by the formula.

$$r_D = \frac{0.116}{D}(arcsec).$$

where D is the diameter of the objective lens or mirror in metres.

Both these resolution criteria are useful in that they can give a useful measure for the capabilities of your telescope, but be warned, in practice the performance of the telescope may be different from both the Rayleigh and Dawes criteria. The reasons for this are:

(i) The visual acuity of the observer.
(ii) Both criteria only apply strictly for objects that are both of the same brightness. The bigger the difference in brightness, the greater is the discrepancy between what is expected and what is actually seen.[43]
(iii) The criteria have been calculated for a light wavelength of 500 nanometres, and thus it follows that a pair of bluish stars can be resolved at a smaller separation than a pair of reddish stars.
(iv) The type of telescope you use has can also change the resolution. A reflecting telescope has about 5% greater resolution than a refractor of the same aperture, owing to diffraction effects at the support and flat in a Newtonian or Dobsonian telescope.
(v) Atmospheric turbulence, or scintillation, usually always stops you from achieving the expected resolutions.

It may seem to you that no matter what is the calculated or expected resolution of your own telescope, you will very rarely achieve it, or even know what it is actually capable of resolving! Take heart, though—there *is* a way to discover what your telescope is actually capable of, namely by observing a number of double stars whose separations are known.

---

[42] A degree can be subdivided into 60 arcminutes. Each arcminute can be further subdivided into 60 arcseconds. This there are 3600 arcseconds in 1 degree. For reference, the noon has an angular diameter of 30 arcminutes.

[43] You may be forgiven in thinking that brighter stars would be easier to resolve than fainter ones, but this is not borne out from experience. In addition, bright double stars are found to be more difficult to resolve as they tend to "dazzle", and reduce the performance of the eye

By undertaking this series of observations, you will be able to determine the performance of your telescope under various conditions, and thus determine the resolution. Of course, the above list of conditions will have to be taken into account, but at least you will know how your telescope behaves, and thus will know what objects cannot be seen, and, more importantly, those which can.

1.9.3 Limiting Magnitude

Having decided what magnifications to use with your telescope, and the resolution you can expect to get, we now turn our attention to the topic of *limiting magnitude*, m_L, or *light grasp*. This determines what is the faintest object you can detect with a given telescope.

Once again, different books give disparate explanations and formulas, thus confusing the issue, and we should not forget that several factors similar to those I listed in the previous section will determine what you can see. And then there is the issue of whether continuous visibility is needed, or whether a fleeting glance can be considered as detection—an important point for the variable-star observer, who needs to make a definite magnitude determination,[44] as opposed to a glancing determination of magnitude.

The important point here is that the bigger the aperture, the greater the amount of light collected, and thus the fainter the objects detected. For example, a telescope with a 5 cm aperture has half the light grasp of a 7.5 cm one, which in turn has just over half the light grasp of a 10 cm telescope. In fact, the theoretical value for light grasp is given by the simple formula.

$$\frac{D^2}{P^2}$$

where D is the diameter of the objective lens or mirror, and P is the diameter of the eye's pupil.

In order to determine the limiting magnitude, a series of observations were carried out several years ago on the faint stars in the Pleiades star cluster, with telescopes of different apertures, and an empirical formula determined. This is

$$m_L = 7.71 + 5 \log D$$

and represents the expected performance of typical observers under normal conditions.

Sometimes, when conditions permit, it may be possible to improve your visual limiting magnitude by using a higher magnification, because this reduces the total amount of light from the sky background.

---

[44] Magnitudes will be discussed in a later section.

Again, it is worth stressing that factors such as observing skill, atmospheric conditions, the magnification used and even the physiological structure of an observer's eyes (we're all different!) can and does influence the limiting magnitude observed, and any figures quoted are approximate.[45]

The table below gives the Rayleigh resolution, r, Dawes resolution, $r_{D,}$ and limiting magnitude, m_L for those telescope apertures most commonly used by amateurs.

D (cm)	5.0	6.0	7.5	10.0	12.5	15.0	17.5	20.0	22.5	25.0
r (arcsec)	2.77	2.3	1.85	1.38	1.11	0.92	0.79	0.69	0.65	0.55
rD (arcsec)	2.32	1.9	1.54	1.16	0.93	0.77	0.66	0.58	0.52	0.46
m_L	11.2	11.5	12.1	12.7	13.2	13.6	13.9	14.2	14.5	14.7

1.9.4 Field of View

The field of view of is an important topic. Field of view defines how much of the sky you see through your equipment, depending on the type of eyepiece you use.

The true field of view of a telescope (TFOV) is given by.

$$TFOV = \frac{FOV\ of\ eyepiece}{Magnification}$$

This equation shows that when the magnification increases, the size of the amount of sky visible decreases. This is why it is so important to centre any object you view in eyepiece before switching to higher magnifications, especially with faint and small objects such as, say, planetary nebulae. If you don't do this and the object is off-centre, switching to a higher magnification will result in your losing the object, and a frustrating time can ensue as you try to find it again. Always centre objects in the eyepiece initially.[46]

Of course, in order to use the above formula, one needs to know, or determine, the field of view of an eyepiece that fortunately is very easy. Locate

---

[45] Many popular astronomy books will tell you that the faintest, or limiting magnitude, for the naked eye is around the 6th magnitude. This may well be true for those of us who live in an urban location. But the truth of the matter is that from exceptionally dark sites with a complete absence of light pollution, magnitudes as faint as 8 can be seen. This will come as a surprise to many amateurs. Furthermore, when eyes are fully dark-adapted, the technique of averted vision will allow you to see with the naked eye up to three magnitudes fainter! But before you rush outside to test these claims, remember that to see really faint objects, either with the naked eye, or telescopically, several factors mentioned in the text will need to be taken into consideration, with light pollution as the biggest evil.

[46] With some observations it may be necessary to off-centre the object in order to locate fainter and more elusive structure within an object. This is good and proper observing technique, but centre it first, then move off.

a star that lies on or very close to celestial equator—α Aquari and δ Orionis are good examples—and set up the telescope so that the star will pass through the centre of the field of view. Now, using any fine controls you may have, adjust the telescope in order to position the star at the extreme edge of the field of view. Turn off any motor drives and measure the time *t* it takes for the star to drift across the field of view. This should take several minutes and seconds, depending on the eyepiece used. Then multiply this time by 15, to determine the apparent field diameter of the eyepiece, conveniently also in minutes and seconds—but minutes and seconds of arc, rather than of time.

If you have to use a star which does not lie on or close to the celestial equator, then the formula

$$15t\cos\delta,$$

where δ is the declination of the star, can be used to find the apparent field diameter of the eyepiece in minutes and seconds of arc.

All of the previous sections have dealt with parameters that are defined by the optics of the system you are using. We will now look at a factor that is beyond the control of most of us—the atmosphere (Math Box 1.9).

Math Box 1.9: Telescope Basics

We have a telescope and eyepiece with the following parameters

Aperture 300 mm
Focal Length 2000 mm
Focal Length of eyepiece 20 mm.
Apparent Field of View of eyepiece 50 degrees
Determine the following
Magnification, the Dawes Limit, and True Field of View.

Magnification is given by

$$magnification = \frac{f_o}{f_e}$$

Thus

$$magnification = \frac{2000}{20} = 100\text{X}$$

The Dawes Limit is given by

$$r_D = \frac{0.116}{D}(arcsec)$$

Thus

$$r_D = \frac{0.116}{0.3} = 0.39\,arcsec$$

The True Field of View is given by.

$$TFOV = \frac{FOV\ of\ eyepiece}{Magnification}$$

Thus

$$TFOV = \frac{50}{100} = 0.5\,degree$$

Thought Question 1.4

You put a cover over the end of your telescope that reduces the aperture, and thus the amount of light entering. This is called stopping down. What is likely to happen to the resulting image?

1.10. Atmospheric Effects

No matter what sort of equipment you have, whether it is a pair of binoculars, a small telescope, or a large-aperture “light-bucket”, all equipped with superb optics, and a rock-solid mount, there is one element that can reduce us all to equals, and this is the atmosphere we live in. Not that I’m against it, of course (we have to breathe), but it really is the bane of astronomers the world over! There are three things to discuss: transparency, seeing and light pollution.[47]

1.10.1 Transparency

The term *transparency* is used to define the clarity of the atmosphere, an important factor when taking long-exposure photographs or measurements. It is dependent on the altitude of your observing site and there are several components which contribute to the transparency: clouds, fog, mist, smoke and particles suspended in the air. Such is the effect of these components

---

[47] I have included light pollution under this heading as without an atmosphere, there wouldn’t be any light pollution!

that even under what may be considered perfect conditions, a star always appears nearly three magnitudes dimmer at the horizon than it would at the zenith.

Other factors can also affect transparency. Living near built-up areas and even aurora in the upper atmosphere can dim the night sky.

1.10.2 Seeing

Seeing[48] is something everyone is familiar with if they trouble to look, for the twinkling of the stars is dependent upon the condition of the atmosphere—whether the air is steady or turbulent. Seeing affects the quality of the telescopic image in that it can cause the image in the telescope to dance about, or deform the image, or even do both simultaneously.

The twinkling, or scintillation, comes about when the temperature of the air changes, altering the air's refractive index and hence causing the image to flicker. Surely everyone has noticed the scintillation that Sirius exhibits when it is close to the horizon on cold winter nights? It twinkles in all the colours of the rainbow as different wavelengths are dispersed by the atmospheric turbulence!

Seeing can also be divided into two components, sometimes referred to as "high" and "low" seeing. High seeing is due to air currents found at altitudes of a 1000 metres and more, and causes the movement of an image in the field of view as mentioned above. Low seeing, as you might expect, depends on local conditions near ground level and also within a dome or telescope; for example, warm air trapped inside a dome or telescope tube causes the air to become unsteady.[49]

Telescopes with different apertures are affected in different ways. One with a large aperture "sees" a larger cross-section of turbulent air and so may produce a more deformed image than a smaller-aperture one.

Several observing scales have been introduced, and the one most widely used is the Antoniadi scale. It was originally intended for lunar and planetary work but is just as applicable for other objects. You can write it down in "shorthand" like this:

---

[48] The topic of seeing is in fact highly complex, but for our purposes, we will ignore all of the mathematics.

[49] A surprising result is that when the transparency is poor, the seeing can be good. For example, if the there is a slight haze in the air, as on Autumn evenings, the atmosphere will be still, and this is often a good time for planetary observation. Similarly, when the transparency is good, and the stars appear as bright sharp points, the seeing can be at its worst. These nights, however, are perfect for observations of nebulae and galaxies.

I. Perfect seeing, no movement whatsoever.
II. Moments of calm lasting several seconds, with some slight undulation.
III. Moderate seeing, accompanied with large air tremors.
IV. Poor seeing, accompanied by constant air tremors.
V. Bad seeing, preventing any worthwhile observing to be made.

1.10.3 Light Pollution

It is a sad reflection on our times that nearly all amateur astronomers are familiar with, and complain of, light pollution. It has grown alarmingly over the past few decades, and although steps are being made to reduce it, it marches ever onward.

The cause of the problem is predominantly street lighting, shining up into the sky and being scattered and reflected throughout the atmosphere. This glow of diffuse light is instantly recognisable as the orange radiance seen over most of the horizon and extending quite high into the sky. From urban sites, it severely limits what can be seen, particularly of the fainter nebulae and galaxies.

The good news is that most of the brighter objects I talk about in this book can be seen from an urban or semi-urban location.

There are two practical solutions to the problem of light pollution,[50]

(i) use filters to block it out,[51]
and/or
(ii) make your observations from a dark location.

The first may be of little use to the binocular observer, but should be a weapon in the arsenal of every telescope user, while the second may be impractical for casual observers or those of us with big telescopes or permanently located observatories

However, observing from a dark sight is a truly awe-inspiring experience. I have observed from several locations throughout the world—usually, but not exclusively, at major professional observatories—and the experience of being unable to discern the constellations because of the plethora of stars visible was breath-taking—as was being able to see most of the Messier objects in Sagittarius with the naked eye! If you ever have the opportunity to observe from a really dark location, then take it![52] It also makes sense to

---

[50] One could also write to your local council, and explain the benefits of a darker sky.

[51] A discussion on light pollution filters can be found in the appendices.

observe after it has rained, as rain removes some of the dust and larger pollutant particles in the air, thus reducing the scattering of streetlights.

The three atmospheric variables mentioned above—transparency, seeing, and light pollution –all have an influence on what can be observed, and thus you should always record them. Any classification you use will always be of a subjective nature, and writing "poor" as a description of the conditions will not usually provide much information, at least for other people.

Finally in our discussion of telescopes and observing, let's now turn our attention to two matters that will allow you, when you have mastered them, to observe faint objects—*dark adaption* and *averted vision*.

1.10.4 Dark Adaption and Averted Vision

The eye is a very complicated optical device. It uses a simple lens to focus light onto the retina, and changes focus by altering the geometry of the lens itself. The retina—the light-sensitive inside back surface of the eye—is composed of two sorts of photosensitive cell, *rods* and *cones*. These cells are packed closely together, stacked rather like the pile in a carpet. The cones are responsible for the perception of colour and for our excellent daytime colour vision; they also enable us to see fine detail. The rods, on the other hand, are sensitive to very low levels of illumination (and also to movement), but produce a low-resolution image and do not discriminate different colours. These different photosensitive cells are mixed together in the retina, but the concentration of each type is not even—there are many more rods toward the edge of the retina, and more cones near the eye's optical axis. Right in the centre of the retina is a small area called the *fovea centralis*, where there are no rods at all and the cones are packed extremely densely. This area is about 600 μm[53] in diameter and is the part of the retina that provides the highest resolution of detail; however, the absence of rods makes it relatively insensitive to light.

The eye has three mechanisms for regulating the amount of light reaching the retina. Short-term and fairly small variations are dealt with by the iris, which opens and closes to adjust the size of the eye's aperture. There is

---

[52] Note that the term "dark sky" refers to a sky that is clear, free of light pollution, and transparent—not one that is very black. In fact, contrary to popular belief, the more light that is seen from stellar objects, the brighter the sky will appear to be. From exceptionally "dark" sites, the light from the stars, Milky Way, zodiacal light, galaxies, and so on., will all combine to brighten the night sky. With such a "dark sky" objects with very faint magnitudes can be observed.

[53] A μm is one millionth of a metre, often called a micron.

also a safety system (rather like that on the Hubble Space Telescope) that rapidly closes our eyelids if the light is too bright. The third mechanism is chemical in nature, and varies the sensitivity of the retina itself, like using fast or slow film in a camera. It is this chemical mechanism that causes the most dramatic change in the eye's sensitivity to light.

When you leave a brightly lit room and go outside and observe you can only see a few stars, but after a period of time, which can be as short as 10 minutes, it becomes apparent that more stars are becoming visible. In fact, after a period of about 15 minutes, the eye is six times more sensitive to low light levels than it was immediately after leaving a lit room. This is the process of *dark adaption*. If you spend even longer in the dark—at least 30 minutes—then the rod cells, located, you will remember, mostly nearer the edge of the retina, can become nearly *one thousand* times more sensitive.

That's why it really is very important to allow time for your eye to become adapted to the darkness before you start to observe. Try to make sure that no bright light can interfere with your observing (as this will destroy your dark adaption and you will have to wait another 30 minutes before you begin observing again). Such is the sensitivity of the eye that even bright-red light can affect your adaption, so a dull-red light should be used. Indeed, some extremely keen observers place thick black cloth over their heads (and yes, it may look silly, but it works really well!) to totally exclude even the minutest trace of light, before beginning observations. It's worth thinking about keeping your eyes dark-adapted, and allowing time: it really does work, as faint objects will be much easier to locate.[54]

A possible problem can arise for double-star observers because of the eye's response to differing levels of illumination. The cones, which are responsible for colour vision, peak at about 550 nanometres, that is, in the

---

[54] It is often useful to be able to determine the night sky's observing conditions (light pollution, haze, cloud cover, transparency) before starting an observing session, so as to determine what type of objects will be visible and even allow you to decide whether observing is viable at all. A good way to do this is to use a familiar constellation, which should be observable every night of the year, and estimate what stars in the constellation are visible. If only the brighter stars are visible, then this would limit you to only bright stellar objects, while if the fainter stars in the constellation can be seen, then conditions may be ideal to seek out the more elusive, and faint objects. A favorite constellation used by many amateurs for just such a technique is Ursa Minor, the Little Bear. If, once outside, you can see υUMi (mag 5.2) from an urban site, then the night is ideal for deep-sky observing. However, if υUMi is not visible, then the sky conditions are not favorable for any serious deep-sky observing, but casual constellation observing may be possible. If the stars δ, ε and ζUMi, located in the "handle" of the Little Bear, are not visible (magnitudes 4.3, 4.2 and 4.3 respectively) then do not bother observing at all, but go back indoors and read this book.

yellow–green region of the spectrum, while the rods peak at about 510 nanometres, the green region of the spectrum. This shift in sensitivity is called the *Purkinje effect* and may result in an observer underestimating the magnitude of a bright and hot star when comparing it with a cool, fainter one. This is also the reason why moonlight appears bluer than sunlight.

When you look directly at an object, you are using the fovea centralis, because it is in the optical centre of the retina, and probably some of the surrounding area of the retina in which there are relatively few rod cells. The human eye has evolved to provide us with high-resolution vision in daylight, along with low-resolution but sensitive night vision, particularly at the periphery—for daytime hunting and night-time avoidance of predators!

Thought Question 1.5

You purchase an enormous reflecting telescope, mirror diameter 300 mm. You also live in a suburban location with a considerable amount of light pollution. Do you think this is a good investment?

Problems

1. Calculate the distance to a star that has a parallax angle of 0.1 arcsec.
2. Convert your answer to problem 1 to light years.
3. Calculate the temperature of a star that emits its energy at a peak wavelength of 91.2 nm.
4. What is the spectral classification of this star?
5. A star has a temperature of about 11,600 K and has a luminosity of about 100 $L_{\odot}$. Determine its radius ratio.
6. A telescope has the following parameters: Aperture 500 mm, Focal Length 1000 mm, Apparent Field of View of eyepiece 50 degrees and an eyepiece of focal length 5 mm. Calculate (a) its Magnification, (b) its Dawes Limit, and (c) its True Field of View.
7. If a star is visually, 10 times brighter than another star, what is the difference in apparent magnitude.

Chapter 2

The Solar System

It may seem strange to have a chapter on the early history of astronomy and the Solar System in a book dedicated to astrophysics, a subject most people would believe applies more to stars and galaxies. However, there is a surprising amount of astrophysics that can be discussed here, as we shall see.

Although a whole book could be written on the subject of Solar System astrophysics, we will limit the discussion mostly to the early history, and to the development of astronomy that led to an understanding of the behavior of planetary orbits and a couple of smaller, but equally important, topics. We will not be covering such areas as planetary geology, planetary atmospheres and the like. There are several books dedicated specifically to those topics listed in the appendices. Naturally, some Bronze Age and Iron Age history will need to be covered, so as to set the scene for the revolution in astronomy that occurred in the Middle Ages.

So, without further ado, let us begin our journey…

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_2

2.1. Early History of Astronomy

2.1.1 The Geocentric Universe

For most of history, the prevalent view was that Earth was the center of the universe,[1] and that all celestial objects revolved around it. Think about this for a minute; all astronomy was naked-eye astronomy and so was limited to the stars, planets, comets, meteors, the Moon and the Sun. Moreover, the real nature of the aforementioned objects was unknown, and either given a supernatural, or religious, identity.

This view is known as the geocentric, or Earth-centered, point of view, and can be easily understood, as it really does seem as if the everything revolves around us; the Sun and Moon and planets rise in the east and set in the west, along with the stars. There is no physical sense of movement of Earth whatsoever, and it took quite a while before it was shown that Earth was revolving on its own accord.[2]

The most obvious historical objects of any astronomical significance must be the magnificent stone henges built throughout Europe and Asia, such as Stonehenge and Brodgar, where stone pillars are aligned, so we believe, to indicate the rising and setting of the summer and winter solstice Sun, etc. But the earliest known written astronomical records are clay tablets from Babylonian and Sumerian[3] civilizations, and later Egyptian[4] hieroglyphs. The former two were especially interested in determining the appearance of the new Moon for their calendar, while the latter focused their attention on Sirius, whose appearance seemed to be connected with the flooding of the Nile, a very important event to them.

---

[1] Don't confuse this "universe" with the one we are familiar with today. The universe then was only all that could be seen with the naked eye and was often referred to as "the Heavens."

[2] This is a great question to ask at public star parties—"How do you know that Earth is revolving, and that rather it is the sky (and all it contains) that is revolving instead. Do we feel Earth move? The solution to this problem is to mention Jean Foucault, who, in 1851, with an actual demonstration using a pendulum, showed the effect of the rotation of Earth.

[3] The Sumerians based their calculation on a base 60 format, not the base 10 we use today. We still use a remnant of this system, our angular measurement scheme, 60°, 60 arc minutes, and 60 arc seconds, and the time system.

[4] The Egyptians, like the Babylonians, kept to a lunar cycle but eventually changed to a 12-month, 30-day system. However, in order to make the new year coincide with the appearance of Sirius, they added extra days to the calendar, giving us the 365-day year.

It is important to note here that these civilizations, and even those that followed, tended to immerse their astronomy within a mystical and religious framework, and it was the action of gods that dictated the events they saw in the sky.

It wasn't until the appearance of the Greeks that things really started to take a more scientific aspect. They, like the Egyptians and Sumerians before them, extended the idea of a domed heaven to that of a giant sphere—the *celestial sphere*—that carried the stars on its inner surface and rotated around a vertical, that is to say, north-south axis.

We still use this idea today, especially when referring to coordinate systems used in astronomy, but we know now of course that the apparent rotation of the celestial sphere is in reality due to the actual rotation of Earth.

The Greeks started to devise some explanations for familiar phenomena that occurred in the sky. For instance, the phases of the Moon were quite familiar to the ancient civilizations, as well as the annual motion of the Sun, but it was the Greeks that realized the position of the Sun and Moon had to coincide for eclipses, whether they be solar or lunar, to occur.

However, there were problems on the horizon. The five planets known at that time—Mercury, Venus, Mars, Jupiter and Saturn—posed problems. They moved independently to the stars, Sun and Moon, along a path called the zodiac, in periods ranging from a quarter of a year for Mercury to 29 years for Saturn. This motion however was not periodic; some objects seemed to slow down, stop, and then reverse their path along the zodiac for a short time, then reverse again and continue in an eastward direction. This *retrograde motion* was even known to the Babylonians.

Furthermore, Mercury and Venus were always found in the vicinity of the Sun, sometimes lagging behind it, and at other times overtaking it. At one time it was even thought that Venus was in fact two stars—the morning star when it rose in the sky before and west of the Sun, and the evening star when it set after and east of the Sun.

2.1.2 The Scientific Method

To present a full and detailed account of the scientific models that were proposed in early history is far beyond the scope of this book, so we shall confine ourselves to just the salient points, events that eventually led to a model of the heavens that was adopted for nearly 1500 years. Before we do that, however, it is important that we discuss what is known as the *scientific method*.

Basically, the scientific method is the way that science is carried out today, irrespective of what field of science we are talking about, and for any scientific model or theory or idea to be taken seriously, the scientific method must be shown to have been applied. A dictionary describes the method thus: "…the scientific method is a method or procedure that has characterized natural science since the seventeenth century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses."

What it comes down to is this—one tries to describe what is observed, with an idea, or a hypothesis, and then one tests this idea, or hypothesis, along with a prediction of what the result should be. Depending on the results of the test, one either adopts the hypothesis or alters it slightly and does the test, or experiment, again, or completely discards the hypothesis as being utterly wrong.

Once the hypothesis has been tested repeatedly, often by different people in different laboratories (but using the same or similar equipment), and is shown to be successful after many, many repeated experiments, only then can the hypothesis be termed a theory.

There are difficulties, however, in following such a formulaic statement of the method, however. Though the scientific method is sometimes presented as a fixed sequence of steps, it is often more helpful to consider the steps more as general principles. In fact, not all steps take place in every scientific inquiry (or to the same degree), and they are not always in the same order. As the Victorian scientist William Whewell said, "invention, sagacity, genius" are required at every step:

What follows are the basics of the scientific method:

Question—The question can refer to the explanation of a specific observation, as in "Why do things fall down when dropped?" but can also be open-ended. This step also means looking up and evaluating evidence from previous experiments, personal scientific observations or assertions and/or related work of other scientists.

Hypothesis—A hypothesis is a conjecture, based on knowledge obtained while formulating the question that may explain the observed behavior of a part of our universe. The hypothesis might be very specific, e.g., Einstein's equivalence principle.

Prediction—This step involves determining the logical consequences of the hypothesis. One or more predictions can then be selected for further testing. The more unlikely that a prediction would be correct simply by coincidence, the more convincing it would be if the prediction were fulfilled.

Testing—This is the part where, say, an experiment is performed to see whether the real world behaves as predicted by the hypothesis. The purpose of the experiment is to determine whether observations of the real world agree with or conflict with the predictions derived from the hypothesis. If they agree, confidence in the hypothesis increases; otherwise, it decreases.

Analysis—This involves determining what the results of the experiment show and deciding on the next actions to take. If the results do not support the hypothesis, a new hypothesis is required; if the experiment supports the hypothesis but the results are not strong enough for high confidence, other predictions from the hypothesis must be tested.

There are a few other components to the scientific method that can be done even when all the iterations of the steps mentioned have been completed:

Replication—If an experiment cannot be repeated to produce the same results, no matter who does the experiment, this implies that the original results were in error. Thus, it will be necessary for the experiment to be performed several times.

External review—The process of peer review involves experts, often in the same field of science, evaluating the experiment, who give their opinions anonymously to allow them to give unbiased criticism. If the work passes peer review, which could require new experiments requested by the reviewers, it would often be published in a peer-reviewed scientific journal.

Data recording and sharing—Scientists must record all data very accurately in order to reduce their own bias and aid in replication by others. They must be willing to supply this data to other scientists who wish to replicate any results.

The most successful explanations of the natural world, ones that seek to explain and make accurate predictions in a wide range of circumstances, are called scientific theories.

2.1.3 Ancient Greek Science

The reason the scientific method was discussed at this point is because the Greeks were the first people that tried to explain what they saw, using mathematics models and not relying on mystical reasons and religious explanations, and thus were using, sort of, the scientific method.

Admittedly, they got it wrong quite a few times, but nevertheless they started the process. To describe in detail the main ideas and introduce all the people that developed them would literally fill up several books, so we shall just give the salient points and end up with the model that was the basis of astronomy for over 1500 years.

Many Greek and Egyptian mathematicians and philosophers developed models that were added to or refined over a period of 500 years, among them Leucippus, Democritus, Pythagoras, Heraclides, Philolaus, Plato, Eudoxus, Aristotle, Hipparchus, and, last but not least, the man who gathered together all these ideas and formed a working model, Claudius Ptolemaeus.[5]

Basically, the explanation, or model, goes something like this.

To begin with, Earth is at the center of the universe and is surrounded by 56 concentric, transparent crystal spheres, rather like the Russian dolls one can buy, where you take the outer one off and inside is a smaller doll, and you take that one off and inside there is a smaller doll…you get the idea.

The outermost sphere is the "celestial sphere," upon which the stars reside. The motion of this outermost sphere was transmitted to an outer sphere of Jupiter; and between the inner sphere of Jupiter and the outer sphere of Saturn lie three additional spheres, and so on and so on with Mars, Venus and Mercury, as well as spheres for the Moon and the Sun. and all these spheres were connected by various linkages. As you can see, it is complicated!

In addition, there were a few other ideas that were adhered to and proved quite difficult to get rid of. One of these was of such appeal that it took a very long time for it to be discarded and replaced. It was this:

> The universe is perfect and unchanging, and thus its constituents are perfect and must move along perfect orbits. Since the circle is the perfect curve and the sphere a perfect solid, it follows naturally that the heavenly bodies, including Earth, are spheres.

Although the model seemed, initially, to correctly describe what was observed, it did have a few problems, as was mentioned earlier, in that Mercury and Venus are always close to the Sun, the retrograde motion of Mars, as well as the changing brightness's of the planets, and so on. To explain these anomalies, corrections and additions were made to the model such that it became very complicated.[6]

Another idea that must be mentioned here is that, surprisingly, the Greeks were aware of the phenomena of parallax—the apparent displacement of an

---

[5] There were, of course, many more people who contributed ideas, but those listed were, more or less, the main players.

[6] Don't think that everything they did was wrong however, Greek mathematicians managed to work out the distance and size of the Moon and Earth, as well as the precession of the equinoxes, to name but a few.

object owing to the motion of the observer, in this case the motion of Earth.[7] They believed that the apparent shift of the stars due to the motion of Earth around the Sun (!) would be visible and measurable. They didn't find any, not surprisingly, as the distances to the stars are so immense that the crude instruments of the Greeks could not detect the parallax shift. However, Aristarchus[8] guessed the reason and pointed out that the orbit of Earth was very small compared to the size of the celestial sphere, so that it was like the center of a sphere of infinite radius and thus immeasurable.

However, not even Archimedes would accept this explanation, as the Greeks could not grasp such concepts as infinity. Thus the idea that Earth revolved around the Sun was discarded, and rather the reverse scenario was the true one—the Sun revolved around a stationary Earth, a concept that was adhered to for a very long time.

2.1.4 *The Ptolemaic System*

The man who collected all these ideas and placed them within a coherent model was Claudius Ptolemaeus, also known as Ptolemy of Alexandra. He is the author of the *Almagest*,[9] a book that contained a full description of all astronomical knowledge of his time as well as his own contributions. It became the astronomical Bible and formed the basis of western astronomy throughout the Middle Ages.

Of particular interest are the introductory chapters that deal with what can be called the postulates of Ptolemaeic astronomy. In them, he provides convincing arguments as to why Earth is spherical, dismisses the idea of a spinning Earth and instead assumes Earth to be immovable at the center of the universe, "a point compared with the surrounding star-sphere."

In the Ptolemaic system, each planet moves uniformly inside a small circle, called an *epicycle*, the center of which moves along the circumference of a larger circle, the *deferent*, with the center of the deferent situated a short distance from the center of Earth, at a different point for each planet. The following diagram, Fig. 2.1, will illustrate this.

---

[7] This topic was covered in detail in Chap. 1

[8] Aristarchus was a lone voice at this time, as he did propose that the Earth, Moon and five planets revolved around a motionless sun. The work in which he put forward this remarkable idea is alas, lost to us, but is reported by such authorities as Archimedes and Plutarch. There can be no doubt that Aristarchus was indeed the first to propose a sun-centered universe.

[9] This is the name given it by its Arabic translators.

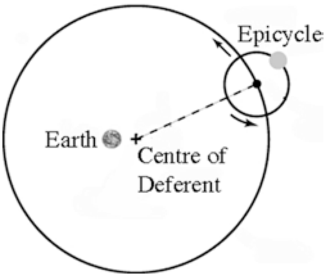


Fig. 2.1. Epicycles and deferents

The whole system was then assumed to rotate slowly around its common axis, thus accounting for the precession of the equinoxes. In this way, it was possible to account for the irregular motions of the planets quite accurately but at the same time preserve the basic idea of motion along circles with constant radii and constant speed.

In this system it was easy to see how the retrograde motion of Mars arose. As the planet moved along its epicycle, it would, at one time be moving, say, in a left direction, Position 1, but at a later time it would be moving towards the right, Position 2, seemingly moving backwards, and at Position 3, it would continue its leftward motion (see Fig. 2.2).

The Moon had a similar setup, but its epicycle revolved in an opposite direction to those epicycles of Mars, Jupiter and Saturn, to account for its differing speed at full and new phase, compared to the other phases.

Mercury and Venus, too, were treated differently from the other planets to account for the fact that they were always seen in close proximity to the Sun. For this reason, the centers of their epicycles had to remain on the same line as that of the Sun from Earth, and their periods on the deferents had to be the same as that of the Sun, namely 1 year (see Fig. 2.3).

In this way, Ptolemy built up his system, adding further epicycles and deferents whenever the data warranted it, until at the end, he had 40 epicycles, including the celestial sphere. This, now very complicated, model remained unchallenged for centuries, with the idea of Earth at the center of the universe.

What followed can be considered the Dark Ages in the development of astronomy in Europe, and even though some enlightened people attempted to continue the work of the Greek mathematicians and philosophers, such as Thomas Aquinas (1225?–1274) and Roger Bacon (1214?–1294), much was lost. Indeed Roger Bacon, who advocated the study of science, was accused of witchcraft by the Church, thrown into prison for 10 years and his work forgotten. It was 400 years before his work was published.

Although the study of astronomy and science languished in Europe, it flourished in the Arabic world. Arab scientists kept the knowledge gathered

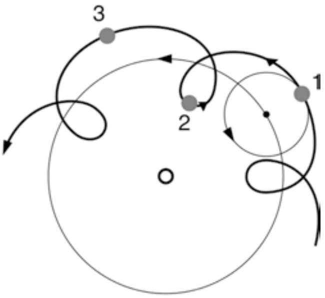


Fig. 2.2. Retrograde motion illustrated

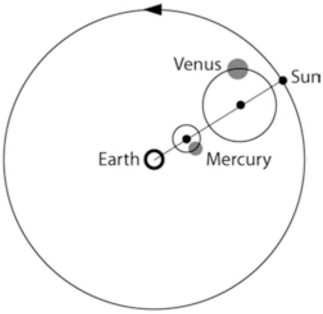


Fig. 2.3. The Earth-Sun line

by the Greeks alive, developed mathematics, especially algebra, translated much of the Greek work into Arabic (which fortunately was later translated into European languages), and it was probably about this time that the first custom-built astronomical observatories were erected. In fact, it can be said, with some justification, that without the flourishing world of Arabic science, and astronomy in particular, much would have been lost, and the further development of astronomy would have taken place at a much later date.[10]

We now move on to when astronomy was resurrected in Europe, and the true development of science in the western world begins.

Thought Question 2.1

Can you think of two reasons why the Geocentric model was believed to be the correct one?

---

[10] Just think of some star names—Rigel, Vega, Betelgeuse, to name but a few—that have Arabic roots.

2.1.5 The Copernican Revolution

We begin our story in the sixteenth century, in Poland, where a man called Niklas Koppernigk, but more famously known as Copernicus, put forward several new ideas, that, although not accepted initially, really did cause a revolution in deposing the then accepted view of the Earth-centered universe.

Copernicus was a canon in the Church, working in the small, sleepy town of Fraunberg, having acquired a well-rounded education, and was proficient in medicine, jurisprudence and astronomy, although there was no indication that he was about to set the scientific world on fire.

He sent a short manuscript, in Latin—*Commentariolus*—to some friends, summing up his ideas in seven propositions, or assumptions, and they are listed here:

- There is no one center of all the celestial circles or sphere.
- The center of Earth is not the center of the universe, but only of gravity and of the lunar sphere.
- All the spheres revolve around the Sun as their midpoint, and therefore the Sun is at the center of the universe.
- The ratio of Earth's distance from the Sun to the height of the firmament is so much smaller than the ratio of Earth's radius to its distance from the Sun that the distance from Earth to the Sun, in comparison with the height of the firmament, is imperceptible.
- Whatever motion appears in the firmament arises not from any motion of the firmament but from Earth's motion.
- What appears to us as motions of the Sun arise not from its motion but from the motion of Earth and our sphere, with which we revolve around the Sun, like any other planet. Earth has, then, more than one motion.
- The apparent retrograde and direct motion of the planets arises not from their motion but from Earth's. The motion of Earth alone, therefore, suffices to explain so many apparent inequalities in the heavens.

As you can see, this is all quite revolutionary. Note that the two most fundamental innovations in the list are,[11] of course, that the Sun and not Earth is at the center of the universe—the *heliocentric system*, and that it is the motion of Earth that accounts for the apparent daily rotation of the celestial sphere, as well as other "apparent inequalities in the heavens."

However, don't make the mistake in thinking that he got it all correct—he didn't. For instance, he still believed that everything moved in circles, that

---

[11] But recall that Aristarchus had proposed these ideas over 1000 years earlier!

there was a sphere of fixed stars beyond the planets, and in fact, to account for the varying velocities of the planets in their orbits, he had no choice but to revert to the epicycles of the Ptolemaic system, since circular motion (or a combination of circular motions) was the only possible one. In the end he had a system of 36 (some historical researchers suggest 38) circles that would completely explain the entire structure of the universe.[12]

Initially, there wasn't much reaction, but over time word slowly spread. He continued, over many years, to refine his measurements, and eventually published a six-volume work entitled *De Revolutionibus*, containing all of his ideas along with expositions of mathematical astronomy, spherical trigonometry, star catalogs, descriptions of planetary orbits, etc.

However, this wonderful endeavor hasn't got a happy ending. Although Copernicus was recognized as a great astronomer, his ideas were ignored, even ridiculed.[13] All the previous classical arguments against an Earth that moved were resurrected and passages from the Bible were cited to refute him. It is said that he died, on May 24, 1543, only a few hours after he had received one of the first copies of his book. It took some time for Copernicus and his ideas to gain acceptance. The real significance of the heliocentric, or Sun-centered, system lies in the immensity of its conception, rather than in the discovery itself. With his concept of a moving Earth, Copernicus laid the cornerstone for modern astronomy.

Thought Question 2.2

Can you think of one big weakness in the Copernican model, that gives rise to its incorrect predictions of planetary positions?

2.1.6 Tycho—The Great Observer

We are now going to briefly discuss a man whose observations laid the groundwork for the first truly mathematical theories of planetary astron-

---

[12] At about the same time, Thomas Digges (1546–1595), an English member of Parliament, mathematician and astronomer, expounded the Copernican system in English, but more importantly, discarded the notion of a fixed shell of immoveable stars and instead proposed an infinite number of stars at varying distances. He was also first to postulate the "dark night sky paradox," later referred to as "Olber's paradox." We shall discuss this paradox in the final chapter.

[13] Only one man believed in the Copernican system, and that was Giordano Bruno. He taught it, defended it with courage and died for it. He was called before the Roman Inquisition, tortured, and then burned alive at the stake.

omy. That man was Tycho Brahe (1546–1601). Tycho had a very eventful life, full of intrigue and a fair amount of danger.

It was well known that Tycho was a great womanizer, to such an extent that in a duel over a woman part of his nose was cut off, and thereafter he wore a false one of silver and gold. What is less known is that he was also an astrologer, an alchemist and had his own theory of planetary motion that disagreed with the Copernican system. In it he stated that the five known planets revolved around the Sun, which in turn revolved around Earth, with the whole celestial sphere turning around Earth once a day.

What is important, however, is that although not a mathematician, he was ingenious at designing and constructing astronomical instruments at his extensive private observatory, along with his outstanding ability as an observer. Over his lifetime, he made innumerable accurate measurements of the positions, in the sky, of the planets and stars. These measurements, and the following analysis of them, are probably, in this author's opinion, the greatest pieces of pre-telescope astronomical work done.

2.1.7 Kepler—The Great Theoretician

Kepler[14] (1571–1630) worked as an assistant to Tycho, and over several years used his observations to develop ideas of his own that he was formulating.

However, the end result was worth the effort, as he developed three laws of planetary motion[15] that are still used today, whether they be for planets, asteroids, comets, and even, in a modified way, for the orbits of stars around the Milky Way.[16]

He realized that he could get an accurate model of the motion of the planets around the Sun[17] if one dismissed two previously held concepts.[18] These were:

- circular motion
- uniform or constant motion.

---

[14] Kepler was also an astrologer, and his mother was tried for being a witch. He obviously had a lively childhood.

[15] Remember that this work was done pre-calculus and was developed using trigonometry and algebra. Amazing!

[16] Kepler did a lot more than just developing his three laws, most of which were incorrect. It is his laws of planetary motion for which he is justly famous.

[17] Except for Mercury, which presented problems until Einstein explained what was going on.

[18] He also dismissed the idea of the layers upon layers of spheres model. He believed forces made the planets move.

By discarding these concepts Kepler was able to successfully describe the behavior of the planets around the Sun.

When one looks at the laws, a couple of them may seem rather odd, and not at all easily understood. However, after giving the formal statement of the laws, we will present easily understood explanations.

The laws are:

1. All planets move in elliptical paths, with the Sun at one focus (the law of ellipses) (Fig. 2.4).
2. An imaginary line that is drawn from the center of the Sun to the center of the planet will sweep out equal areas in equal intervals of time (the law of equal areas).
3. The squares of the periods of revolution of the planets around the Sun are proportional to the cubes of their mean distances from the Sun (the law of harmonies) (Fig. 2.5).

$$P^2 = a^3$$

Here are explanations of these rather formally written laws.

Kepler's first law—sometimes referred to as the law of ellipses—tells us that planets orbit the Sun in a path described as an ellipse. A more mathematical description of an ellipse would be "an ellipse is a special arc in which the sum of the distances from every point on the curve to two other points is a constant, the two other points being known as the *foci* of the ellipse."[19] The closer together these foci are, the more closely the ellipse

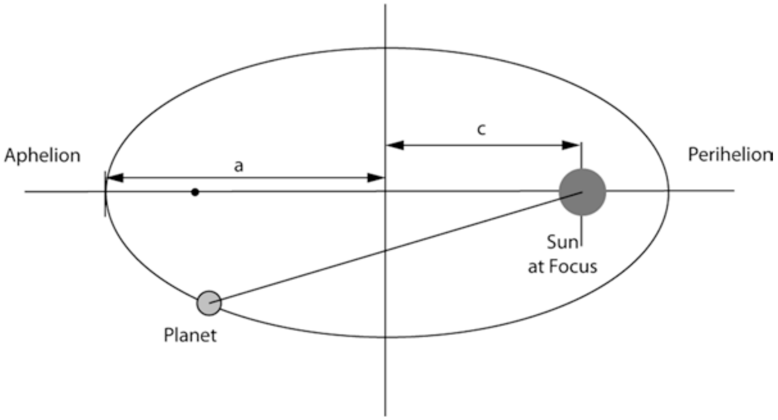


Fig. 2.4. Diagram to illustrate the parameters of Kepler's First Law

---

[19] The Sun is at one focus, there is nothing at the other, it is just a mathematical entity.

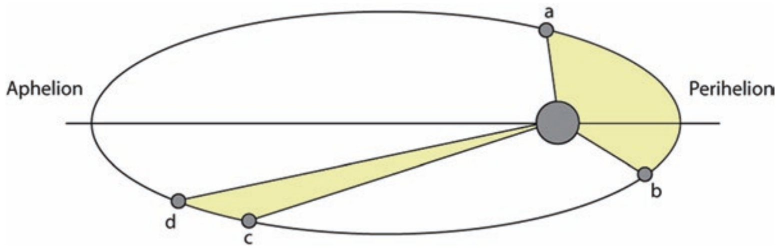


Fig. 2.5. Diagram to illustrate the parameters of Kepler's Second Law

resembles a circle. In fact, a circle is the special case of an ellipse in which the two foci are at the same location. Thus, Kepler's first law is very simple: All planets orbit the Sun in an elliptical orbit with the Sun being located at one of the foci of that ellipse.

Notice in the diagram illustrating the first law that there are two additional quantities, a and c. The former is the distance from the center of the ellipse to the end of its longest axis, and is called the semi-major axis, a. The latter quantity, c, is the distance from the center of the ellipse to a focus (either one). Using these quantities one can get a measure of the *ellipticity*, e, of the ellipse, which quantifies how elliptical the ellipse is.[20] In other words, is the orbit almost circular, slightly squashed, or like a thick cigar? An example is given in Math Box 2.1.

Math Box 2.1: Eccentricity of an Ellipse

The eccentricity of an ellipse is given by

$$e = \frac{c}{a}$$

An asteroid is discovered with the following parameters:

$$a = 3.00 \times 10^8 \text{ km}$$

$$c = 1.05 \times 10^8 \text{ km}$$

Therefore, the eccentricity is given by

$$e = \frac{1.05 \times 10^8}{3.0 \times 10^8}$$

$$e = 0.35$$

---

[20] I apologize for the oddness of this sentence, but there isn't really any other way to write it.

The eccentricity of Earth's orbit is currently about 0.0167; Earth's orbit is nearly circular. Over millennia, the eccentricity of Earth's orbit has varied from nearly 0.0034 to almost 0.058 as a result of gravitational attractions typical of the planets. Mercury has the greatest orbital eccentricity of any planet in the Solar System, with an e of 0.2056. Before its demotion, Pluto was considered to be the planet with the most eccentric orbit, $e = 0.248$, and the Moon's value is 0.0549. Many of the asteroids have orbital eccentricities between 0 and 0.35, with an average of 0.17. These comparatively high eccentricities are believed to be due to the influence of Jupiter and to past collisions.

The second law, although appearing a tad strange, is actually quite easy to understand. Kepler's second law—sometimes referred to as the law of equal areas—describes the speed at which a planet will move while orbiting the Sun. Again, refer to the diagram under the second law. Look at the points a, b, c and d, and imagine that it takes, say, 1 month to go from position a to position b, and it takes the same amount of time to travel in its orbit from position c to position d. It doesn't matter what time interval one takes as long as they are both the same. The law tells us the colored regions that the line sweeps out will be of equal area. In the example given, the area enclosed, from a to b, is the same size as the area from c to d.

So what, I hear you say? Well, this time look at the length of arc from a to b, and then c to d. Notice that the length of the arc from a to b is much larger than the arc from c to d. But, and this is an important but, the time taken for the planet to go from a to b is the same as for the planet to move from c to d. In order for it to do that, it must be moving *faster* at positions a to b, and *slower* at positions c to d. Thus the speed at which any planet moves through space is constantly changing. A planet moves fastest when it is closest to the Sun—*perihelion*—and slowest when it is furthest from the Sun—*aphelion*.

Finally, Kepler's third law—sometimes referred to as the law of harmonies—compares the orbital period and radius of the orbit of a planet to those of other planets. Unlike the first two laws that describe the orbital motion of a single planet, the third law makes a comparison between the motion characteristics of different planets. The comparison being made is that the ratio of the squares of the periods to the cubes of their average distances from the Sun is the same for every one of the planets.

As an example, consider Mars and Earth:

Planet	Period (s)	Average distance (m)	T^2/P^3 (s^2/m^3)
Earth	3.156×10^7	1.4957×10^{11}	2.977×10^{-19}
Mars	5.93×10^7	2.278×10^{11}	2.975×10^{-19}

You can see immediately that the T^2/P^3 ratio is the same for Earth as it is for Mars. In fact, every planet has nearly the same T^2/P^3 ratio.

Planet	Period (year)	Average distance (au)	T^2/P^3 (year^2/Au^3)
Mercury	0.241	0.39	0.98
Venus	.615	0.72	1.01
Earth	1.00	1.00	1.00
Mars	1.88	1.52	1.01
Jupiter	11.8	5.20	0.99
Saturn	29.5	9.54	1.00
Uranus	84.0	19.18	1.00
Neptune	165	30.06	1.00

It is very important to note however that in this analysis we always use units that relate to Earth, i.e., the average distance value is given in astronomical units where 1 AU is equal to the distance from Earth to the Sun—1.4957×10^{11} m—and the orbital period is given in units of Earth-years where 1 Earth year is the time required for Earth to orbit the Sun—3.156×10^7 s. See Math Box 2.2 for more examples.

Math Box 2.2: Kepler's Third Law

Kepler's third law states the period squared is related to the average distance cubed. Providing one uses units of years and astronomical units, the law can be stated thus:

$$P^2 = a^3$$

where P is the period and a is the average distance.

An asteroid has a period of 8 years. Calculate its average distance from the Sun.

So:

$$P^2 = 8^2 = 64$$
$$a^3 = 64$$
$$a = \sqrt[3]{64}$$
$$a = 4$$

Thus the average distance of the asteroid from the Sun is 4 AU.

A comet has a mean distance of 200 AU from the Sun; using Kepler's third law, determine its period.

$$P^2 = a^3$$

$$
\begin{aligned}
a^3 &= (200)^3 = 4000 \\
P^2 &= 4000 \\
P &= \sqrt{4000} \\
P &= 63.25
\end{aligned}
$$

Thus, the period of the comet is 63.25 years.

2.1.8 Galileo—The Great Experimenter

Our penultimate character, in this story of discovery, is someone known to us all. Most everyone has heard of Galileo Galilei (1564–1642) and the stories of him dropping things from the Leaning Tower of Pisa,[21] making pendulums and rolling balls down inclined surfaces. However, what interests us is his work in astronomy and the literally devastating effect it had on the geocentric model.

Galileo observed the sky with several homemade telescopes, achieving a magnification of 33×. Surprisingly, it is now believed he wasn't the first person to do so. But, and this is what matters, he was the first person to do so and *publish* his work. It doesn't matter if one makes an Earth-shattering discovery or develops a mind-blowing theory if no one gets to hear about it!

He published his observations in his famous book *Siderius Nancius*, known to us as *The Starry Messenger*. In this he presented ideas stemming from his observations, and it was these that not only eventually put the nail in the coffin for the geocentric model but, alas, got him into trouble with the Inquisition.[22] Nevertheless, by the time the Inquisition had finished burning his books in Italy, and stopped just short of burning him, many copies had traveled throughout Europe, allowing others to see his work and expand on the ideas.

Here are the most important observations he made over several years, which changed everything.

He looked at the Moon and demonstrated the existence of lunar mountains, much to the chagrin of the Aristotelians, who had assumed the Moon to be a perfectly crystalline sphere. In addition, he attributed the visibility of the "old Moon," what we now call the new Moon, to Earthshine, sunlight reflected from Earth.

---

[21] Probably not leaning at the time, and probably just a fable.

[22] Even though he was a deeply religious man, it was another book by Galileo—*Dialogue Concerning the Two Chief World Systems—Ptolemaic and Copernican*, that got him into trouble with the Inquisition.

He observed that the Milky Way, previously thought to be an agglomeration of stellar matter in the atmosphere, was now seen to be an endless collection of stars.[23] He saw stars that could not be seen with the naked eye, as they were too faint. This observation was the first step in a long process that culminated in the correct description of the Milky Way as a galaxy.

Considered to be one of his most spectacular discoveries, occurring on January 7, 1610, he observed the moons of Jupiter. Observing over several nights he saw the moons change position, and correctly deduced they were orbiting Jupiter themselves.

These discoveries raised a storm. Kepler wrote to Galileo longing for a telescope to see the moons for himself, and some colleagues refused to believe it (Florentine astronomer Francesco Sizzi), while others refused to even look through the telescope for themselves (philosopher Giulio Libri). The significance of this discovery was much more important than the existence of "additional planets." It gave credence to the Copernican model that Earth was not the center of the universe but is only a planet with a Moon. The inference was obvious. If Jupiter, a planet, had moons, then Earth, with its known Moon, was just another planet, and not unique.[24] The four moons he discovered, now known as Io, Europa, Ganymede and Calisto, are referred to as the Galilean moons, in Galileo's honor.

Galileo also looked at Saturn, and although his telescope couldn't resolve the rings, he understandably believed he saw two moons, one on each side of the planet.

A series of observations that once again caused much consternation to the Aristotelians was his discovery of sunspots. Their perfect Sun was covered with spots that, over a short time, formed and changed, and by timing their movement it was implied that the Sun was rotating.

Finally, one set of observations conclusively showed that one planet was not orbiting Earth but was orbiting the Sun, and by inference so was Earth. These were his observations of Venus. One of the main objections to the Copernican model was the apparent absence of particular phases for Venus and Mercury. However, when Galileo saw all the phases this vindicated the Copernican system and ruled out utterly the Ptolemaic model. The following series of diagrams will show what phases were expected, and what Galileo did see (Fig. 2.6).

Even though his work was slowly gaining acceptance, it nevertheless caused great consternation to the established Church, and inevitably, he was called to Rome to attend the Inquisition, even though he was 70 years of

---

[23] He also looked at the Pleiades star cluster and saw 36 stars, whereas only 7(?) can be seen with the naked eye.

[24] The Moons also provided more data for Kepler's Third Law.

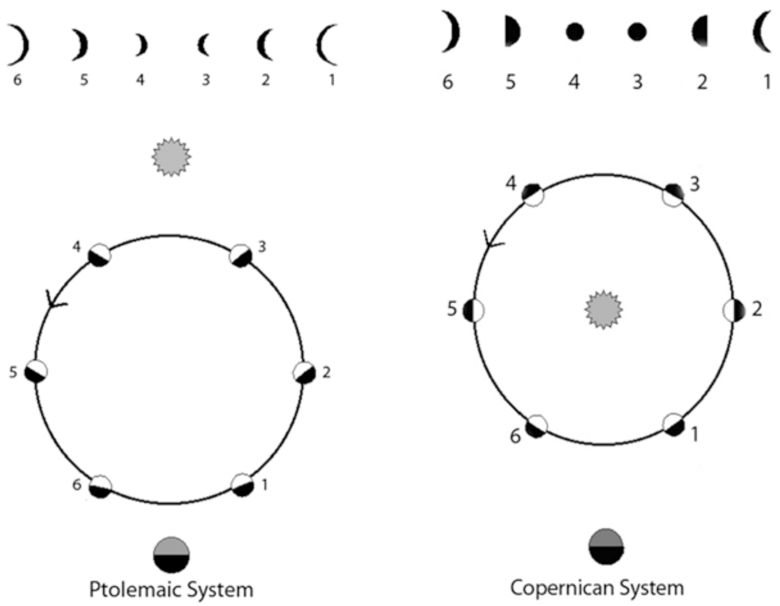


Fig. 2.6. Galileo's description of the phases of Venus

age. The sordid details need not concern us here,[25] but the end result will. He was forced to recant his views, and sentenced to house arrest for the rest of his life. His book was recalled and destroyed, but luckily several copies were smuggled out and reprinted in Leyden, Holland.

Galileo was at heart an experimental physicist, and his contributions to science opened the heavens to further investigations, a process that continues to this day.

2.1.9 Newton—The Genius

We now turn our attention the last player in this saga, who, building on the foundations of the previous participants, put the science of astronomy on a firm footing. We are, of course, talking about Isaac Newton (1642–1727).

Not only did the great man work on and publish on such diverse topics as light, optics, calculus and telescopes,[26] he also put forward the first serious proposal as to why planets move around the Sun and why apples fall to

---

[25] Everyone should read the details of this trial, as it shows the sordid depths the Church was willing to sink to in its attempts to prevent the truth from being told.

[26] And along the way, developed the reflecting telescope, the main instrument of choice for most amateur astronomers for the past 150 years.

the ground. What concerns us here, however, is his work on the forces that move the planets and how they move. This work was published in his magnificent three-volume opus entitled *Principia*.

Before we discuss these laws of motion, we should define a few concepts. You will probably already know these, but one or two may surprise you, and we will discuss them after listing them.

- *Speed*—the rate at which an object moves, i.e., the distance traveled per unit time [m/s; mi/hr].
- *Velocity*—an object's speed in a certain direction, e.g., "10 m/s moving east".
- *Acceleration*—a change in an object's velocity, i.e., a change in either speed or direction is an acceleration [m/s^2].

The first concept speaks for itself. The second may appear a bit odd. After all, who says "I am traveling down the road at 40 km a minute in a north-easterly direction?[27]" But this relates to the last concept that shows that acceleration need not be an increase or decrease in speed; it can also mean a change in direction, and this is where orbits come into play. For instance, imagine a moon in a circular orbit around a planet, with a uniform, or constant, speed. Because it is moving in a circular orbit, it is changing its direction at every instant of time, and a change in direction is acceleration. So the moon is undergoing acceleration,[28] not by changing its speed but by changing its direction!

Newton labored for several years, which resulted in his three laws of motion. Once again we will present them as a formal statement, and then explain them in a less formal manner.[29]

1. A body at rest or in motion at a constant speed along a straight line remains in that state of rest or motion unless acted upon by an outside force.
2. The change in a body's velocity due to an applied force is in the same direction as the force and proportional to it, but is inversely proportional to the body's mass.
3. For every applied force, a force of equal size but opposite direction arises.

The first law can be explained like this. Imagine a spaceship in space, with no other objects, be they planets, stars or whatever, anywhere nearby; in fact, assume the spaceship is the only thing around for millions of light years. Now, according to the law, the spaceship will, if already motionless,

---

[27] The type of person who has a compass in their car perhaps.

[28] The direction of acceleration is towards the planet, a consequence that need not concern us here, as we would have to delve into vector analysis.

[29] Newton credited the first two laws to Galileo and the last to Christopher Wren and Christian Huygens.

remain motionless forever, unless something acts upon it, for instance, the gravity of a planet that appeared nearby, or an impact from a micrometeorite. In addition, consider the same spacecraft in the same unlikely scenario, but this time moving with a constant speed, neither increasing nor decreasing its speed. If there is nothing else around, it will continue to move at that constant speed, in a straight line, forever!

In reality, of course, the planets, asteroids, the Sun, and the moons do affect a spacecraft in the Solar System; everything in fact, will alter its motion. Ever wondered about those tiny little rockets on space missions to the planets? Well, they are there to make course corrections to the vehicle as it progresses to its target, as its motion is constantly being affected by the gravitational force of everything in the Solar System.

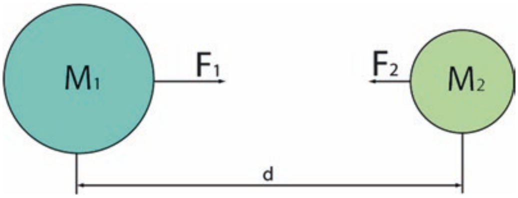
The second law can be thought of like this. Imagine you have a cricket ball and a cannonball, both the same size, but naturally the latter will have more mass. They are then thrown with exactly the same force. The cricket ball will travel further than the cannonball. Similarly, if you have two cricket balls of precisely the same mass, but one is thrown with greater force, it will travel a larger distance.

The final law explains something that many people will have experienced for themselves. Both you and a friend are standing on ice. You push your friend away, and not only do they move away from you, shouting expletives with arms waving, but you move in the opposite direction to them. Another example is a rocket launch. The rocket is propelled upward by a force equal and opposite to the force with which the gas exhaust is expelled out of its back.

Newton then went on to perhaps his greatest triumph, showing that gravity was the force that made the planets move in a stately motion around the Sun. He was able to show that the gravitational attraction between any two objects is dependent firstly on the masses of the two objects and secondly on their distance apart.

A more mathematical definition would be that the force is proportional to the product of the masses, and inversely proportional to the square of their distance apart. Mathematically:

$$F_1 = F_2 = G\frac{(M_1 M_2)}{d^2}$$



M_1—Mass of first object (kg).
M_2—Mass of second object (kg).
d—distance apart (m).
G—Gravitational constant = 6.67×10^{-11} N•(m/kg)^2.

What this means in plain English is that the larger the masses, the larger the force of attraction, but with the caveat that as the distance between the two objects increases, the force of attraction decreases. Note that the distance, *d*, is from the center of the mass of the objects, which can be considered to be approximately true for spherical objects such as planets and large moons, but not for potato-shaped asteroids. The gravitational constant, denoted by the letter *G*, is an empirical physical constant involved in the calculation of gravitational force between two bodies. The resulting force will be in the units called Newtons[30]—*N*. An example of the formula in action can be seen in Math Box 2.3.

At this point it is well worthwhile looking at the formula in some detail. Notice how one can work out the gravitational force perfectly for two objects and get a precise solution. However, it is not so simple for three or more objects,[31] and in fact there is no one single formula for any system that has more than two objects. Don't think for one second that the forces and motions cannot be worked out in this scenario, they can, but it involves much more rigorous mathematics. Just imagine, if you will, the complexity of, say, accurately calculating the gravitational effect of all the planets in the Solar System, along with their attendant moons, the Sun and the asteroids, on the behavior of a spacecraft, all moving so that their distances apart constantly change. It is complicated.

Also, consider this scenario. Imagine that, for some reason, the Sun were to completely and utterly disappear.[32] Then, with Newton's concept, Earth would "react" to this instantaneously! This means that using Newton's description, gravity propagates faster than light. Thus you can immediately see that although we can use Newton's formula to get very accurate results, it cannot be the correct description of gravity. In fact, this became apparent early on when astronomers tried to explain the strange motion of Mercury around the Sun and measure the deflection of light rays by gravity. We had to wait for another genius to give us the true description of gravity, and that genius was Einstein.

---

[30] 1 Newton, N, is the force of Earth's gravity on a mass of about 0.102 kg, or, 1 N = 0.10197 kg × 9.80665 m/s^2

[31] A partial solution has been worked out for three objects, but nothing as exact as Newton's formula.

[32] I don't mean be eaten by a black hole. I mean vanish!

Thought Question 2.3

On the surface of the Earth you weigh 100 kg. You then visit the International Space Station. How would your weight and mass differ?

However, the formula works very well when one is considering velocities much smaller than the speed of light, and with situations where the gravitational force is small, or the masses involved are small.

There is no doubt that Isaac Newton was a genius, and even though he understood *how* the planets moved due to gravity, he did admit he didn't know what gravity was. Nevertheless, his insight started a stream of discovery that still flows today, and although his life was, at times, beset with controversy and intrigue—he was after all an alchemist—he was a towering intellect.

Math Box 2.3: Newton's Universal Law of Gravitation

Assume you are on Earth, with a mass of 70 kg, the Earth's Moon, mass 7.35 × 10^{22} kg, and the Andromeda Galaxy, mass 1.41 × 10^{42} kg. The Earth-Moon average distance is 3.8 × 10^{8} m. The Earth-Andromeda Galaxy distance is 2.4 × 10^{22} m.

Calculate the gravitational force of attraction between the following:

You and the Moon.

You and the Andromeda Galaxy.

Determine how many times greater the stronger force is to the weaker.

$$F_1 = F_2 = G\frac{(M_1M_2)}{d^2}$$

$$F_1 = F_2 = 6.67\times10^{-11} N\bullet(m/kg)^2\,\frac{(70\ kg\times7.35\times10^{22}\ kg)}{(3.8\times10^{8}\ m)^2}$$

$$F_1 = F_2 = 6.67\times10^{-11}\ N\bullet(m/kg)^2\,\frac{(5.15\times10^{24}\ kg)}{(1.45\times10^{17}\ m)}$$

$$F_1 = F_2 = 2.37\times10^{-3} N$$

$$F_1 = F_2 = 6.67\times10^{-11}\ N\bullet(m/kg)^2\,\frac{(70\ kg\times1.41\times10^{42}\ kg)}{(2.4\times10^{22}\ m)^2}$$

$$F_1 = F_2 = 6.67 \times 10^{-11} \ N \bullet (m / kg)^2 \frac{(9.88 \times 10^{43} \ kg)}{(5.76 \times 10^{44} \ m)}$$

$$F_1 = F_2 = 1.14 \times 10^{-11} N$$

Ratio of stronger force to weaker force:

$$\frac{2.37 \times 10^{-3}}{1.14 \times 10^{-11}} = 2.08 \times 10^{8}$$

Put another way, the force of gravitational attraction between you and the Moon is nearly 200 million times stronger than the force between you and the Andromeda Galaxy.

2.1.10 The Greenhouse Effect

Many of us have heard of the Greenhouse Effect, but few actually understand what it means. It is in fact a simple concept but has immense implications for us on the Earth. The effect occurs, surprisingly, on any planet or moon, in the Solar System, that has an atmosphere, to a greater or lesser extent.

Here's what happening.

Every second, the Sun emits a vast amount of energy, but the amount that actually hits the Earth's atmosphere is about 1368 W/m^2 (watts per square meter).[33] This amount of energy is reduced by the atmosphere however, so that the amount that arrives at the surface is closer to 1000 W/m^2 assuming clear conditions when the Sun is at the zenith. Incidentally, the light from the Sun at the top of Earth's atmosphere is composed of about 50% infrared light, 40% visible light, and 10% ultraviolet light.[34]

This solar energy falls on the surface and, not unexpectedly, is absorbed and heats up, as the surface, and oceans are rather good at this.[35] If we make an assumption that there is no atmosphere, then as the Earth's surface heats up a point would occur when this solar energy is emitted back into space as infrared radiation, due to the Earth acting as a blackbody (see the Sect. 1.7 –

---

[33] Also known as the Solar Constant.

[34] The atmosphere in particular filters out over 70% of solar ultraviolet, especially at the shorter wavelengths.

[35] We're all familiar with hot pavements or sidewalks, and hot sandy beaches, in the summer.

Size of Stars). This occurs when a balance is achieved between the amount of energy being radiated out into space equals the amount received from the Sun.

Now for a surprise, the temperature at which this balance occurs, has been calculated to be about 270 K, or – 23 °C, which is far below the freezing point of water. You may have noticed that we do not live on a giant snowball, so something must be happening that prevents this low temperature, and that is obviously an atmosphere. In some manner, the atmosphere holds in the infrared light and warms the surface of the Earth.

However, it is not just the fact that we have an atmosphere that is responsible, but rather some particular gases that make up the atmosphere. In the Earth's atmosphere these gases comprise of Carbon Dioxide, CO_2, water vapour, and Methane, CH^4, and are known as Greenhouse gases.[36] Basically, these gases absorb the infrared radiation that has been emitted from the Earth's surface, and more-or-less prevent it from escaping back into space for a short while. Thus, by "trapping" some the heat, they warm the surface of a planet. The gases do not hold onto the heat indefinitely, otherwise the surface would continue to heat up and get hotter but instead the trapped heat is slowly re-emitted from the gases, that's why it is cooler at night. It is important to note that it is a thin layer of the atmosphere that absorbs the infrared radiation, and that this layer is the one in contact with the Earth's surface.

This then is the Greenhouse Effect.[37]

On the Earth, the relative contributions to the Greenhouse Effect are.

- water vapor, 36–70%
- carbon dioxide, 9–26%
- methane, 4–9%
- ozone, 3–7%

The Greenhouse Effect occurs on any planet that has an atmosphere with Greenhouse gases. Thus, Venus, Mars, Jupiter, Saturn, Uranus, and Neptune will all feel its effect, depending on the concentration of the gases present. See Table 2.1.

Mars, which has a very thin atmosphere, with an average surface pressure is only about 610 pascals (0.088 psi) which is less than 1% of the Earth's value. It is primarily composed of carbon dioxide (95%), molecular nitrogen (2.8%) and argon (2%). It also contains trace levels of water vapor,

---

[36] Clouds also absorb and emit infrared radiation and thus affect the radiative properties of the atmosphere.

[37] The term "Greenhouse effect" is actually a misnomer since in an actual greenhouse, the heating is due to the reduction of convection, while the "greenhouse effect" works by preventing absorbed heat from leaving the planet through radiative transfer.

Table 2.1 The Greenhouse effect on the terrestrial planets

Planet	"No greenhouse effect" Average surface temperature	Actual surface temperature (average)
Venus	−43 °C	470 °C
Earth	−18 °C	15 °C
Mars	−55 °C	−50 °C

oxygen, carbon monoxide, hydrogen and noble gases. Therefore, the weaker greenhouse effect in the Martian atmosphere, about 5 °C, versus 33 °C on Earth can be explained by the low amount of the greenhouse gases.[38]

Looking at Venus, we see a very thick atmosphere with a pressure of 600 pascals (1350 psi), or about 90 times that of the Earth. The atmosphere is mainly composed of 96.5% carbon dioxide, 3.5% nitrogen, and traces of other gases,[39] most notably sulphur dioxide. Due to the large amount of Greenhouse gases and immense pressure, a Runaway Greenhouse Effect is believed to have occurred.

It goes something like this. Early in the history of Venus, it may have had global ocean. As the energy output of the early Sun increased, the amount of water vapor in the atmosphere subsequently increased, thus increasing the temperature and consequently increasing the evaporation of the ocean, but also increasing the Greenhouse Effect. This means it got hotter, eventually giving rise to the situation where the oceans boiled, and all of the water vapor entered the atmosphere. This scenario may explain why there is so little water vapor in the atmosphere of Venus today.

The Greenhouse may also have a part to play on the Jovian planets, albeit a very small one. However, analysis of the data from both the Cassini mission and Huygens probe suggests that the contribution due to the Greenhouse effect is substantial on the Saturnian moon, Titan which is the only nitrogen-rich dense atmosphere in the Solar System aside from Earth's.

---

[38] Recently, unexpected levels of methane were detected in the Martian atmosphere, which could be interpreted as a biosignature for life on Mars. However, the it is highly controversial and lacks a scientific consensus as there are some suspicions that suggest it may instead be caused by the terrestrial contamination or a misinterpretation of measurement raw data.

[39] In September 2020, it was announced that phosphine, a potential biomarker, had been detected in the atmosphere of Venus. However, the detection of phosphine was suggested to be a possible false positive in October 2020, but in April 2021, further analysis of the data and new observations suggest that the initial conclusion may have been correct after all. Time will tell.

Here we leave our theoretical exploration[40] of the Solar System and now concentrate, albeit briefly, on the observational aspects of the Solar System.

Thought Question 2.4

Venus and Mars have about the same percentage of the Greenhouse gas CO_2 *in their atmospheres, yet their temperatures are very different. What could be the main reason for this temperature discrepancy?*

2.2. Observing the Solar System

Observing the planets, their moons, the asteroids and the Sun is a wonderful pastime for amateur astronomers, and many have devoted most of their observing time to just such a passion, but there is no way we can give it the full coverage it deserves in this book. Each planet really needs its own book, and to that end we have listed such books in the appendices. However, there are a few things one can look out for, without any optical equipment at all, except the naked eye, and so that is what will be presented here.[41] In fact, regard them as observing challenges, now known as the "The Inglis Naked-Eye Planetary Phenomena Observing Challenge," TINEPOC for short.[42]

The positions for the planets can easily be found online, or by using planetarium software.[43]

2.2.1 The Moon

Observing our Moon can be a lifelong study, and will of course need a telescope, but there is one specific observation that can be done without resorting to optical aid—to glimpse the very young or very old Moon. In fact, many observers spend an inordinate amount of time doing just this. Try to locate the

---

[40] We shall not bother with the Titus-Bode law, as it is not believed to be a law at all but rather a rule, and possibly a coincidence.

[41] Solar observing has its very own techniques and equipment. Suffice to say NEVER look at the Sun through a telescope, or even just with the naked eye. You'd be a fool to do so.

[42] Perfect!

[43] Naturally, not all the planets will be mentioned here, as this is a naked-eye exercise.

Moon either as soon as possible after it is new, or as late as possible before it becomes new. It helps if you have a good horizon view to either the east or west.

2.2.2 Mercury

The planet closest to the Sun has its own attendant problems, because of its location in space. At greatest elongation, which varies between 18° and 28° due to its elliptical orbit, it can only be glimpsed for about 1 h (considerably less when its elongation is the minimum value), either before sunrise or after sunset. Therefore, the sky will still be bright, and so the challenge here is, basically, to just find it. Needless to say you will need a more or less completely unobstructed view of the horizon. That's to say, no trees, houses, breweries, etc., that could obscure the view. It's tricky, but once glimpsed, you'll wonder why you've never seen it before, and the first time you actually locate it will be an event you won't easily forget.

2.2.3 Venus

Believe it or not, it is possible to observe the phases of Venus with the naked eye. When a phase is at its most extreme, those lucky individuals blessed with exceptionally acute eyesight can see it. It will be, however, at the limit of human perception. This is because the angular resolution of the naked eye is about 1 minute of arc, whereas the apparent disk of Venus' extreme crescent measures between 60.2 and 66 seconds of arc, depending on the distance from Earth. Of course, perfect atmospheric conditions will be necessary. This is not something that is hearsay but comes from many substantiated reports from observers worldwide.[44]

2.2.4 Jupiter

OK, this section isn't actually about Jupiter itself but rather its Galilean moons, and even then we are talking about only Ganymede and Calisto. Unknown to many amateur astronomers is the fact that all four moons are

---

[44] The author, although plagued with poor eyesight now, did manage to observe an elongated Venus with the naked eye when I was much younger and had all bodily parts in working order.

bright enough, apparent magnitudes between 4.6 and 5.6, when Jupiter is at opposition, to be hypothetically visible from Earth without a telescope—if only they were further away from Jupiter. However, the problem that arises is twofold. Firstly, Io and Europa are too close to Jupiter to be resolved with the naked eye. However, the maximum angular separations of Ganymede and Calisto are 351 arc seconds and 618 arc seconds, respectively, and thus are the likeliest targets for potential naked-eye observation. The second problem is the glare from Jupiter itself, which floods the eye with light, thus preventing observation of the satellites. In order to remedy this try obscuring Jupiter with an object, e.g., a tree branch, telephone pole or anything similar that is perpendicular to the plane of the moons' orbits.[45]

2.2.5 *Uranus*

The last challenge is, like with Mercury, to just locate Uranus. At opposition, Uranus has a magnitude of around 5.7, and thus, from a very dark sight will be within reach of those with excellent eyesight. The problem here is that due to its faintness, it will appear to be just like a faint star, set among many other faint stars, and so a prerequisite for a successful identification is to have a good knowledge of the sky in which it will be (hopefully) observed. This is where experienced amateurs excel, as they know the night sky intimately.

For all the above observations, good eyesight is essential, along with transparent skies and little or no light pollution.

Now let us leave the Solar System and explore the universe—stars, galaxies, and even further afield.

Thought Question 2.5

Which naked-eye planet is missing from the above list, and why?

Problems

1. An exoplanet orbits its star in 45.66 days. What is its distance from the star?
2. An asteroid is discovered with the following parameters: $a = 5.00 \times 10^8$ km, and $c = 1 \times 10^8$ km. Determine its eccentricity.

---

[45] I have tried this and truly believe that although I didn't see actual separate moons, I did see an elongation of Jupiter, the Moons being located on either side of Jupiter.

3. The distance between the Earth and Moon (d) is 3.84×10^8 m. What is the magnitude of gravitational force each exerts on the other? Use the following data; Earth's mass (m_E) = 5.97×10^{24} kg, Moon's mass (m_M) = 7.35×10^{22} kg, and Universal constant (G) = 6.67×10^{-11} N m^2/kg^2.

Chapter 3

Spectroscopy and the Spectral Sequence

3.1. Spectra and Spectroscopy

Spectroscopy is an amazing subject; from just looking at the light from an object, we can tell how hot it is, how far away it is, in which direction it is moving,[1] if it is rotating, and (from all this data) infer its age, its mass, how long it has left to live, etc. From this point on in the book, a star will be referred to by its spectral classification.

Determining a star's classification is a theoretically easy task, although it can be difficult in practice. What is needed is a spectroscope. This is an instrument that looks at the light from a star in a special way by utilizing either a prism or a diffraction grating for analysis.

You're probably already aware that white light is in fact a mixture of many different colors, or wavelengths, so it's safe to assume that the light from a star is also a mixture of colors. Indeed it is, but usually with an added component. Using a spectroscope mounted at the eyepiece end of the telescope,[2] light from the star can be collected and photographed (these days with a CCD camera). The end result is something called a *spectrum* (pl.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_3

spectra). Many amateur astronomers are now making some very good observations of star's spectra.

Basically, a spectrum is a map of the light coming from a star. It consists of all of the emitted light, spread out according to wavelength (color) so that the different amounts of light at different wavelengths can be measured. Red stars have a lot of light at the red end of the spectrum, while blue stars have a correspondingly larger amount at the blue end.[3] This band of color is referred to as the *continuous spectrum.* However, the important point to note is that in addition to this light, there will be a series of dark lines superimposed upon this rainbow-like array of colors. These are called *absorption lines*, and they are formed in the atmosphere of the star. In a few rare cases, there are also bright lines, called *emission lines*. Although comparatively rare in stars, these lines are very prominent in nebulae.

The electrons in the atoms located in the surface layers of a star can only have very specific energies, rather like the specific heights of the rungs of a ladder. Sometimes an electron in an atom of, say, hydrogen, can be "knocked" from a lower energy level to a higher energy level, maybe by a collision with another atom, or by gaining energy from the high-energy light, say, UV, emitted from a nearby star. Eventually, however,[4] it will fall back down to a lower level. The energy that the atom loses when the electron returns back to its original level must go somewhere, and it often goes to emitting a photon of light. This emitted photon has a unique property—it has the exact amount of energy that the electron gained in the first place, and then loses, which in turn means that the photon has a very specific wavelength and frequency.

When hydrogen gas is heated to a high temperature, the number of collisions between atoms can continually bump electrons to higher energy levels, and an *emission line spectrum* results. This consists of the photons that are emitted as each electron falls back to lower levels.

The origins of the absorption lines are due to the differing amounts of elements in the cooler atmosphere of the star (recall that in addition to hydrogen and helium, there are additional elements, or metals, present, although in minute quantities). Not only are photons emitted, but they can also be absorbed. This process causes the electrons to jump up in energy to a higher level. But this can only happen if the photon has the precise amount of energy required. Too much, or too little, by even a minuscule amount, and the photon will not interact with the electron.

---

[1] We can only easily deduce whether an object is moving away from us or toward us. To measure if it is moving laterally to us requires some complicated measurements.

[2] Some spectroscopes place the prism or grating in front of the telescope, and thus the light from *every* star in the field of view is analyzed simultaneously. This is called an *objective spectroscope*. The drawback is the considerable loss of detail (i.e., information about the stars), but initial measurements can be made.

[3] See Chap. 1 and the discussion of Wien's Law.

In hydrogen gas, an electron moving from level 2 to level 1 will emit a photon that has a wavelength of 121.6 nm; an electron absorbing a photon of this wavelength will jump from level 1 to level 2. Such jumps from different levels are called *transitions*. Thus, in the above example, an electron undergoes a transition from level 1 to level 2, with absorption of a photon of wavelength 121.6 nm. Figure 3.1 shows the allowed energy levels of hydrogen and the wavelengths that occur for downward transitions. Also shown are the absorption and emission spectra.

Note that in Fig. 3.1, the dark absorption lines and the bright emission lines occur at exactly the same wavelengths, regardless of whether the hydrogen is emitting or absorbing the light. Emission lines are simply the

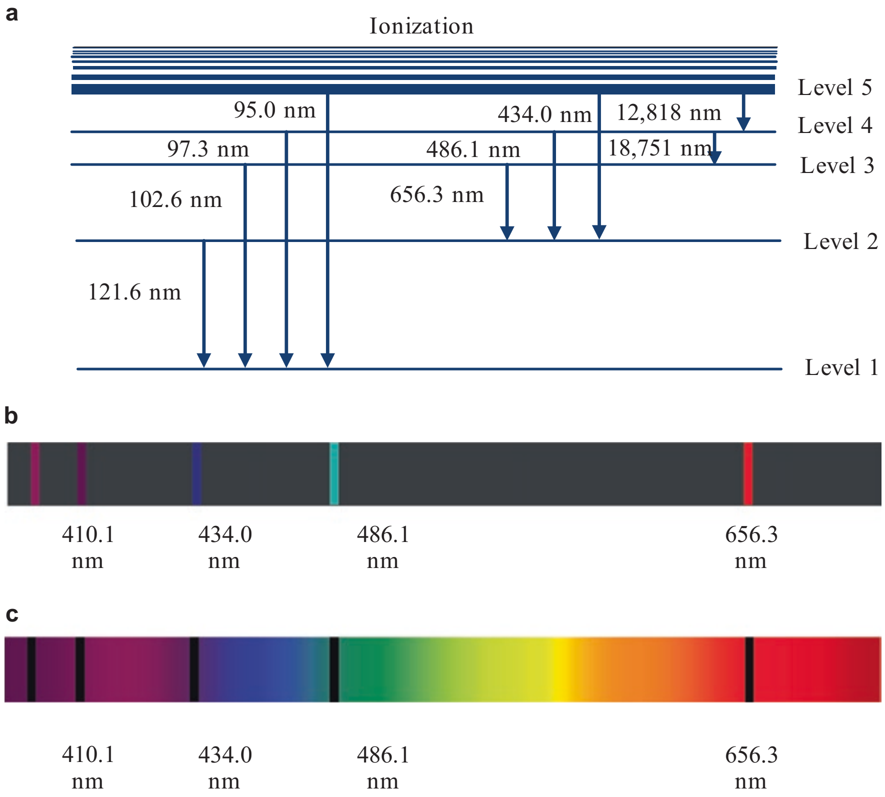


Fig. 3.1. Allowed energy levels of hydrogen. (**a**) The wavelengths of various energy level transitions in hydrogen (the transitions shown are only a few of the many that occur). (**b**) Visible emission line spectra, showing transitions that occur from high energy levels downward to level 2 for hydrogen. (**c**) Absorption line spectra, showing transitions that arise from energy level 2 to higher levels. These absorption and emission lines of hydrogen are called the *Balmer lines*

result of downward jumps, or transitions, of electrons between the energy levels, while absorption lines are upward transitions.

What must be stressed here is that for an electron to make a transition as an absorption event, the amount of energy involved must be precisely the right amount for the transition to occur. Too much or too little and the transition will not occur; it is a very exact phenomenon. Let's look at this in more detail.

Imagine that an electron has an initial (lower) amount of energy E_{i} (i corresponding to its initial energy). The possibility exists that, under the right circumstances, the electron can make a transition to a higher energy level, E_{h} (h corresponding to its higher energy). The energy needed to make the transition, E_{hi}, is thus $E_{\mathrm{h}} - E_{\mathrm{i}}$. In the same manner, the energy emitted when an electron makes a transition from a higher to a lower level is exactly the same value, only this time the energy is not absorbed but emitted, or released, by the electron.

The famous physicist Niels Bohr devised a way to determine the wavelength of light needed to make a transition specifically for the hydrogen atom, using quantum mechanics. This method makes use of the levels that the electrons are in, and an example of this technique is given in Math Box 3.1.

Another very important point to understand is that the emission will occur in *any* direction. What this means is that energy absorbed by the electron(s) will not be entirely replaced by emitting energy, but rather the emitted energy will be given off in all directions. Thus, looking directly at the origin of the light, say, a star through a cloud of gas, we will only observe the absorption spectrum, whereas the emission spectrum will be seen from any direction. (Fig. 3.2 may help in understanding this concept.)

The energy levels of electrons in each chemical element are unique—a "fingerprint" that results in each element having its own distinct spectral lines. Hydrogen is a very simple element, with only one electron, but in those elements with many electrons and energy levels, the corresponding spectra can be very complex.[5]

The factor that determines whether an absorption line will arise is the temperature of a star's atmosphere. A hot star will have different absorption lines than a cool star. Thus, examining its spectrum and measuring various aspects of the absorption lines determines the classification of a star. A very important point that should be emphasized is that primarily the temperature of the atmosphere and not the temperature of the core determine the spectral classification of a star. The structure of the absorption lines themselves can also be examined, and this gives further information on pressure, rotation, and even the presence of a companion star.

---

[4] In most cases this is a very short time, a few millionths of a second.

[5] Do not confuse the energy of a photon with the brightness of emission lines. The energy is what determines the wavelength, or color, of the line, whereas the total number of photos emitted at that wavelength determines the brightness of the line.

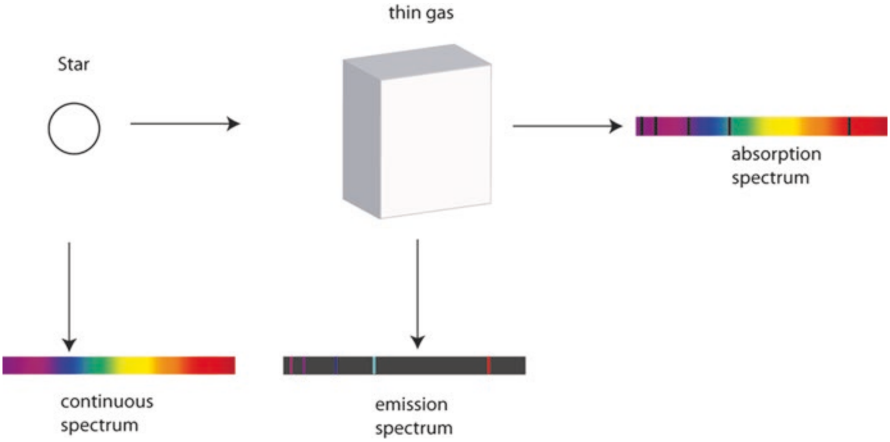


Fig. 3.2. Different types of spectra; a continuous, emission and absorption spectra are observed by viewing the source of light (energy) and a thin gas (nebulae) at various inclinations or directions

Math Box 3.1 Hydrogen Atom Transitions

The formula for determining the wavelength of light of any spectral line of the hydrogen atom is given by

$$\lambda = \frac{1}{R}\left(\frac{n_h^{\,2} n_i^{\,2}}{n_h^{\,2} - n^2_{\;i}}\right)$$

where R is the Rydberg constant (a combination of other fundamental physical constants) 1.097 × 10^7 m^{-1}.

nh is the quantum number of a higher level.

ni is the quantum number of an initial level.

λ is the wavelength of light.

Determine the wavelength of light emitted when an electron makes a transition from level 3 to level 2.

$$\lambda = \frac{1}{R}\left(\frac{n_3^{\,2} n_2^{\,2}}{n_3^{\,2} - n^2_{\;2}}\right)$$

$$\lambda = 9.11 \times 10^{-8}\left(\frac{36}{9-4}\right)$$

$$\lambda = 656.467 \ \text{nm}$$

This is the famous hydrogen alpha (“aitch α”) line, probably the most important spectral line in astrophysics, which is responsible for the characteristic red color of emission nebulae. So often imaged, very rarely observed visually.

Thought Question 3.1

You use a camera to obtain an image of a nebula. After processing accurately, you see it has a lot of blue colors, but hardly any red. What does this tell you about the region in and around the nebula? (Note: it is not a Planetary Nebula).

3.2. Stellar Classification

We have seen how stars are distinguished by their spectra (and thus temperature). Let's now think about the spectral type. For historical reasons a star's classification is designated by a capital letter; thus, in order of *decreasing* temperature:[6]

OBAFGKMLTY
R
N
S

The sequence goes from hot blue stars, types O and A, to cool red stars, types K, M and L. In addition, there are rare hot stars called Wolf-Rayet stars, classes WC and WN, exploding stars, Q, and peculiar stars, P. The star types R, N and S actually overlap class M, and so R and N have been reclassified as C-type stars, the C standing for carbon stars.

Over the past few years, due to an increase in the abilities and sophistication of spectroscopes, detectors and telescopes, several new classes, L, T and Y, have been added to the sequence. These new classes refer to the infrared spectra of very cool red stars, and the type of star referred to as a brown dwarf. Type T is also referred to as *methane dwarfs* with an abundance of the compound in its spectra, while Y stars are *sub-brown dwarfs* and *substellar objects*. This latter name is for those strange objects that

---

[6] A brilliant astronomer, Cecilia Payne-Gaposchkin, discovered how the now familiar order, OBAFGKM, really works. She found that all stars are made primarily of hydrogen and helium and that a star's surface temperature determines the strength of its spectral lines. For instance, O stars have weak hydrogen lines because, due to their high temperature, nearly all the hydrogen is ionized. Thus, without an electron to "jump" between energy levels, ionized hydrogen can neither emit nor absorb light. On the other hand, M stars are cool enough for molecules to form, resulting in strong molecular absorption lines.

cross the divide between a star and a planet. Not many have been discovered and so much has yet to be learned about them.

Furthermore, the spectral types themselves are divided into ten spectral classes beginning with 0, 1, 2, and 3 and so on up to 9. A class A1 star is thus hotter than a class A8 star, which in turn is hotter than a class F0 star. Further prefixes and suffixes can be used to illustrate additional features:

Star with emission lines (also called *f* in some O-type stars)	*e*
Metallic lines	*m*
Peculiar spectrum	*p*
Variable spectrum	*v*
Star with a blue or red shift in the line (for example, P-Cygni stars)	*q*

Also for historical reasons, the spectra of the hotter star types O, A and B are sometimes referred to as early-type stars, while the cooler ones (K, M, L, C and S) are later-type. F and G stars are designated intermediate-type stars.

Thought Question 3.2

Can you think of a reason as to why there are three spectral classes of Brown Dwarfs and not just one?

Because the spectral type is so important, we will explain further how its surface temperature affects the appearance of a spectrum. We will consider the Balmer lines of hydrogen, mainly because these are by far the easiest to understand.

Hydrogen gas makes up 75% of a star, yet the Balmer lines do not always show up in a star's spectrum. The Balmer absorption lines are produced when an electron undergoes a transition from the second energy level to a higher level, by absorbing a photon with the correct amount of energy. If, however, the star is hotter than about 10,000 K, the photons coming from the star's interior have such a high energy that they can easily knock electrons out of hydrogen atoms in the star's atmosphere. This is the process of *ionization*. Now that the hydrogen atom has lost its electron, it cannot produce absorption lines. So, the Balmer lines will be relatively weak in the spectra of such hot stars (for example, O-type stars, up to type B2).

On the other hand, if the atmosphere of a star is cooler than 10,000 K, most of the hydrogen atoms are in the first energy state. Many of the photons passing through the atmosphere do not have enough energy to boost the electron from the first to the second energy level. Therefore, very few atoms will have electrons in the second level, and only these few electrons will absorb the photons characteristic of the Balmer lines. This results in the lines being almost absent from the spectrum of cool stars, such as M0 and M2 stars.

For the Balmer lines to be prominent, a star must be hot enough to excite the electrons out of level 1 (also known as the *ground state*), but not so hot that the hydrogen becomes ionized. If a star has a surface temperature of around 9000 K, it will have the strongest hydrogen lines (for example, the A0 to A5 stars).

The Balmer lines of hydrogen become increasingly prominent as you go from type B0 to A0. From A0 through to F and G class, the lines weaken and almost fade away. The Sun, a G2 star, has a spectrum dominated by lines of calcium and iron.

Finally, a star can also be classified by its *luminosity,* which is related to its intrinsic brightness, with the following system:

Hypergiants	0
Luminous supergiants	Ia
Intermediate luminous supergiants	Iab
Less luminous supergiants	Ib
Bright giants	II
Normal giants	III
Subgiants	IV
Main sequence (dwarfs)	V
Subdwarfs	VI
White dwarfs	VII

There are, of course, situations where the star's classification borders on that of another. For instance, a star classified as Ia-0 is a very luminous supergiant, verging on being a hypergiant.

It's evident that astronomers use a complex and very confusing system! In fact, several classes of spectral type are no longer in use, and the luminosity classification is also open to confusion. It will not surprise you to know that there is even disagreement among astronomers as to whether, for example, a star labeled F9 should be reclassified as G0! Nevertheless, it is the system generally used, and so it will be adhered to here. Examples of classification are:

α Boötes (Arcturus)	K2IIIp
β Orionis (Rigel)	B8Ia
α Aurigae (Capella)	G8 III
P Cygni	B1Iapeq
Sun	G2V
α Orionis (Betelgeuse)	M2 Iab

Table 3.1 Spectral classification

Spectral-type	Absorption lines	Temperature	True color	Notes	Brightest wavelength (color)	Examples
O	Ionized helium (HeII)	35,000 K+	Blue-white	Massive, short-lived	<97 nm (ultraviolet)	Stars of Orion's Belt
B	Neutral helium first appearance of hydrogen	20,000 K	Blue-white	Massive and luminous	97–290 nm (ultraviolet)	Rigel
A	Hydrogen lines singly ionized metals	10,000 K	White	Up to 100 times more luminous than the Sun	290–390 nm (violet)	Sirius
F	Ionized calcium (CaII), weak hydrogen	7000 K	Yellow-white		390–480 nm (blue)	Polaris
G	CaII prominent, very weak hydrogen	6000 K	Yellow	Sun is G-type	480–580 nm (yellow)	Alpha Centauri A, Sun
K	Neutral metals, faint hydrogen, hydrocarbon bands	4000–4700 K	Orange		580–830 nm (red)	Arcturus
M	Molecular bands, titanium oxide (TiO)	2500–3000 K	Red	Most prolific stars in galaxy	>830 nm (infrared)	Proxima Centauri, Betelgeuse
L	Metal hydrides, alkali metals	1300–2400 K	Scarlet (?)	L-type giants may form through collisions	>1000 nm (infrared)	LSR 1610-0400 (subdwarf) V838 Monocerotis (supergiant)
T	Methane	500–1300 K	Magenta (?)	Possibly very numerous	>1000 nm (infrared)	CFBDS 1448 (with exoplanet!)
Y	Ammonia (?)	<500 K	Black (?)	Crosses divide between planet/stars?	>1550 nm (infrared)	WISE 18828 + 2650 (~300 K; temp. of human body)

We conclude this section on spectral classification by explaining what the spectral type actually *refers* to.[7] You will recall that the classification was based on the detection of absorption lines, which in turn depend on the temperature of a star's atmosphere. Thus, the classification relies on the detection of certain elements in a star, giving rise to a temperature determination for that star. The classification can be summarized best by Table 3.1.

It is interesting to point out that the distribution of stars throughout the galaxy may not be what you assume. A casual glance at the stars you see in the night sky will give you several O- and B-type, a few A-type, some F- and G-type, a smattering of K- and more M-types. You may then think this is a fair picture of the type distribution throughout the remainder of the galaxy. You would be wrong! As we shall see in later sections, the vast majority of stars in our galaxy—over 76% of them—are the faint, cool and red M-type stars. The bright and hot O-type main sequence stars number less than 0.00003%. For every O-type star, there are about 1.7 million M-types!

Thought Question 3.3

A fellow amateur astronomer shows you a spectrum of a star she has taken which shows considerably more lines due to Calcium than those of Hydrogen. She then states, "obviously the star has more Calcium than Hydrogen". Is she correct?

3.3. Amateur Astronomical Spectroscopy

During most of its history, spectroscopy was firmly in the domain of the professional astronomer, at university research departments or at major observatories. Then, over time, smaller observatories at colleges were able to build for themselves usable spectroscopes. It was really only in the latter part of the twentieth century that it began, hesitantly, to appear in the hands of the amateur astronomer. Even then, it was usually done only by those individuals who were able to build spectroscopes for themselves.

However, in the twenty-first century, a revolution has occurred, where, for literally the same price as a premium eyepiece, spectroscopic equipment can be bought, and installed, for use at the telescope.

---

[7] Usually only the classes O, A, B, F, G, K and M are listed, as these are stars that can be observed with amateur equipment. The other classes are used and defined as and when they are needed.

Leading the revolution has been one man on a mission to make the subject acceptable to every amateur astronomer, and this is Tom Field. His very simple equipment, with associated software, allows one to take spectra of stars, nebulae, supernovae and even quasars, and then process the data to get meaningful results.

All the spectra in this book have been obtained using Field's "Star Analyzer" and software from amateur astronomers around the world.[8]

Let's now look at some spectra taken with such equipment, along with a detailed list of possible stars that can be used for yourself, as examples of the spectral sequence.

Look at Fig. 3.3 for a moment. Notice how the M-type stars have very broad lines at the red end of their spectrum. These are lines due to molecules

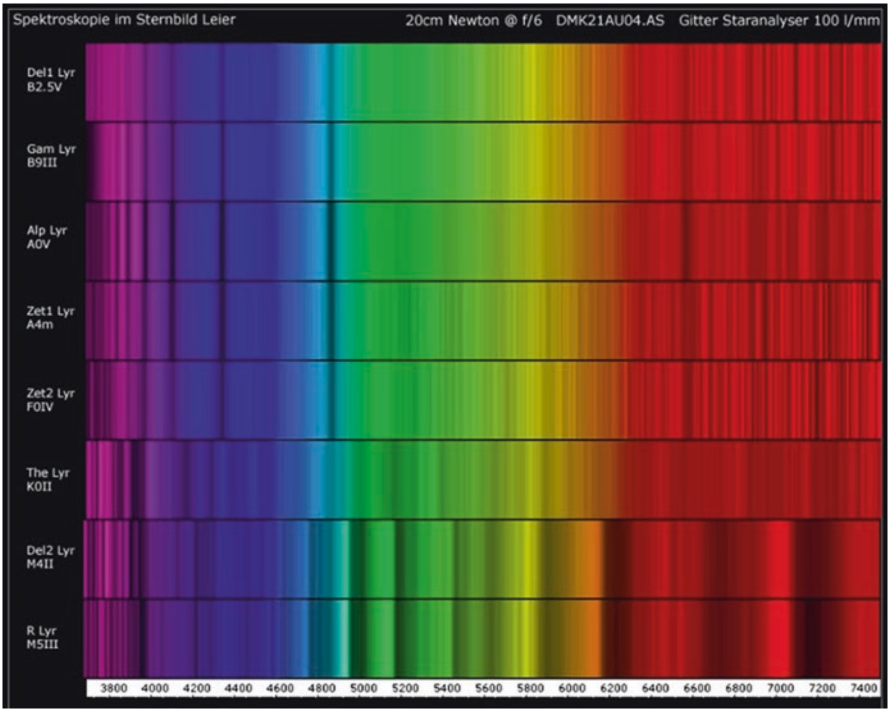


Fig. 3.3. Examples of the spectral sequence. The photo shows the spectra of several stars of differing spectral class. These spectra were obtained with a 20-cm Newtonian telescope, along with a star analyzer. (Spectrum courtesy of Torsten Hansen. Used with permission)

---

[8] There are, of course, many other spectroscopes available for the amateur astronomer, but I have mentioned Tom's due to its simplicity, price, and effectiveness. It is really very good, and no, I didn't get one free for mentioning him.

that can survive the cooler temperatures in the atmosphere of a red, M-type star. In addition, one can see that in an A-type star, the hydrogen Balmer alpha and beta absorption lines are stronger than for any other star, for the very simple reasons mentioned earlier.

Now let’s look at some observable examples. The month is the best for viewing.

–	HD 93129A	10h 43.9m	−59° 33′	February
7.0 m	−7.0 M	O3 If		Carina

An extraordinary star! This supergiant star, lying at a distance of about 11,000 light years, shines about five million times as brightly as the Sun. With a mass of 120 $M_{\odot}$ it is believed to be one of the most luminous stars in the entire Milky Way Galaxy.

θ Orionis C	θ Ori	05h 35.3m	−05° 23′	December
4.96 m	−5.04 M	O6		Orion

A member of the famous Trapezium multiple-star system in the Orion Nebula, this is a fairly new star, maybe only several thousand years old, and as a consequence most of its light is emitted at ultraviolet wavelengths. It is at a temperature of 45,000 K and has a diameter ten times that of the Sun.

15 Monocerotis	HD47839	06h 40.9m	+09° 54′	December
4.66vm	−2.3 M	O7		Monoceros

Both a visual binary and a variable star, this is located in the star cluster NGC 2264, which in turn is encased in a diffuse nebula.

Plaskett’s Star	HD47129	06h 37.4m	+06° 08′	December
6.05 m	−3.54 M	O8		Monoceros

Plaskett’s Star is actually composed of two stars, a spectroscopic binary system with an estimated mass of approximately 110 that of the Sun’s, making it one of the most massive known.

Gamma Cassiopeiae	γ Cas	00h 56.7m	+60° 43′	October
2.15vm	−4.22 M	B0 IV		Cassiopeia

A peculiar star in that it has bright emission lines in its spectrum, indicating that it ejects material in periodic outbursts. It is the middle star of the familiar W-shape of Cassiopeia.

Mirzim	β CMa	06h 22.7m	−17° 57′	December
1.98vm	−3.96 M	B1 II		Canis Major

This is the prototype of a class of variable stars now classified as β Cepheid stars, which are pulsating variables. The magnitude variation however is too small to be observed visually.

Algenib	γ Peg	00h 13.2m	+15° 11′	September
2.83vm	−2.22 M	B2 V		Pegasus

This is a member of the type β CMa (Canis Majoris) variable star and is at the southeastern corner star of the famed square of Pegasus.

Achernar	α Eri	01h 37.7m	−57° 14′	October
0.45vm0.45 vm	−2.77 M	B3 V		Eradinus

Achernar is a hot and blue star. Alas, it lies so far south that it can never be seen from more northern locations.

Aludra	η CMa	07h 24.1m	−29° 18′	January
2.45 m	7.51 M	B5 I		Canis Major

A highly luminous supergiant with an estimated luminosity 50,000 times that of the Sun.

Electra	17 Tau	03h 44.9m	+24° 07′	November
3.72 m	−1.56 M	B6 III		Taurus

Located within the Pleiades star cluster.

Alcyone	η Tauri	03h 47.5m	+24° 06′	November
2.85 m	−02.41 M	B7 III		Taurus

Alcyone is the brightest star in the Pleiades star cluster, with a luminosity of about 350 times that of the Sun.

Maia	20 Tauri	03h 45.8m	+24° 22′	November
3,87 m	−1.344 M	B8 III		Taurus

Yet another lovely blue star in the Pleiades cluster. This one has a luminosity of about eight times that of the Sun.

Eta Sagitai	ε Sgr	18h 24.2m	−34° 23′	June
1.79 m	−1.44 M	B9.5 III		Sagittarius

A brilliant orange star at a distance of 125 light years with a luminosity of 250 times the Sun's.

Nu Draconis^1	ν^1 Dra	17h 32.2m	+55° 11′	June
4.89 m	2.48 M	Am		Draco

A classic double-star system visible in binoculars and small telescopes. The stars are nearly identical in magnitude and stellar class and have a lovely white color.

Alhena	γ Gem	06h 37.7m	+16° 23′	December
1.93 m	−0.60 M	A0 IV		Gemini

The star is relatively close at about 58 light years, with a luminosity of 160 Suns.

Castor	α Gem	07h 34.6m	+31° 53′	January
1.43 m	0.94 M	A1 V		Gemini

Part of the famous multiple-star system and fainter brother to Pollux. The visual magnitude stated is the result of combining the magnitudes of the two brighter components of the system, 1.9 and 2.9.

Deneb	α Cyg	20h 41.3m	+45° 17′	August
1.25vm	−8.73[9] M	A2 I		Cygnus

The faintest star of the Summer Triangle (the others being Altair and Vega). A supergiant star with a definite pale blue color. The prototype of a class of pulsating variable stars.

Denebola	β Leo	11h 49.1m	+14° 34′	March
2.14vm	1.92 M	A3 V		Leo

Several companion stars are visible through a variety of instruments. The star has only recently been designated a variable.

Delta Leonis	δ Leo	11h 14.1m	+20° 31′	March
2.56 m	1.32 M	A4 V		Leo

Also called Zozma, this lies at a distance of 80 light years, with a luminosity of 50 Suns.

Ras Alhague	α Oph	17h 34.9m	+12° 34′	June
2.08 m	1.30 M	A5 III		Ophiucus

---

[9] This value is in question. The data are awaiting reassessment.

An interesting star for several reasons. It shows the same motions through space as several other stars in the so-called Ursa Major Group. It also shows interstellar absorption lines in its spectrum. Finally, measurements show an oscillation, or wobble, in its proper motion, which would indicate an unseen companion star.

2 Mon	HD 40536	05h 59.1m	−09° 33′	December
5.01 m	0.02 M	A6		Monoceros

The star lies at a distance of over 1900 light years, with a luminosity of 5000 Suns.

Alderamin	α Cep	21h 18.6m	+62° 35′	August
2.45 m	1.58 M	A7 IV		Cepheus

This is a rapidly rotating star, which results in the spectral lines becoming broad and less clear. It also has the dubious distinction of becoming the Pole Star in A. D. 7500.

Gamma Herculis	γ Her	16h 21.8m	+19° 09′	May
3.74 m	−0.15 M	A9 III		Hercules

An optical double system, lying at a distance of 144 light years and with a luminosity 46 times the Sun's.

Canopus	α Car	06h 23.9m	−52° 41′	December
−0.62 m	−5.53 M	F0 I		Carina

The second-brightest star in the sky. Its color is often reported as orange or yellow, as it is usually seen lying low down in the sky and is thus apt to be affected by the atmosphere. Its true color is white.

b Velorum	HD 74180	08h 40.6m	−46° 39′	January
3.84 m	−6.12 M	F3 I		Vela

This star is unremarkable except that its luminosity has been calculated to be that of 180,000 Suns!

Zubenelgenubi	α^1 Lib	14h 50.7m	−15° 60′	May
5.15 m	3.28 M	F4 IV		Libra

An easily resolvable double star, α^1 is also a spectroscopic binary. The colors are a nice faint yellow and pale blue.

Algenib	α Per	03h 24.3m	+49° 52′	November
1.79 m	−4.5 M	F5 I		Perseus

The star lies within Melotte 20, a loosely bound stellar association, also known as the Perseus OB-3, or Alpha Persei Association. About 75 stars with magnitudes down to 10 are contained within the group. All are stellar infants, only 50 million years old, lying 550 light years away. The metallic lines increase through the F class, especially the H and K lines of ionized calcium.

Polaris	α UMi	02h 31.8m	+89° 16′	October
1.97vm	−3.64 M	F7 I		Ursa Minor

An interesting and famous star, even though it is only the 49th brightest star in the sky. It is a Cepheid Variable Type II (the W Virginis class); it will be closest to the celestial pole in A. D. 2102, and it is a binary star (the companion reported as being pale blue).

β Vir	HD 102870	11h 50.7m	+01° 46′	March
3.59 m	3.40 M	F8 V		Virgo

A close star at 34 light years, only three times as luminous as the Sun.

Sadal Suud	β Aqr	21h 31.6m	−05° 34′	August
2.90 m	−3.47 M	G0 I		Aquarius

A giant star, and a close twin to α Aqr, it lies at a distance of 990 light years and is 5000 times more luminous than the Sun.

Sadal Melik	α Aqr	22h 05.8m	−00° 19′	August
2.95 m	−3.88 M	G2 I		Aquarius

Although it has the same spectral class and surface temperature as the Sun, α Aqr is a supergiant star, whereas the Sun is a main sequence star.

Ras Algethi	α^2 Her	17h 14.7m	+ 14° 23′	June
5.37 m	0.03 M	G5 III		Hercules

As stated later a beautiful double star with colors of ruddy orange and blue-green. The spectral class refers to the primary of α^2 Her, which is a spectroscopic double and thus visually inseparable with any telescope.

Algeiba	γ^2 Leo	10h 19.9m	+19° 50′	February
3.64 m	0.72 M	G7 III		Leo

A famous double; most observers report orange-yellowish colors, but some see the G7 star as greenish.

β LMi	HD 90537	10h 27.8m	+36° 42′	February
4.20 m	0.9 M	G8 III		Leo Minor

A constellation in which there is no star given the classification α, β LMi has the misfortune of not even being the brightest star in the constellation; that honor goes to 46 LMi.

β Cet	HD 4128	00h 43.6m	−17° 59′	October
2.04 m	−0.30 M	G9.5 III		Cetus

This star lies at a distance of 60 light years, with a luminosity of 42 Suns.

Gienah	ε Cyg	20h 46.2m	+33° 58′	August
2.48 m	0.76 M	K0 III		Cygnus

Marking the eastern arm of the Northern Cross, this star is a spectroscopic binary. In K-class stars, the metallic lines are becoming more prominent than the hydrogen lines.

ν^2 CMa	HD 47205	06h 36.7m	−19° 15′	December
3.95 m	2.46 M	K1 III		Canis Major

This star lies at a distance of 60 light years, with a luminosity of seven times that of the Sun.

Enif	ε Peg	21h 44.2m	+09° 52′	August
2.38vm	−4.19 M	K2 I		Pegasus

This star lies at a distance of 740 light years, with a luminosity of 7450 times that of the Sun. The two faint stars in the same field of view have been mistakenly classified as companions, but analysis has now shown them to be stars in the line of sight.

Almach	γ^1 And	02h 03.9m	+42° 20′	October
2.33 m	−2.86 M	K3 III		Andromeda

This is a famous binary star where the colors are gold and blue, although some observers see orange and greenish blue. Nevertheless, the fainter companion is hot enough to truly show a blue color. It is also a binary in its own right, but not observable with amateur instruments.

ζ^2 Sco	HD 152334	16h 54.6m	−42° 22′	June
3.62 m	0.3 M	K4 III		Scorpius

The brighter of the two stars in this naked-eye optical double-star system, the orange supergiant star contrasts nicely with its slightly fainter blue supergiant companion.

ν^1 Boö	HD 138481	15h 30.9m	+40° 50′	May
5.04 m	−2.10 M	K5 III		Boötes

The star lies at a distance of 385 light years and has a luminosity of 104 Suns (see also Aldebaran).

Mirach	β And	01h 09.7m	+35° 37′	October
2.07 m	−1.86 M	M0 III		Andromeda

In this stellar class, the bands of titanium oxide are strengthening. This red giant star is suspected of being slightly variable, like so many other stars of the same type. In the field of view is the galaxy NGC 404.

Antares	α Sco	16h 29.4m	−26° 26′	May
1.06vm	−5.28 M	M1 I		Scorpio

A giant star measured to be some 600 times the diameter of our Sun, it is a gloriously colored star of fiery red, contrasting nicely with its fainter green companion.

Scheat	β Peg	23h 03.8m	+28° 045	September
2.44vm	−1.49 M	M2 II		Pegasus

Marking the northwestern corner of the Square of Pegasus, this is a red irregular variable star. It is noted for having been one of the first stars to have its diameter measured by the technique of interferometry, at 0.021 arc seconds. Being variable, its size oscillates, to a maximum diameter of 160 Suns.

Eta Persei	η Per	02h 50.7m	+55° 54′	November
3.77 m	−4.28 M	M3 I		Perseus

A yellowish star in an easily resolved double-star system. The color contrasts nicely with its blue companion.

Gacrux	γ^A Crucis	12h 31.2m	−57° 07′	March
1.59 m	−0.56 M	M4 III		Crux

The top star of the Southern Cross, this is a giant star. γ^A and γ^B do not form a true binary, as they are apparently moving in different directions.

Ras Algethi	α^1 Her	17h 14.6m	+14° 23′	June
3.03vm	−2.32 M	M5 II		Hercules

A fine double-star system. The M5 semi-regular star is an orange supergiant, in contrast to its companion, a blue-green giant. However, it must be pointed out here that it can be resolved only with a telescope and not in binoculars, as the two stars are less than 5″ apart. The changes in brightness are attributed to actual physical changes in the star, as it increases and decreases in diameter.

Mira (at maximum)	o Cet	02h 19.3m	−02° 59′	October
2.00vm	−3.54 M	M5		Cetus

For full details on Mira, see the section on long period variables.

Mira (at minimum)	o Cet	02h 19.3m	−02° 59′	October
10vm	−0.5 M	M9		Cetus

For full details on Mira, see the section on long-period variables.

θ Apodis	HD 122250	14h 05.3m	−76° 48′	April
5.69vm	−0.67 M	M6.5 III		Apus

This is a semi-regular variable with a period of 119 days and a range of fifth to nearly eighth magnitude. The titanium bands are now at their strongest.

Thought Question 3.4

You live in an area that has considerable light pollution due to streetlights. You take several spectra of several different objects but notice two distinct spectral lines that always appear in your spectra. What lines do you think they could be, and their origin?

3.4. Redshift and Blueshift

Let us introduce a concept here that we will meet several times later in the book, especially when we deal with high-energy objects, and again when we discuss cosmology. This is the topic of redshift.

Basically, redshift is a phenomenon that occurs when an object, be it a star, nebulae, galaxy, or in fact anything, moves either towards or from an observer. The result is that the light from the moving object actually changes color. If it is moving from the observer, the light is reddened, hence *redshift*; if it is moving toward the observer, the light becomes bluer, hence *blueshift*.

In addition, the faster the object moves, in either direction, the greater the shift in color. So that objects, galaxies for example, that have a tremendous velocity with respect to us on Earth, become very red, visually, and so this effect has to be taken into account when measurements are made, in order to get a true measurement of the color (and other data) of the galaxy.

Note that the effect only occurs if the object is moving radially toward or from us. If it is moving across our line of sight, then no such effect occurs.

Where the effect really comes into its own is as a tool for measuring accurately the velocities of objects. What one does here is not look at the color but rather a specific spectral line, and measure how much it has shifted from its normal, non-moving wavelength. This may seem complicated, but in reality for bright objects it is a very easy procedure.

One obtains the spectra of, say, hydrogen, in the observatory (thus it is stationary with respect to the observer) and then measures the spectra of an object in space (moving with respect to the stationary observatory), which will hopefully have hydrogen in its spectra.[10]

Comparing the spectra will show a measurable shift in the wavelength of the hydrogen lines, and one usually uses Hα, the very important Balmer line we mentioned earlier. Finally, one uses a simple mathematical relationship that relates the amounts of shift to that of the velocity. By measuring the shift one can easily determine the velocity of an object. An example is shown in the following Math Box 3.2.

In many books the phenomena of redshift has been likened to the Doppler effect, where sound waves from moving objects change frequency. While this is superficially acceptable, it must be stressed that the actual cause of the redshift is not the same as that in the Doppler effect. The redshift under discussion is due to the subtle and elegant phenomenon of special relativity, and although the mathematics is not too complicated, it is detailed and would require significant background theory to fully appreciate it. Thus, we will not cover it here.

Finally, the description we have covered here for redshift concerns objects moving *through* space, that is to say objects that have a velocity far below that of the speed of light. When an object has a significant amount of velocity due to the expansion of space, an entirely new description, and new formula, must be used.

---

[10] In reality several spectra are taken in the observatory, not just that of hydrogen, in order to ensure that some spectral lines will be recognizable in the object's spectrum.

Math Box 3.2 Redshift

The redshift of an object is the difference between the observed wavelength of a spectral line and its rest wavelength.

$$\text{Redshift} = z = \frac{\lambda_{observed} - \lambda_{rest}}{\lambda_{rest}}$$

Example:

A neutron star has an observed Hα line at 662.9 nm. The rest wavelength of Hα is 656.3 nm.

Calculate the redshift of the neutron star and its velocity of recession.

$$z = \frac{662.9 - 656.3}{656.3} = 0.010$$

The redshift of the neutron star is 0.010.

For objects where $v \ll c$, and when z is much less than 1, its velocity can be calculated by

$$v = c \times z$$

where c = speed of light, 3.0×10^8 meters per second. Thus

$$v = c \times z = \left(3.0 \times 10^8\,\text{m/s}\right) \times 0.01 = 3000\ \text{km per second}$$

The neutron star's velocity of recession is 3000 km/s.

Thought Question 3.5

An astronomer measures the redshift of a stellar-looking object. It has a tremendous redshift of 7.54 which implies it lies at a comoving distance of 29.36 billion light years. Using this information can you suggest what type of object it is?

Problems

1. Determine the wavelength of light emitted when an electron makes a transition from level 4 to level 2.
2. What is this line commonly called?

3. A nearby galaxy has an observed Hα line at 675.5 nm. The rest wavelength of Hα is 656.3 nm. Calculate the redshift of the galaxy and its velocity of recession.
4. The highest redshift discovered is of galaxy GN-z11 with a value $z = 11.09$. Calculate its recessional velocity.
5. Do you think this is correct?

Chapter 4

The Hertzsprung-Russell Diagram

4.1. Introduction

Most of us are familiar with graphs, having seen, for example, ones that display height as a function of age, or temperature as a function of the time of year. So a similar approach was pursued for the characteristics of stars. The graph that is used universally is called the *Hertzsprung-Russell Diagram*. It is, without a doubt, one of the most important and useful diagrams in the whole of astronomy.

In 1911, the Danish astronomer Ejnar Hertszprung plotted the absolute magnitude of stars (a measure of their luminosities) against their colors (a measure of their temperature). Then, in 1913, the American astronomer Henry Norris Russell independently plotted spectral types (another way to measure temperature) against the absolute magnitude. They both realized that certain previously unsuspected patterns were emerging, and furthermore, an understanding of these patterns is of *crucial* importance to the study of stars. In recognition of the pioneering work of these astronomers, the graph is now known as the Hertszprung-Russell, or H-R, diagram. Figure 4.1 is a typical H-R diagram. Each dot on the diagram represents a star whose properties, such as spectral type and luminosity, have been determined. We will now explore the key features of the diagram.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical
Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_4

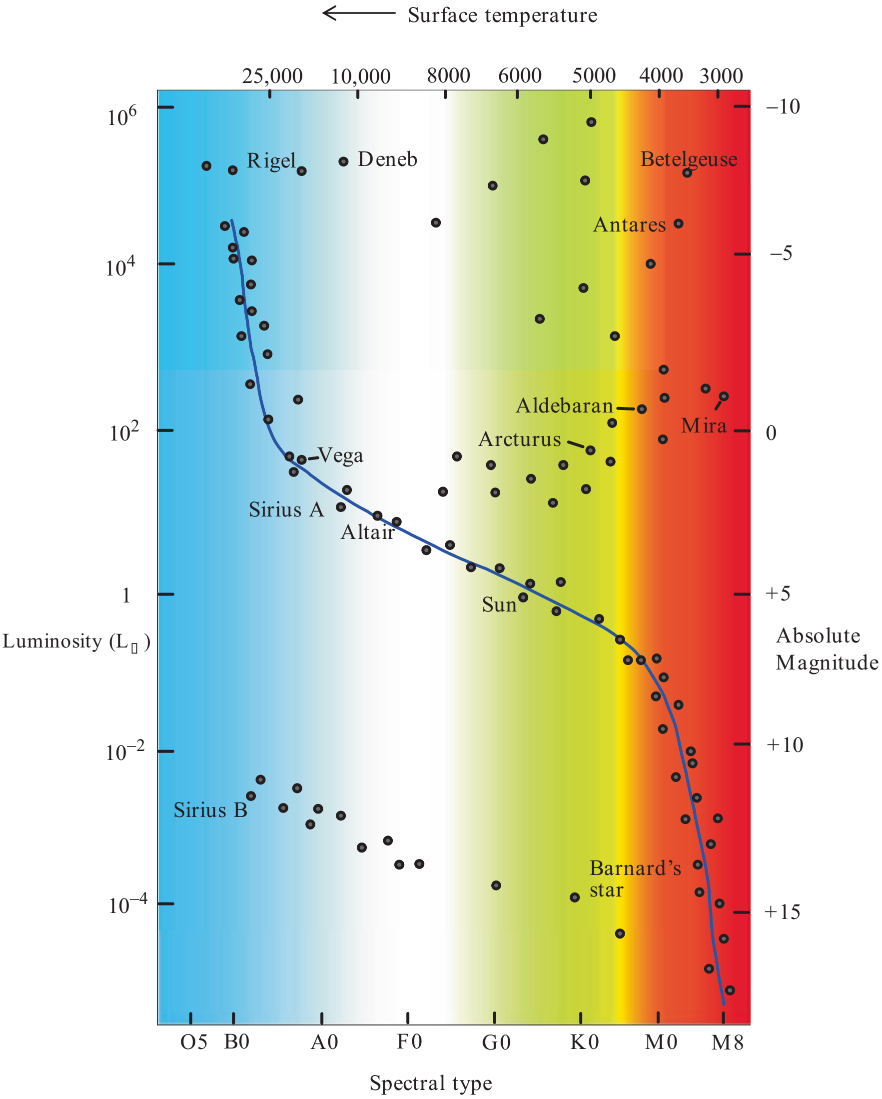


Fig. 4.1. The Hertzsprung-Russell diagram

Luminosity is plotted against spectral type for a selection of stars. Some of the brighter stars are shown. Each dot represents a star whose spectral type and luminosity have been determined. Note how the data are grouped in just a few regions, indicating a correlation. The main sequence is the continuous blue line. Surface temperature and absolute magnitude are also shown. In addition,

- The horizontal axis represents stellar temperature, or, equivalently, the spectral type.
- The temperature increases from right to left. This is because Hertzsprung and Russell originally based their diagram on the spectral sequence OBAFGKM,[1] where hot O-type stars are on the left and cool M-type stars are on the right.
- The vertical axis represents stellar luminosity and is measured in units of the Sun's luminosity, $L_\odot$.
- The luminosities cover a wide range, so the diagram makes use of the logarithmic scale, whereby each tick mark on the vertical axis represents luminosity ten squared times greater than the prior tick mark.
- Each dot on the H-R diagram represents the spectral type and luminosity of a single star. For example, the dot representing the Sun corresponds to the Sun's spectral type of G2 with a luminosity of $L_\odot = 1$.

Note that because luminosity increases upward on the diagram and surface temperature increases leftward, stars near the upper left corner are hot and luminous. Similarly, stars near the upper right are cool and luminous, while stars near the lower right are cool and dim. Finally, stars that are near the lower left corner are hot and dim.

4.2. The H-R Diagram and Stellar Radius

The H-R diagram can also directly provide important information about the radius of stars, because a star's luminosity depends on both its surface temperature and its surface area, or radius. You will recall that the surface temperature determines the amount of power emitted by the star *per unit area*. Thus, a higher temperature means a greater power output per unit area. So, if two stars have the same temperature, one may be more luminous than the other if it is larger in size. Stellar radii must perforce increase as we go from the high-temperature, low-luminosity corner on the lower left of the H-R diagram to the low-temperature, high-luminosity upper right-hand corner. This is shown in Fig. 4.2.

The first thing to notice on the H-R diagram is that the data points (or stars) are not scattered at random but appear to fall into distinct regions. This immediately implies that the surface temperature (or spectral type) and luminosity are related! The several groupings can be described thus.

The band that stretches diagonally across the H-R diagram is called the *main sequence*, and it represents about 90% of all the stars in the night sky.

---

[1] The new classes of stars, L, T and Y are not plotted, as these are only seen if using infrared detectors, and we are mainly concerned with stars in the visual range.

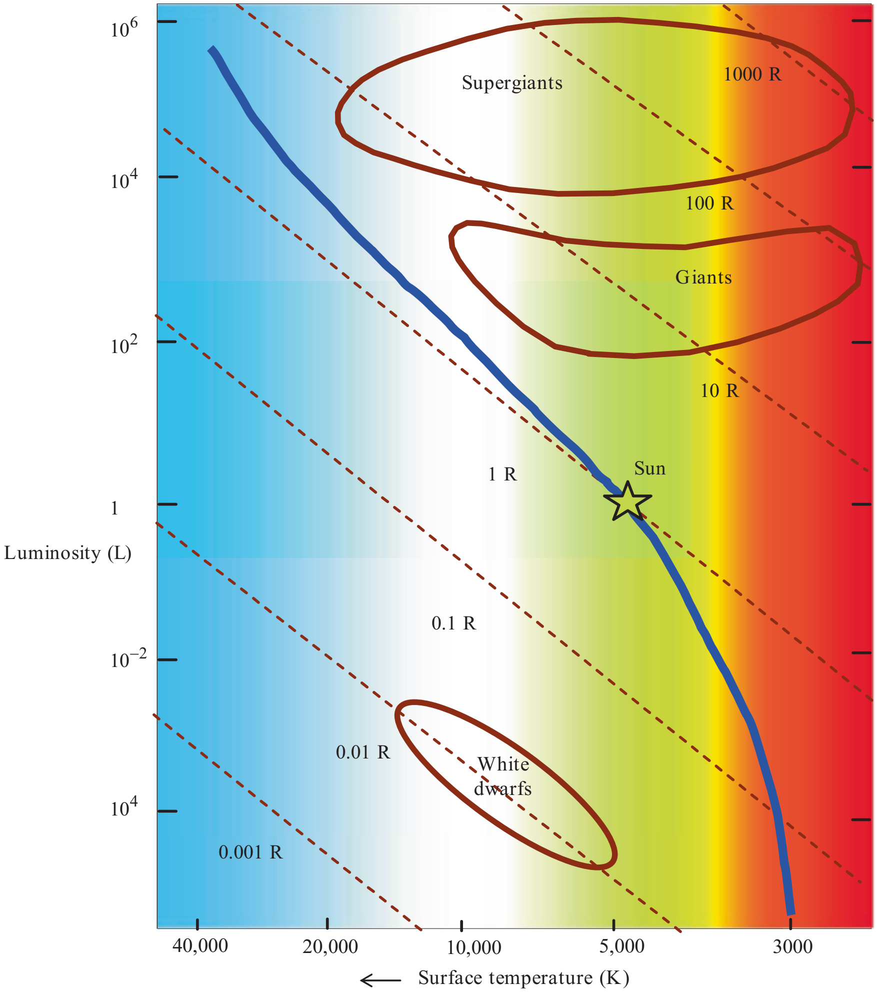


Fig. 4.2. Size of stars on an H-R diagram

It extends from the hot and luminous blue stars in the upper left corner to the cool dim red stars in the bottom right. Any star residing in this part of the H-R diagram is called a *main sequence star*. Note that the Sun is a main sequence star (spectral-type G2, absolute magnitude +4.8, luminosity 1 $L_\odot$). Later in the book we will discuss how the stars on the main sequence are undergoing hydrogen burning (thermonuclear fusion, which converts hydrogen to helium) in their cores.

The stars in the upper right are called *giants*. These stars are both cool and luminous. Recall that in an earlier section we discussed the Stefan-Boltzmann law, which told us that a cool star will radiate much less energy per surface area than a hot star. So for these stars to be as luminous as they appear, they must be immense, and so they are called *supergiants.* They can be anywhere from 10 to 100 times as big as the Sun. Figure 4.2 shows this, where stellar radii have been added to the H-R diagram. Most giant stars are about 100–1000 times more luminous than the Sun and have temperatures of about 3000–6000 K. Many of the cooler members of this class are reddish in color and have temperatures of 3000–4000 K. These are often referred to as *red giants*. Some examples of red giants are Arcturus in Boötes and Aldebaran in Taurus.

In the upper extreme right corner are a few stars that are even bigger than the giants. These are the supergiants, which have radii up to 1000 $R_\odot$. Giants and supergiants make up about 1% of all the stars in the night sky. Antares in Scorpius and Betelgeuse in Orion are two fine examples of supergiant stars. Although nuclear fusion is taking place in these stars, it is significantly different in both character and position than the reactions taking place in the stars on the main sequence.

The stars in the lower left of the H-R diagram are much smaller in radius and appear white in color. These are the *white dwarfs.* As you can see from the H-R diagram, they are hot stars with low luminosities; therefore, they must be small and hence the dwarf aspect to their name. They are faint stars, so they can only be seen with telescopes, and they are approximately the same size as Earth. There are no nuclear reactions occurring within white dwarfs; rather, they are the still-glowing remnants of stars. White dwarfs account for about 9% of all stars in the night sky.

The dashed diagonal lines indicate stars of different radii. At a given radius, the surface temperature increases (moving from right to left), and luminosity increases. Notice the main sequence and the Sun's place on it. A very average star.

4.3. The H-R Diagram and Stellar Luminosity

Thought Question 4.1

I have stated above that the Sun is an average star. Is this an accurate statement?

The temperature of a star determines which spectral lines are most prominent in its spectrum. So classifying a star by its spectral type is essentially the same as by its temperature. But a quick glance at an H-R diagram will reveal that stars can have similar temperatures but in fact very different luminosities.

Consider this example: a white dwarf star could have a temperature of 5700 K, but so could a main sequence star, a giant or a supergiant. It all depends on its luminosity. Therefore, by examining a star's spectral lines, one can determine to which category the star belongs. A general rule of thumb (for stars of spectral types B through F), is the more luminous the star, the narrower the lines of hydrogen. The theory behind the phenomenon is quite complex, but suffice to say that these measurable differences in spectra are due to differences in stars' atmospheres, where the absorption lines are produced. The density and pressure of the hot gas in the atmosphere affect the lines, and hydrogen in particular. If the pressure and density are high, the hydrogen atoms collide more frequently, and they interact with other atoms in the gas. The collisions cause the energy levels in the hydrogen atoms to shift, resulting in broadened hydrogen spectral lines.

In a giant luminous star, the atmosphere will have a very low pressure and density due to the star's mass being spread over such an enormous volume. Therefore, the atoms (and ions) are relatively far apart. This means that collisions between them are far less frequent, which produces narrow hydrogen lines. In a main sequence star, the atmosphere is denser than a giant or supergiant, with the collisions occurring more frequently, thereby producing somewhat broader hydrogen lines.

In an earlier section describing stellar classification, we saw that we can ascribe to a star a luminosity class. We can now use this to describe the region of the H-R diagram where a star of a particular luminosity will fall. This is shown in Fig. 4.3.

Knowing both a star's spectral type and its luminosity allows an astronomer to immediately know where on the main sequence it will lie. For instance, a G2 V star is a main-sequence star with a luminosity of about 1 $L_\odot$ and a surface temperature of about 5700 K. In a similar vein, Aldebaran is a K5 III star, which tells us immediately that it is a red giant star with a luminosity of about 375 $L_\odot$ with an accompanying temperature of about 4000 K.

Dividing the H-R diagram into luminosity classes allows distinctions to be made between giant and supergiant stars.

Thought Question 4.2

There are several bright red stars in the night sky visible to the naked eye. Are these Giants, Supergiants, or Main Sequence stars?

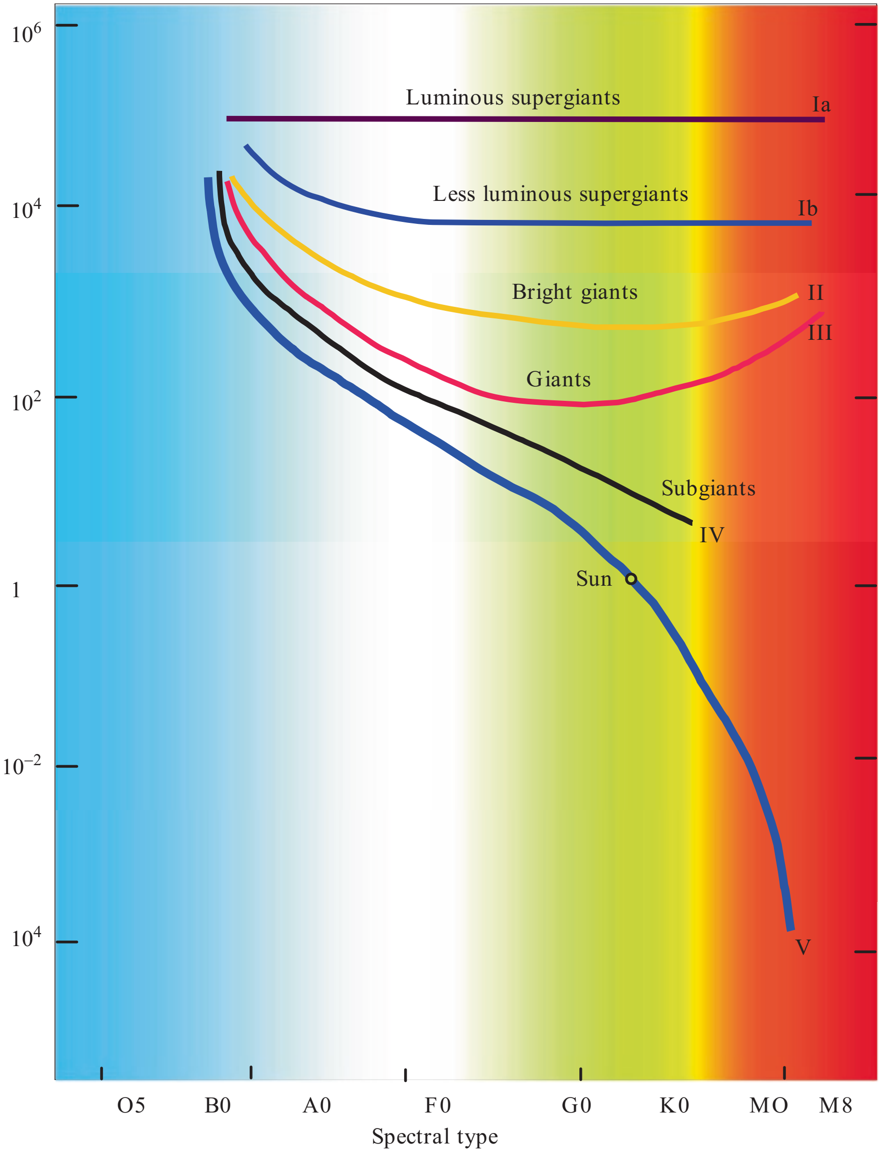


Fig. 4.3. Luminosity classes

4.4. The H-R Diagram and Stellar Mass

The most common trait of main sequence stars is that, just like our Sun, they undergo nuclear fusion in their cores to convert hydrogen to helium. And because most stars spend most of their lives doing this, it naturally follows that the majority of stars spend their time somewhere on the main sequence. But even a cursory glance at the H-R diagram will tell you that an enormous range of luminosities and temperatures are covered.

The question that arises is, why such a large range?

Astronomers have determined the masses of stars by using binary star systems, and they discovered that a star's mass increases as we move upward along the main sequence (Fig. 4.4). The O-type stars at the upper part of the diagram, that is, hot and luminous stars, can have masses as high as 100 times that of the Sun—100 $M_\odot$. At the other end of the main sequence, the cool and faint stars have masses as low as 0.1 times that of the Sun—0.1 $M_\odot$.[2] This orderly distribution of stellar masses along the main sequence tells us it is a star's *mass* that is the most important attribute of a hydrogen-burning star. The mass has a direct bearing on a star's luminosity because the weight of a star's outer layers will determine how fast the hydrogen-to-helium nuclear reaction will proceed in the core. A 10 $M_\odot$ star on the main sequence will be more than 1000 times more luminous than the Sun (i.e., 1000 $L_\odot$).

However, this mass-surface temperature relationship is just a little more subtle than the preceding paragraph would indicate. Generally, very luminous stars must be either very large or have a very high temperature, or even a combination of both. Those stars on the top left of the main sequence are some thousands of times more luminous than the Sun, but they are only about ten times larger than the Sun. Therefore, their surface temperatures must be significantly hotter than the Sun's to account for such high luminosities. Bearing this relationship in mind, we can now say that those main sequence stars that are more massive than the Sun must have correspondingly higher temperatures, while those with lower masses must have lower surface temperatures. Thus, you can now see why the main sequence on the H-R diagram goes diagonally from upper left to lower right.

Thought Question 4.3

Look at the Main Sequence. What changes the most, the mass or the luminosity?

The H-R diagram is one of the most fundamental tools in all of astronomy. We will use it throughout the remainder of the book in sections that deal with stars, as it provides a means for us to determine and follow many

---

[2] Over the past several years, astronomers have discovered that the low mass, faint M-type dwarf stars are far more numerous than any other type of star. We have just not been able to see them up until now.

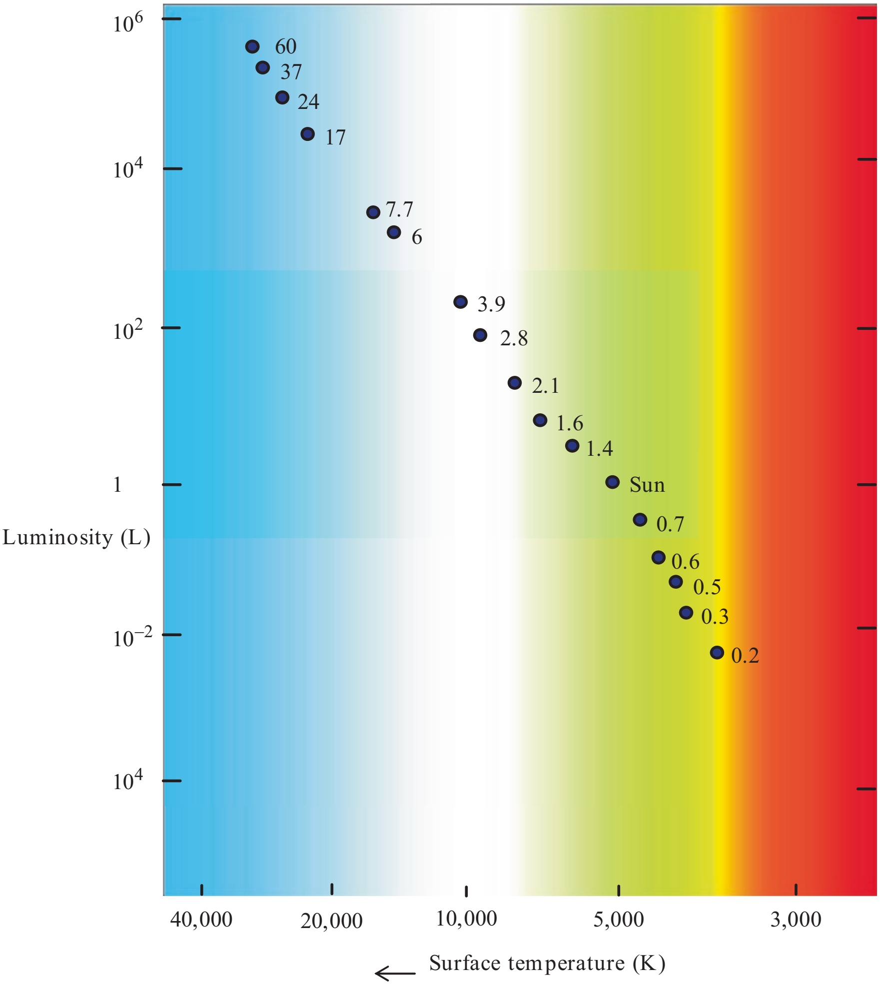


Fig. 4.4. Mass and the main sequence. Each filled-in circle is a main sequence star. The number is the star's mass in solar masses ($M_\odot$). As you move up the main sequence from lower right to upper left, the mass, luminosity and temperature increase

of the paths that stars take during their lives—from star birth all the way to star death.

Problems

1. Using Fig. 1, how many times greater is the luminosity of Aldebaran compared to the Sun?
2. Using Fig. 1, how many times smaller is the luminosity of Barnard's star compared to the Sun?

Chapter 5

The Interstellar Medium and Protostars

5.1. Introduction

When we look up into the night sky we see stars, and not much else. So we get the impression that between the stars, space is empty. There doesn't seem to be any sort of material that lies between one star and another. At the same time, we know intuitively that this cannot be true, for if space were empty, from what did stars form? This then leads us to the conclusion that perhaps space is not quite so empty, but filled with some sort of material that, to our eyes, is all but invisible yet is responsible for providing the source material for stars.

In fact, space is anything but empty; it is filled with gas and dust. This is known as the *interstellar medium* (ISM). The ISM is made up of gas (mainly hydrogen and Helium) and dust (which accounts for about 1% of the mass of the gas). The dust, not to be confused with dust on Earth, consists of other elements that are not hydrogen, such as carbon, silicon, etc., and their compounds, CO, HCN, etc. Think of it more like tiny grains of material, akin to sand, although far smaller.

The material that makes up the ISM is not spread evenly throughout space. There are regions that are dense and regions that are not so dense. Similarly, there are areas of the ISM that are hot, and other areas that are cooler. Thus, the two most important parameters concerning the ISM are the temperature and a quantity we call the number density, n. The latter is just

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_5

the number of particles per unit volume (per cubic meter), and it can be individual atoms, neutral, ionized, combined in molecules, or a mixture of all four. But because there is far more hydrogen in the ISM than anything else, we can say, to a good approximation, that the particle density, n, is just the number of hydrogen atoms per cubic meter, and this we call n_H.[1]

The important point to realize is the enormous range of temperatures and number densities that occur in the ISM. There can be as few as 100 particles per cubic meter ($n = 100\ \text{m}^{-3}$) to about 10^{17} per cubic meter ($n = 10^{17}\ \text{m}^{-3}$). Similarly, the temperature can be as low as 10 K and as high as a few million K. To get a feel for these ranges, look at Fig. 5.1 It shows the ranges of

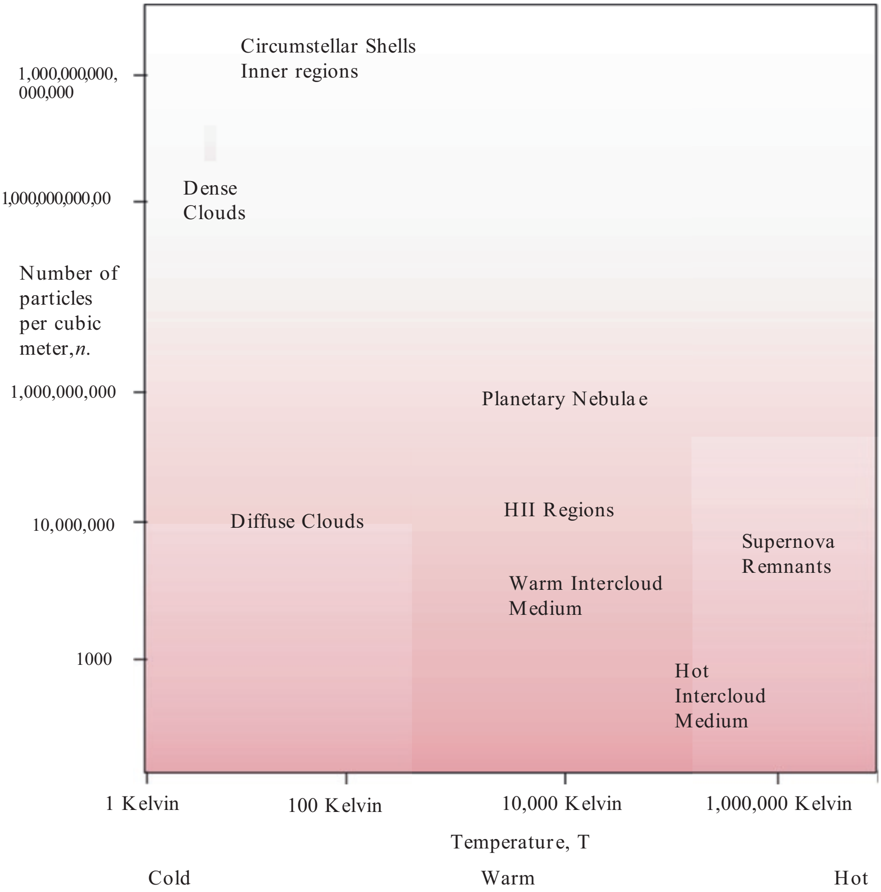


Fig. 5.1. Regions in the interstellar medium

---

[1] An important distinction is that n is not the same as the number of hydrogen atoms per cubic meter, nH. If the hydrogen is in a molecular form, H_2, then the number of separate particles is $n_H/2$.

temperature and number density, and the names we give to the correspondingly different regions in the ISM.

Let's look at this diagram in more detail; what we call the *intercloud medium*, whether hot or warm, actually accounts for most of the ISM. The interesting thing is that all other regions of the ISM are located within the intercloud medium. The regions are:

Hot intercloud medium—this is widespread and, although hot, of extremely low density and consisting of mainly ionized hydrogen. Fortunately for amateur astronomers, it does not obscure our view of space, as we can see through it. For similar reasons, the *warm intercloud medium* is also transparent.

All other regions on the diagram (and thus in the ISM) present a much more visual aspect and so are important to us as observers. They can divided into two groups: those regions in the ISM that are concerned with star formation, namely the *diffuse* and *dense clouds* and the *HII regions*,[2] and those that deal with star death—*planetary nebulae*, *supernova remnants*, and *circumstellar shells*.

We will discuss all of these regions in considerable detail in this and later chapters because they are objects of interest to the amateur astronomer…all, that is, except for the diffuse clouds, as these are transparent to visible light. There are, however, methods to allow observations of these clouds—radio astronomy to measure the hydrogen 21 cm line, microwave telescopes to measure the CO molecule, and infrared telescopes to measure the far infrared emission of the dust.

As an astronomer, you will have already observed the interstellar medium, but perhaps without realizing it. As previously mentioned, the ISM is composed of gas (mainly hydrogen[3]) and dust, so it is, from an observer's viewpoint, invisible; however, there are places in the galaxy where certain conditions tend to aggregate the material, and these denser-than-average regions are indeed visible to the amateur astronomer. We know them as *nebulae*.

5.2. Nebulae

Nebulae are actually disparate in nature, even though many of them have a rather similar appearance. They are associated with the areas of star formation,[4] cover several aspects of a star's life, and end with the process of star

---

[2] As mentioned earlier in this book, HII is pronounced "aitch 2."

[3] Recall that the ISM is made up of about 74% hydrogen (by mass), 25% helium and the rest metals.

[4] And in some cases, star death, namely supernovae remnants, covered later.

death. This section will just cover three main types of nebulae: emission, reflecting, and dark, all associated with the birth of a star. In addition, we are fortunate from an observing point of view, because these objects abound in the night sky, and some are spectacular objects indeed.

5.3. Emission Nebulae

These clouds of gas are associated with very hot O- and B-type stars, which produce immense amounts of ultraviolet radiation. They typically have masses of about 100–10,000 solar masses. This huge mass, however, is usually spread over a correspondingly large area (possibly a few light years across), so the actual density of the gas is extremely low (maybe only a few thousand hydrogen atoms per cubic centimeter). Usually, these very luminous stars are actually born within and from the material of the clouds, and so many emission nebulae are “stellar nurseries.” Radiation from the stars causes the gas (usually hydrogen) to undergo a process called *fluorescence*, and it is this that is responsible for the glow observed from the gas clouds.

The energy provided by ultraviolet radiation from the young and hot stars ionizes the hydrogen. In other words, energy—in this case, in the form of ultraviolet radiation—is absorbed by the atom and transferred to an electron that is sitting comfortably in what is called an energy level or orbital shell.[5] Having gained extra energy, the electron can leave the energy level it is in, and in some instances actually break free from the atom. When an atom loses an electron, the process is called *ionization*.

If electrons escape from their parent atoms, the hydrogen cloud will contain some hydrogen atoms without electrons—ionized hydrogen (also known as protons), and a corresponding number of free electrons. Eventually,[6] the electrons recombine with the atoms, but an electron can’t just settle down back to the state it was originally in before it absorbed the extra energy. It has to lose the extra energy that the ultraviolet imparted. For this to happen, the electron moves down the atomic energy levels until it reaches its original level, losing energy as it goes. In hydrogen (the most common gas in the nebula, you remember), an electron moving down from

---

[5] Our simple model of an atom has a central nucleus with electrons orbiting around it, somewhat like planets orbiting a sun. Electrons with a lot of energy are in the outer orbits, while electrons with less energy are closer to the nucleus. Not all orbits are allowed by quantum mechanics. To move up to higher energy levels, electrons need a very specific amount of energy; too much or too little, and an electron will not move.

[6] The time spent before recombining is very short—millionths of seconds—but also depends on the amount of radiation present and the density of the gas cloud.

the third energy level to the second emits a photon of light at 656.3 nm (we covered this in Chap. 3). This is the origin of the famous “hydrogen alpha line,” usually written as *H-alpha* and pronounced “aitch alpha.” It is a lovely red-pink color and is responsible for nearly all the pink and red glowing gas clouds seen in photographs of emission nebulae.[7]

When electrons move down from other energy levels within the atom, other specific wavelengths of light are emitted. For instance, when an electron moves from the second level to the first, it emits a photon in the ultraviolet part of the spectrum. This particular wavelength is called the *Lyman alpha line* of hydrogen, which is in the ultraviolet part of the spectrum.

It is this process of atoms absorbing radiation to ionize a gas, with electrons subsequently cascading down the energy levels of an atom, that is responsible for nearly all of the light we see from emission nebulae. If a gas cloud is particularly dense, the oxygen gas in it may be ionized, and the resulting recombination of the electron and atom produces the doubly ionized lines, at wavelengths of 495.9 nm and 500.7 nm.[8]

Emission nebulae are sometimes called HII regions, pronounced “aitch two.” This astrophysical term refers to hydrogen that has lost one electron by ionization. The term HI, or “aitch one,” refers to hydrogen that is unaffected by any radiation (that is, neutral hydrogen). The doubly ionized oxygen line mentioned above is termed OIII (“oh three”); the “doubly” means that *two* of the outermost electrons have been lost from the atom by ionization.[9]

The shape of an emission nebula is dependent on several factors: the amount of radiation available, the density of the gas cloud, and the amount of gas available for ionization. When there is a significant amount of radiation, coupled with a small and low-density cloud, then all of the cloud will likely be ionized, and thus the resulting HII region will be of an irregular shape—just the shape of the cloud itself. These nebulae are thus referred to as *matter bounded*.

If the cloud of gas is large and dense, however, then the radiation can penetrate only a certain distance before it is used up—that is, there is only a fixed amount of radiation available for ionization. In this case, the HII region will be a sphere,[10] often surrounded by the remaining gas cloud, which is not fluorescing. These nebulae are thus referred to as *radiation bounded*.

---

[7] Unfortunately, the red glow is usually too weak to be seen at the eyepiece.

[8] These lines are a rich blue-green color and, under good seeing conditions and with clean optics can be glimpsed in the Orion Nebula, M42.

[9] In some astrophysical contexts, such as in the center of quasars, conditions exist that can give rise to terms such as Fe23. The amount of radiation is so phenomenal that the atom of iron (Fe) has been ionized to such an extent, it has lost 22 of its electrons!

[10] This is often called the *Stromgren sphere*, named after the astronomer Bengt Stromgren, who did pioneering work on HII regions.

Many of the emission regions that are irregular in shape include M42 (the Orion Nebula), M8 (the Lagoon Nebula) and M17 in Sagittarius. Two of those exhibiting a circular shape, and thus are in fact spherical, are M20 (the Trifid Nebula) and NGC 2237 (the Rosette Nebula).

After a suitable period of time, usually several million years, the group of young O- and B-type stars located at the center of the nebulae will be producing so much radiation that they can in effect sweep away the residual gas and dust clouds that surround them. This produces a "bubble" of clear space surrounding the cluster of stars. Several emission regions show this. For example, NGC 6276 and M78 show the star cluster residing in a circular clear area within the larger emission nebula.

Thought Question 5.1

A fellow astronomer tells you that he has seen emission lines of Hydrogen around the star Aldebaran. Do you think he is correct?

Let's now look at a few examples of the brighter emission nebulae. Note, however, that from an observational viewpoint, many emission nebulae are faint and have a low surface brightness, making them not exactly difficult objects to observe, but rather featureless and indistinct (though in some instances the brighter nebulae do show several easily seen features). Therefore, clear nights and clean optics are a high priority.

There exist several classification schemes for emission nebula, but only one is used here. It is a measure of the visibility of the nebula as seen on the photographic plates from the Palomar Observatory Sky Survey (POSS). The photographic brightness is assigned a value from 1 through 6; those nebulae rated 1 are just barely detectable on the plate, while those quoted at a value of 6 are easily seen. In the context of this book, it is just the measure of the difficulty (or ease) of observation and is given the symbol •.

The size of an object is also given in arc seconds and is indicated by the symbol ⊕. Where a value of ⊕ is given as *x*|*y*, then the object is approximately *x* arc seconds long by *y* arc seconds wide. However, as larger apertures will show more and more of the nebula, what is given may not be what you observe.

The month given is the best time of the year to observe, when it will be at its highest in the sky.

The magnitude is also given where appropriate,[11] but on those occasions where a magnitude is quoted, treat it with caution!

---

[11] To determine the true visual magnitude of nebulae can be fraught with difficulties. Assigning a magnitude to a nebula (or any extended object, including galaxies) is performed thus: it is treated as if all the light from the object originates from a single point—the *integrated magnitude*. So an object that has, say, an integrated magnitude of 5, will not look as bright as a fifth magnitude star.

Gum 4	NGC 2359	07h 18.6m	−13° 12′	January
–	• 2 to 5	⊕ 9∣6′		

This nice object to observe, known as the Duck Nebula or, more recently, as Thor's Helmet, lies about 9° to the northeast of Sirius and is a bright emission nebula easily seen in telescopes of aperture 20 cm. It consists of two patches of nebulosity, with the northern patch being the larger and less dense. Using an OIII filter will greatly improve the appearance of the emission nebula, showing its delicate filamentary nature. The shape of the nebula is due to the presence of the central star, which is a Wolf-Rayet star, an extremely hot giant believed to be in a brief, pre-supernova stage of evolution.

Messier 20	NGC 6514	18h 02.3m	−23° 02′	June
6.3 m (9.0 m)	• 1 to 5	⊕ 20∣20′		

This emission nebula, or HII region, is also known as the Trifid Nebula and can be glimpsed as a small hazy patch of nebulosity. It is difficult to locate on warm summer evenings unless the skies are very transparent. In fact, determining its magnitude presents something of a problem due to the presence of several bright stars located within the nebula. With aperture of around 15 cm, the nebula is easy to see, along with its famous three dark lanes (classified as Barnard 85) that give it its name and which radiate outwards from the central object. An O8-type star is the power source for the nebula. The northern nebulosity is in fact a reflection nebula, and thus harder to observe.

Messier 8	NGC 6523	18h 03.8m	−24° 23′	June
5.8 m (4.6 m)	• 1 to 5	⊕ 46∣32′		

The Lagoon Nebula is undoubtedly one of the finest emission nebulae in the entire sky, and thought by many to be the premier emission nebula of the summer sky. Visible to the naked eye on summer evenings, binoculars will show a vast expanse of glowing green-blue gas split by a very prominent dark lane. Using light filters and telescopes of aperture 30 cm will show much intricate and delicate detail, including many dark bands. The Lagoon Nebula is located in the Sagittarius-Carina spiral arm of our galaxy, at a distance of around 5400 light years. A favorite target with those equipped with CCD cameras.

Messier 17	NGC 6618	18h 20.8m	−16° 11′	June
6.0 m	• 1 to 5	⊕ 40∣30′		

This is a magnificent object in binoculars and is perhaps a rival to the Orion Nebula, M42, in the summer sky, and is known as the Swan or Omega Nebula. Not often observed by amateurs, which is a pity, as it offers much.

With telescopes the detail of the nebula becomes apparent, and with the addition of a light filter it can in some instances surpass M42. Certainly, it has many more dark and light patches than its winter cousin, although it definitely needs an OIII filter for the regions to be fully appreciated. Unlike the Trifid Nebula, the stars responsible for nebulae are obscured within the cloud itself.

Messier 16	IC 4703	18h 18.8m	−13° 49′	June
6.4 m	• 1 to 5	⊕ 35|30′		

To clear up any confusion that may occur, Messier 16 is in fact a star cluster, NGC 6611, but associated with it is the famous Eagle Nebula, also known as the Star Queen Nebula. Oddly, even withstanding its fame, it is not often observed visually. Although it can be glimpsed in binoculars, the cluster can appear as a hazy patch with the naked eye, telescopic observation is needed to see any detail. The “Black Pillar” and associated nebulosity are difficult to see, even though they are portrayed in many beautiful photographs. (A prime example of astronomical imagery fooling the amateur into thinking that these justifiably impressive objects can easily be seen through a telescope.) Nevertheless it can be spotted by an astute observer under near-perfect conditions.

Caldwell 27	NGC 6888	20h 12.0m	+38° 21′	July
7.2 m	• 1 to 5	⊕ 18|103		

Although visible in binoculars, the Crescent Nebula will require a dark location and perhaps even a light filter to make its location that much easier. With good conditions, the emission nebula will live up to its name, having an oval shape with a gap in the ring on its southeastern side. The nebula is in fact a stellar wind bubble, and is the result of a fast-moving stellar wind from a Wolf-Rayet star that is sweeping up all the material that it had previously ejected during its red giant stage.

–	IC 5067–70	20h 50.8m	+44° 21′	August
8.0 m	• 1 to 5	⊕ 60|50′		

This nebula, also known as the Pelican Nebula, lies close to the North America Nebula and has been reported to be visible to the naked eye but is easily glimpsed in binoculars as a triangular, faint, and hazy patch of light. Remember, it can be seen best with averted vision and the use of light filters. The Pelican is an object of much study, as it consists both of star-forming regions and evolving gas clouds.

Caldwell 20	NGC 7000	20h 58.8m	+44° 20′	August
4.0 m	• 1 to 5	⊕ 120|100′		

A famous emission nebula, located just west of Deneb, the North America Nebula is magnificent in binoculars, melding as it does into the stunning star fields of Cygnus. Providing you know where and what to look for, the nebula is visible to the naked eye. With small- and large-aperture telescopes details within the nebula become visible, though several amateurs have reported that increasing aperture decreases the nebula's impact. The dark nebula lying between it and the Pelican Nebula is responsible for their characteristic shape. Until recently, Deneb was thought to be the star responsible for providing the energy to make the nebula glow, but recent research points to several unseen stars being the power sources.

–	IC 1396	21h 39.1m	+57° 30′	August
3.5 m	• 3 to 5	⊕ 170|40′		

One of the few emission nebulae visible to the naked eye (under perfect seeing conditions, of course!), and easily spotted in binoculars, this is an enormous patch of nebulosity, over 3°, spreading south of the orange star Mu (μ) Cephei. As usual, a telescope will lessen the impact of the nebula, but the use of filters will help to locate knots and patches of brighter nebulosity and dark dust lanes. The use of dark adaption and averted vision will enhance the observation of this giant emission nebula.

Caldwell 19	Sh2-125	21h 53.4m	+47° 16′	August
9.0 m (7.2 m)	• 3 to 5	⊕ 12|12′		

This nebula, also known as the Cocoon Nebula, is very difficult to find and observe. It has a low surface brightness and appears as nothing more than a hazy amorphous glow surrounding a couple of ninth-magnitude stars. The dark nebula Barnard 168 (which the Cocoon lies at the end of) is surprisingly easy to find, and thus can act as a pointer to the more elusive emission nebula. The whole area is a vast stellar nursery, and recent infrared research indicates the presence of many new and protostars within the nebula itself. Located with the nebula is the star cluster IC 5146.

Caldwell 11	NGC 7635	23h 20.7m	+61° 12′	September
10 m	• 1 to 5	⊕ 16|9′		

This is a very faint and strange nebula to observe, even in telescopes of medium and larger aperture. The "bubble" part, size 3 × 3 arc minutes, is actually located in a larger emission nebula of around 16 × 9 arc minutes. An eighth-magnitude star within the emission nebula and a nearby seventh-magnitude star hinder its detection, owing to their combined glare. Research suggests that a strong stellar wind from a star pushes material out and in turn ionizes the "bubble" also heating up a nearby molecular cloud. It really does bear a striking resemblance to a soap bubble. Located nearby is the classic star cluster M52, so anyone observing the cluster will also be observing the nebula! Just to confuse matters, it is also known as the Bubble Nebula and Sharpless 162.

–	NGC 604	01h 34.5m	+30° 48′	October
14 m	• 3 to 5	⊕ 60|35′		

Now for something very special. This may come as quite a surprise to many observers, but this is possibly the brightest emission nebula that can be glimpsed that is actually located in another galaxy. It resides in M33, in the constellation Triangulum. It appears as a faint hazy glow some 10′ northeast of M33's core. Owing to M33's low surface brightness (which often makes it a difficult object to find), the emission nebula may be visible while the galaxy isn't! It is estimated to be about 1000 times bigger than the Orion Nebula.

–	NGC 1499	04h 00.7m	+36° 37′	November
6 m	• 1 to 5	⊕ 160|50′		

This emission nebula presents a paradox. Some observers state that it can be glimpsed with the naked eye, others that binoculars are needed. The combined light from the emission nebula results in a magnitude of 6, but the surface brightness falls to around the 14th magnitude when observed through a telescope. Most observers agree, however, that the use of filters is necessary, especially from an urban location and when the seeing is not ideal. Clean optics is also a must to locate this nebula. Glimpsed as a faint patch in binoculars, with telescopes of aperture 20 cm, the emission nebula is seen to be nearly 3° long. Whatever optical instrument is used, it will remain faint and elusive. Also known as the California Nebula.

Messier 42	NGC 1976	05h 35.4m	–05° 23.5′	December
4.0 m	• 1 to 5	⊕ 85|60′		

Yes, here it is… the sky's most famous emission nebula. Visible to the naked eye as a barely resolved patch of light, the Orion Nebula shows detail from the smallest aperture upwards. It is really one of those objects where words cannot describe the view seen. In binoculars its pearly glow will show structure and detail, and in telescopes of aperture 10 cm the whole field will be filled. The entire nebulosity is glowing, owing to the light (and thus energy) provided by the famous Trapezium group of four stars located within it. What is also readily seen along with the glowing nebula are the dark, apparently empty and starless regions. These are still part of the huge complex of dust and gas, but are not glowing by the process of fluorescence. Instead they are vast clouds of obscuring dust. This emission nebula is one of the few that shows definite color, and many any observers report seeing a greenish glow, along with pale gray and blue. The British amateur astronomer Don Tinkler has this to say about M42: "The size of M42 always amazes me. Under dark skies it seems endless, with no edge or boundary. A wonderful nebula and a celestial showpiece." It has been reported that with very large apertures of 35 cm a pinkish glow can be seen. Located within the nebula are the famous Kleinmann-Low sources and the Becklin-Neugebauer object, which are believed to be dust-enshrouded young stars. The whole nebula complex is a vast stellar nursery, and in fact the stars AE Aurigae, 53 Arietis, and μ (Mu) Columbae, believed to have been formed in the nebula, are currently moving away from it at velocities greater than 100 km/s. Such stars are termed runaway stars. Messier 42 is at a distance of 1344 light years and is about 40 light years in diameter. Try to spend a long time observing this object. You will benefit from it, and many observers just let the nebula drift into the field of view. Truly wonderful!

CALDWELL 49	NGC 2237-39	06H 32.3M	+05° 03′	DECEMBER
–	• 1 to 5	⊕ 80|60′		

This giant emission nebula, known as the Rosette Nebula, has the dubious reputation of being very difficult to observe. But this is wrong. On clear nights it can be seen with binoculars. It is over 1° in diameter (~1.3°), and thus covers an area of sky four times larger than a full Moon! With a large aperture and light filters the complexity of the nebula becomes readily apparent, and under perfect seeing conditions dark dust lanes can be glimpsed. The brightest parts of the emission nebula have their own NGC numbers: 2237, 2238, 2239, (2244 is a cluster embedded within it) and 2246. It is a young nebula, perhaps only half a million years old, and star formation may still be occurring within it. Images show the central area containing the star cluster NGC 2244, along with the "empty" cavity caused by the hot young stars blowing the dust and gas away. Also known as the Rosette Molecular Complex (RMC).

NGC 2024	SH2-227	05H 41.7M	–01° 51.5′	DECEMBER
–	• 2 to 5	⊕ 30|30		

This observationally difficult object lies next to the famous star ζ (Zeta) Orionis (Altinak), which is unfortunate, as the glare from the star makes observation difficult. It can however be glimpsed in binoculars as an unevenly shaped hazy and faint patch to the east of the star, providing the star is placed out of the field of view. With large telescopes and filters the emission nebula is a striking object and has a shape reminiscent of a maple-leaf. Sometimes known as the Flame Nebula.

Caldwell 46	NGC 2261	06h 39.2m	+08° 44′	December
–	• 1 to 5	⊕ 3.5|1.5′ variable		

Caldwell 46, also known as Hubble's Variable Nebula, is easily seen in telescopes of 10 cm as a small, comet-like nebula, which can even be seen from the suburbs. Larger apertures just amplify what is seen with little detail visible. What we see is the result of a very young and hot star clearing away the debris from which it was formed. The star R Monocerotis (buried within the nebula and thus invisible to us) emits material from its polar regions, and we see the north polar emissions, with the southern emission blocked from view by an accretion disc. The variability of the nebula, reported in 1916 by Edwin Hubble, is due to a shadowing effect caused by clouds of dust drifting near the stars. It was also the first object to be officially photographed with the 200-in. Hale Telescope.

NGC 1555	Sh2-238	04h 21.8m	+19° 32′	November
–	• 2 to 5	⊕ 1|7′ variable		

This object was included despite its difficulty to locate and observe because it is so interesting. The famous but incredibly faint emission nebula known as Hind's Variable Nebula is located to the west of the famous star T Tauri, the prototype for a class of variable star. The nebula was much brighter in the past. With a large aperture, it will appear as a small faint hazy patch. When (and if!) located, it does bear higher magnification well. It may become brighter in the future, so it is well worth looking for in the hope that it makes a reappearance.

Sharpless 2-276		05h 31m	–04° 54′	December
6.5 m?	⊕ 600′			

In the author's own humble opinion, what we have here is possibly the greatest observing challenge to the naked-eye observer. Known as Barnard's Loop, and often mentioned in books but very rarely observed, this is a huge arcing loop of gas located to the east of the constellation Orion. It encloses both the sword and belt of Orion, and if it were a complete circle it would be about 10° in diameter. Barnard's Loop is currently believed to have originated from a supernova that occurred 2–3 million years ago. In addition, it may also have given rise to several runaway stars that include AE Aurigae, m (Mu) Columbae and 53 Arietis. It continues to glow due to a group of hot young stars in the Orion OB1 Association. Observationally, the eastern part of the loop is well defined, but the western part is exceedingly difficult to locate and likely has never been seen visually, only being observed by the use of photography or using a CCD. Impossible to see through a telescope, recent rumors have emerged that it has been glimpsed by a select few, by using either an OIII filter or an ultra-high-contrast filter. Needless to say, perfect conditions and very dark skies will greatly heighten the chances of it being seen.

Thought Question 5.2

What type of spectra would you expect from an emission nebula?

5.4. Dark Nebulae

Dark nebulae (also known as the dense clouds, mentioned earlier in the chapter) by nature differ from all others in one major respect—they do not shine. In fact, when you observe them, you are actually not seeing them by any light-emitting process but rather by their light-blocking ability. They are vast clouds of gas molecules, such as H_2, HCN, OH, CO and CS, as well as dust grains. These grains, however, bear no resemblance to the dust we see on Earth. They are microscopic in size, believed to be in the region of 20–30 nm. However, ice (either water ice—H_2O—or ammonia ice—NH_3) may condense on them, forming a "mantle," which then increases their size up to 300 nm. Dust grains are shaped like long spindles, and in some cases, they rotate. The actual composition of the grains is a topic of vigorous debate but is believed to be composed, in various unknown amounts, of carbon in the form of graphite, along with silicon carbides and silicates of magnesium and aluminum.

The formation of the dust grains is thought to take place in the outer regions of stars—in particular, the cool supergiants and the R Corona Borealis-type stars. Dense molecular clouds are thus believed to be another possible star formation site. The temperature of the grains is thought to be about 10–100 K, which is cool enough to allow the formation of molecules. In a typical dark nebula, there may be anywhere from 10^4 to 10^9 particles made up of atoms, various molecules and dust grains.

Due to their vast size, the nebulae appear dark and so are very effective in scattering all of the light, with the result that hardly any reaches the naked eye. The process of scattering the light is so effective, for instance, that visible light emitted from the center of our galaxy is nearly 100% extinguished by the dust clouds between its center and us. This is why the appearance of the central region in visible light is still a mystery. The scattering and absorption of light is known as *extinction*.

Don't be confused by thinking that these clouds of dust grains are very dense objects. They are not. Most of the material in the cloud is molecular hydrogen (along with carbon monoxide, which is responsible for their radio emission), and the resulting density is low. There is also some evidence to suggest that the dust grains present in the clouds have different properties from those in the interstellar medium.

Many dark nebulae are actually interacting with their environs, as witnessed by the spectacular images taken by the Hubble Space Telescope of M16 in Serpens. The images show dust clouds containing dense regions, or globules, resisting the radiation pressure from close, hot young stars, with the result that many of the globules are trailing long tails of material. The area near the Horsehead Nebula in Orion is also famous for its image of the radiation from the supergiant stars of Orion's Belt impacting on the dark clouds to either side of the Horsehead, with the result that material is ionized and streaming from the cloud's surface.

Most of the dark clouds listed herein have vastly different shapes, and this is due to several reasons:

- The cloud could have originally been spherical in shape, and thus present a circular image to us, but hot stars in its environment will have disrupted this by radiation pressure and stellar winds.
- Shock fronts from nearby supernovae can also have an effect.
- Gravitation from other clouds, stars and even that of the Milky Way itself can cause these effects.
- It is also thought that magnetic fields may have some limited effect.
- Many of the dark clouds are part of a much larger star forming region, thus new stars will themselves influence and alter cloud shape.

Let's now look at a few examples of dark nebulae.

The "opacity" of a dark nebula is a measure of how opaque the cloud is to light, and thus how dark it will appear. There is a rough classification system that can be used; a value of 1 for a dark nebulae indicates that it only very slightly attenuates the starlight from the background Milky Way; conversely, a value of 6 means that the cloud is nearly black, and it is given the symbol ★.

Observing dark nebulae can be a very frustrating pastime. The best advice we can offer is to always use the lowest possible magnification. This will enhance the contrast between the dark nebulae and the background star field. If a high magnification is used, the contrast will be lost, and you will only see the area surrounding the dark nebula, not the nebula itself. Dark skies are a must with these objects, as even a hint of light pollution makes their detection an impossible task.

Barnard 228	–	15h 45.5m	–34° 24′	May
⋆6	⊕ 240⇆20′			

A long band of dark nebula, easily spotted in binoculars lying halfway between Psi (ψ) and Chi (χ) Lupi. Best seen in low-power, large-aperture binoculars, as it stands out well against the rich background star field.

Barnard 59, 65-7	LDN 1773	17h 21.0m	–27° 23′	June
⋆6	⊕ 300⇆60′			

This large dark nebula, also known as the Pipe Nebula (Stem) and Lynds Nebula 1773, is visible to the naked eye because it stands out against a star-studded field and is best viewed with lower-power binoculars. With the unaided eye, it appears as a straight line, but under magnification its many variations can be glimpsed.

Barnard 78	LDN 42	17h 33.0m	–25° 35′	June
⋆5	⊕ 200⇆150′			

This is part of the same dark nebula as above. Known as the Pipe Nebula (Bowl), the bowl appears as a jagged formation, covering over 9°. The whole region is studded with dark nebula and is thought to be part of the same complex as that which encompasses Rho (ρ) Ophiuchi and Antares, which are more than 700 light years from it.

Barnard 86	LDN 93	18h 03.0m	–27° 53′	June
⋆5	⊕ 6′			

Barnard 86 is located within the Great Sagittarius Star Cloud. It is a near-perfect example of a dark nebula, appearing as a completely opaque blot against the background stars. Also known as the Ink Spot.

Barnard 87, 65-7	LDN 1771	18h 04.3m	–32° 30′	June
⋆4	⊕ 12′			

Not a distinct nebula, but it stands out because of its location within a stunning background of stars. Visible in binoculars as a small, circular dark patch, it is best seen in a small telescope of about 10–15 cm. Also known as the Parrot Nebula.

Barnard 103	LDN 497	18 h 39.4 m	–06° 41′	July
⋆6	⊕ 40|15′			

Easily seen at the northeast edge of the famous Scutum Star Cloud. This is a curved dark line that can be glimpsed in binoculars, but best seen at apertures of around 10–15 cm.

Barnard 110-1 18h 50.1m −04° 48′ July
⋆6 ⊕ 11′

The contrast between the background star clouds and the darkness of the nebula is immediately seen as a complex of dark nebulae that can be viewed in binoculars.

Barnard 142-3 19h 39.7m +10° 31′ July
⋆6 ⊕ 45′

This is an easily seen pair of dark nebulae, visible in binoculars. They appear as a cloud with two “horns” extending towards the west. The nebulae contrast very easily with the background Milky Way and so are fine objects. With a rich field telescope and large binoculars, the dark nebulae actually appear to be floating against the star field.

Barnard 145 20h 02.8m +37° 40′ July
⋆4 ⊕ 35|8′

Visible in binoculars, this is a triangular dust cloud that stands out well against the impressive star field. As it is not completely opaque to starlight, several faint stars can be seen shining through it.

Barnard 343 20h 13.5m +40° 16′ July
⋆5 ⊕ 13|6′

Easily seen as a “hole” in the background Milky Way, this is an oval dark nebula, which although glimpsed in binoculars is at its best in telescopes.

Lynds 906 20h 40.0m +42° 00′ August
⋆5 ⊕ – –

This is probably the largest dark nebula of the northern sky. It is an immense region, easily visible on clear moonless nights just south of Deneb and lying just at the northern boundary of the Great Rift, a collection of several dark nebulae that bisects the Milky Way. The rift is of course part of a spiral arm of the galaxy. To get an idea of what it would look like from a view outside the Milky Way, check out photographs of other galaxies such as NGC 891 in Andromeda. Also known as the Northern Coalsack.

Barnard 352 20h 57.1m +45° 54′ August
⋆5 ⊕ 20⇆10′

Visible in binoculars as a well-defined triangular dark nebula, this is part of the much more famous North America Nebula; this dark part is located to the north.

Barnard 33 05 h 40.9 m −02° 28′ December
⋆4 ⊕ 6⇆4′

Often photographed but rarely observed, this famous nebula, also known as the Horsehead Nebula, is very difficult to see visually in any telescope. It is a small dark nebula that is seen in silhouette against the dim glow of the emission nebula IC 434. Both are very faint and require perfect seeing conditions. Such is the elusiveness of this object that even telescopes of 40 cm are not guaranteed to pick it up.

Thought Question 5.3

Where would you look for new stars, and with what type of telescope?

5.5. Reflection Nebulae

The final classification of nebulae is *reflection nebulae*. As the name suggests, these nebulae shine by the light reflected from stars within them, or from nearby stars. Like the emission nebula, these vast clouds consist of both gas and dust, but in this case, the concentration of dust is far less than that found in emission nebulae. One of the characteristics of particles, or grains, that are so small (in proportion to the wavelength of light) is their property of selectively scattering light of a particular wavelength. If a beam of white light shines upon a cloud containing the grains, the blue light is scattered in all directions, a phenomenon similar to that seen in Earth's sky[12] (hence its blue color). This is one reason that reflection nebulae appear so blue on photographs; it is just the blue wavelengths of the light from (usually) hot blue stars nearby.

To be scientifically accurate, the nebula should be called scattering nebulae instead of reflection nebulae, but the name has stuck. An interesting

---

[12] Note that scattering of water molecules, and not dust, is responsible for the blue sky on Earth.

property of the scattered light is that the scattering process itself polarizes the light, useful in the studies of grain composition and structure.

But that's not all. If a star that lies behind a dust cloud is observed, some of its blue light is removed by the process discussed above, and an effect known as *interstellar reddening* occurs, which makes the light from the star appear redder than it actually is. This leads to a further phenomenon associated with dust grains called *interstellar extinction*, which should be mentioned because it affects *all* observations. Astronomers noticed that the light from distant star clusters was fainter than expected, and this was due to dust lying between the cluster and us. This in fact makes all objects fainter than they actually are and leads to an underestimation of their luminosity and an overestimation of their distance. Thus, interstellar extinction must be taken into account when making measurements.

Several reflection nebulae reside within the same gas clouds as emission nebulae. The Trifid Nebula is a perfect example. The inner parts of the nebula are glowing with a telltale pink color, indicative of the ionization process responsible for the emission, whereas further out from the center, the edge material is definitely blue, signposting the scattering nature of the nebula.

Visually, reflection nebulae are very faint objects with a low surface brightness, so they are not easy targets. Most require large-aperture telescopes with moderate magnification to be seen, but a few are visible in binoculars and small telescopes, and the brightest are the ones listed below. Note that excellent seeing conditions and very dark skies are required.

Caldwell 4	NGC 7023	21h 01.6m	+68° 10′	August
–	• 1 to 5	⊕ 18|18′		

Though small, this is a very nice, easy to observe reflection nebula. It has a star of seventh magnitude at its center, which can aide or hinder observation, depending on your inclination. However, what makes the reflection nebula easy to detect is its location; it is surrounded by a larger area of dark nebulosity, probably part of the same nebula complex. The contrast between the background stars, the dark nebula and the reflection nebula makes for a very interesting region.

NGC 1333	LBN 741[13]	03h 29.3m	+31° 25′	November
–	• 3 to 5	⊕ 9|7′		

This is a nice, easily seen reflection nebula and appears as an elongated hazy patch. Larger aperture telescopes will show some detail along with two fainter dark nebulae, Barnard 1 and 2, lying toward that north and south of the reflection nebula.

---

[13] This signifies it is the 741st object in the *Lynd's Catalogue of Bright Nebulae*.

MEROPE NEBULA	NGC 1435	03H 46.10M	+23° 46′	NOVEMBER
–	• 2 to 5	⊕ 30|30′		

This faint patch of reflection nebula is located within the most famous star cluster in the sky—the Pleiades. The nebula surrounds the star Merope; one of the brighter members of the cluster, and under perfect conditions, it can be glimpsed with binoculars. It has even been seen with a finder telescope, that, admittedly, was fitted with a deep sky filter. It is a comet-shaped cloud described by W. Tempel in 1859 as resembling "a breath on a mirror." Several other members of the cluster are also enshrouded by nebulosity, but these require exceptionally clear nights, and, incidentally, clean optics, as even the slightest smear on, say, a pair of binoculars, will reduce the chances to nil. There is considerable debate ongoing about the visibility of this nebula, as seen with the naked eye! Several observers have reported that it is visible under perfect conditions. Also known as Tempel's Nebula.

CALDWELL 31	IC 405	05H 16.2M	+34° 16′	DECEMBER
–	• 2 to 5	⊕ 37|19′		

This is a very difficult nebula to observe. It is actually several nebulae, including IC 405, 410 and 417, plus the variable star AE Aurigae. Using narrowband filters are justified with this reflection nebula, as they will highlight the various components. Can be thought of as a most definitive a challenge for the observer. Also known as the Flaming Star Nebula.

NGC 1999	LBN 979[14]	05H 36.5M	–06° 43′	DECEMBER
–	• 1 to 5	⊕ 2|2′		

In telescopes of aperture 20 cm, this small but bright reflection nebula resembles a planetary nebula, and even has a 10.5 magnitude star, V380 Orionis, in its central region. The nebula lies about 1° south of M42.

MESSIER 78	NGC 2068	05H 45.4 M	00° 05′	DECEMBER
8.0 m	• 1 to 5	⊕ 8|6′		

This is a bright but small reflection nebula that can be glimpsed in binoculars. Some observers say it resembles a fan shape, whereas others liken it to a comet. There are two 10th-magnitude stars located within the nebula that can give the false impression of two cometary nuclei. With a large-aperture telescope and high magnification, some very faint detail can be glimpsed along the eastern edge of the nebulosity, but excellent seeing will be needed. This is part of a group of reflection nebulae with NGC 2064 about 7′ to the southwest and NGC 2067 6′ to the west-northwest.

---

[14] This signifies it is the 979th object in the *Lynd's Catalogue of Bright Nebulae.*

Thought Question 5.4

What is the most obvious observational aspect of a Reflection Nebula?

5.6. Molecular Clouds

We have seen that interstellar space is filled with gas and dust, and that in certain locations concentrations of this material give rise to nebulae. But the location of these nebulae are not, as one might expect, entirely random. The areas that give rise to star formation are called *molecular clouds*. These clouds are cold, perhaps only a few degrees above absolute zero, and occupy enormous regions of space.

Due to the conditions within them, molecular clouds allow the formation of several molecules [e.g., carbon monoxide (CO), water (H_2O), and hydrogen molecules (H_2)[15]]. Although the most abundant molecule in a cloud, molecular hydrogen is very difficult to observe because of the low temperature. On the other hand, CO can be detected when certain portions of the cloud are 10–30 K above absolute zero. It is these molecules that allowed molecular clouds to be discovered by two radio astronomers—Philip Solomons and Nicholas Scoville—who, in 1974, found traces of carbon monoxide molecules in the Milky Way.

Molecular clouds are truly gigantic and contain vast amounts of hydrogen. They can range from 10^5 to 2×10^6 solar masses, and have diameters anywhere from 12 to 120 pc, or about 40 to 390 light years. The total mass of molecular clouds in our galaxy is thought to be about 5 billion solar masses.[16] But even though these molecular clouds are so vast, do not be fooled into thinking that we are talking about something that resembles, in structure, conditions similar to a foggy day, with hydrogen and dust being so dense that you can't see your hand in front of your own eyes. If we could go inside one of these clouds, there would be about 200 or 300 hydrogen molecules per cubic centimeter. This is not a lot, even though it is several thousand times greater than the average density of matter in our Galaxy. Even more staggering, it is 10^{17} times less dense than the air we breathe.

An example of all these types of nebulae can be seen in the Figs. 5.2 and 5.3, at the end of the chapter.

---

[15] Other molecules such as ammonia (NH_3) and alcohol (CH_3OH!!!) have also been detected.

[16] In areas where the average density exceeds, say, a million solar masses, clouds referred to as giant molecular clouds can form.

Fig. 5.2. Emission Nebula, Messier 42 – The Orion Nebula, and above it is NGC 1875, a reflection Nebula. You can see dark nebula embedded within both

Astronomers have deduced that molecular clouds and CO emission are intimately linked, and by looking at the areas in our galaxy where CO emission originates, we are in fact looking at those areas where star formation is taking place. Because the molecular clouds are, by comparison with the rest

Fig. 5.3. The Eagle Nebula, Messier 16

of the ISM, heavy and dense, they tend to settle toward the central layers of the Milky Way. This has produced a phenomenon we have all seen—the dark bands running through the Milky Way. Surprisingly, it was found that the molecular clouds in which star formation occurs outline the spiral arms of the Galaxy and lie about 1000 pc apart, strung out along the arms rather like pearls on a necklace.[17] However, spiral arms of galaxies are not the only place where star formation can occur. There are several other mechanisms that can give rise to stars, as we shall see in the chapter on stars.

Thought Question 5.5

Imagine the Earth was inside a Molecular cloud. What would the night sky look like to the naked eye?

---

[17] Molecular clouds can be found outside of spiral arms, but current ideas suggest that the spiral arms are regions where matter is concentrated, due to gravitational forces. The molecular clouds pass through the arms and are 'squeezed.' This dense region then gives rise to star-forming regions.

5.7. Protostars

We have included the topic of protostars in this chapter and not the following because we are still discussing large, diffuse clouds of gas and dust, although briefly, before they are turned into stars proper. So let's begin by looking at the mechanisms by which stars are believed to form.

As we said, space is full of gas and dust, and local concentrations of this material give rise to nebulae. But how do stars form in these regions? It may seem obvious in hindsight that a star will form in those clouds where the gas and dust are particularly dense and thus will allow gravity to attract the particles. An additional factor that will assist in formation is a very low temperature of the cloud. A cold cloud means that the (thermal) pressure of the interstellar medium is low. A cold temperature is not only helpful but in fact a prerequisite, as if a cloud has a high (thermal) pressure, it will tend to overcome any gravitational collapse. It is a delicate balancing act between gravity and pressure, whereby if gravity dominates, stars form. In fact, as we shall see in the next chapter on star birth, an outside agency is thought to be necessary to initiate star birth—a trigger mechanism!

From our discussion earlier in the book, you should have realized by now that there is one place where conditions like those just mentioned arise: the dark nebulae located within the molecular clouds. As a cloud contracts, pressure and gravity permitting, the dust and gas cloud becomes very opaque and is the precursor region to star formation. These regions are often called *Barnard objects*, after the astronomer who first cataloged them, Edward Barnard.[18] There are also even smaller objects sometimes located within a Barnard object. These resemble small, spherical dark blobs of matter and are referred to as *Bok globules*, named after astronomer Bart Bok. It may help you to think of a Bok globule as a Barnard object but with its outer layers, which are the less-dense regions, dispersed.

Radio measurements of Bok globules indicate that their internal temperature is a very low 10 K, and their density, although only about 100–20,000 particles (dust grains, gas atoms and molecules) per cubic centimeter, is considerably greater than that found in the ISM. The size of these objects can vary considerably; there are no standard sizes, but on average, a Bok globule is about 1 pc in diameter, with anywhere from 1 to 1000 $M_\odot$. The larger Barnard object, on the other hand, can have a mass of about 10,000 $M_\odot$, with a diameter of about 10 pc. As you can imagine, the sizes of these objects vary greatly and are determined by the local conditions in the ISM.

Now, if conditions permit, the densest areas within these objects and globules will further contract under gravitational attraction. A consequence of this contraction is a heating up of the blob of material; however, the cloud can radiate this thermal energy away, and in doing so prevent the pressure from building up high enough to resist the contraction. During the early phase of collapse, the temperature remains below 100 K, and the thermal energy is transported from the warmer interior to the exterior of the cloud by convection, causing the cloud to glow in infrared radiation. This ongoing collapse has the effect of increasing the cloud's density, but this makes it difficult for the radiation to escape from the object. Consequently, the central regions of the cloud become opaque, which traps nearly all of the thermal energy produced by the gravitational collapse.

Trapping the energy results in a dramatic increase in both pressure and temperature. The ever-increasing pressure fights back against the overpowering crush of gravity, and the now-denser fragment of cloud becomes a *protostar*—the seed from which a star is born. At this stage, the protostar may resemble a star, but it is not really a star, as no nuclear reactions occur in its core.

The time taken for the above scenario to occur can be extremely short, in an astronomical sense—maybe of the order of a few thousand years. The protostar is still quite large. For example, after, say, 1000 years, a protostar of 1 $M_\odot$ can be 20 times larger than the Sun's radius, $R_\odot$, and be about 100 times as luminous, 100 $L_\odot$.

5.8. The Jeans Criterion

You might think from the previous sections that star formation is a pretty straightforward process, and that if there is enough material (i.e., gas and dust) and a long enough period of time, the only possible outcome is the formation of a star. You would be wrong!

Remember that an interstellar cloud, however large (or small), performs a delicate balancing act between the gravitational attraction from all the cloud's particles, which is trying to collapse the cloud, and the thermal energy (think of it as heat of the cloud), which is trying to resist this collapse. If one is more dominant than the other, either a star will form, or it won't.

The question to ask yourself is "what decides whether gravity wins?" This is where the *Jeans Criterion*[19] comes into play. In a cloud with a spe-

---

[18] See the section on dark nebulae for observable examples of Barnard objects.

[19] James Jeans (1877–1946) was a British astronomer, and was the first person to mathematically describe the necessary conditions for star collapse.

cific density, temperature, and mass, these criteria describe the smallest-sized cloud and its minimum mass where gravity could overcome the thermal pressure and so result in collapse. As you can imagine, some quite involved equations are used; however, these are not too difficult, so we can make approximations to them.

The critical mass of the cloud is known as the Jeans Mass, M_J, and the critical size, the Jeans Length R_J. The Jeans Mass is the mass of the cloud whose radius is the Jeans Length. See Math Box 5.1.

From the above description, you can see that there are a few conditions that make cloud collapse more likely: the cloud has a very low temperature (the cooler the cloud, the better the chances of collapse), and the cloud is denser (a dense cloud has a better chance of collapse than one that is very thinly spread out). Thus, the dark, dense clouds discussed earlier would be ideal locations for cloud collapse. Indeed, in the darkest, densest clouds, only a few solar masses of material are necessary for collapse.

Some dense, dark clouds have within them even denser areas, called clumps and cores, which may have masses ranging from 0.3 to 10^3 solar masses, and thus can satisfy the Jeans Criteria on their own. So now we have the situation of a large dense, dark cloud collapsing, while inside it there are clumps collapsing as well!

But, as you can imagine, things are far more complicated than the picture we have just drawn for you. As clouds, cores and clumps collapse, they tend to warm up. This acts to inhibit the gravitational collapse; however, this brief hiccup is overcome, and collapse continues.

One point that needs to be mentioned is that most of the diffuse clouds in the interstellar medium are not close to the Jeans Criteria, so some sort of mechanism, or trigger, is needed to change the conditions. In fact, what is needed is something that will increase a cloud's density (i.e., an event that will compress the cloud material into a smaller volume of space[20]). Once a trigger pushes the cloud closer to, and possibly over, the Jeans limits, then cloud collapse can begin (we discuss these possible triggers in the next chapter).

Finally, imagine a massive cloud that does not initially satisfy the Jeans Criteria, but then something causes the cloud to collapse. Areas within the large cloud may now satisfy the criteria, and so they themselves start to contract. In a cloud of several hundred to several thousand solar masses, there can be a lot of clumps, and this *fragmentation*, as it is called, could eventually give rise to a cluster of stars. Thus, this may be a possible scenario for the formation of open star clusters.

---

[20] At this point, the increase in density is thought to be a more important condition than a commensurate increase in temperature.

The Jeans Criteria are a good starting point in the description of cloud collapse, and today there exist far more sophisticated models that perhaps more accurately describe what is going on. Nevertheless, as a starting point, they adequately describe the possible beginning of star formation.

Let us now move on to the next chapter that will discuss those objects we (hopefully) can see any and every night of the year—stars.

Math Box 5.1 Jeans Length and Jeans Mass

The Jeans Length is approximately given by:

$$R_J \approx (kT/Gm^2n)^{1/2}$$

where k is Boltzmann constant = 1.3806×10^{-23} J K^{-1}

T is temperature in Kelvin
G is gravitational constant = 6.67×10^{-11} Nm^2 kg^{-2}
m is mass of hydrogen atom = 1.67×10^{-27} kg
n is number of particles (number density)

Example: If an interstellar cloud has a temperature of 50 K, and there are 10^{11} hydrogen atoms per cubic meter, determine the Jeans Length and Jeans Mass.
Using the above formula, we get:

$$R_J \approx \left[\frac{(1.38\times10^{-23})\times(50)}{(6.67\times10^{-11})\times(1.67\times10^{-27})^2\times(10^{11})}\right]^{1/2}$$

$$R_J \approx 6 \times 10^{15\,\mathrm{m}}$$
$$\approx 0.2 \text{ parsec}$$

The Jeans Mass can easily be estimated by multiplying the density by the volume:

$$M_J = (4\pi/3)(1.67\times10^{-27})(10^{11})(6\times10^{15})^3$$
$$1.5 \times 10^{32} \text{ kg}$$
$$\sim 75 \text{ solar masses.}$$

Thus, in a cloud with a temperature of 50 K that has 10^{11} atoms per cubic meter, ~75 solar masses of material is the minimum amount needed for gravitation to overcome any thermal pressure, with a radius of about 0.2 parsecs.

Thought Question 5.6

Two protostars, one with a mass 15 times larger than the other are formed at the same time. Which will become a main sequence star first?

Problems

1. An interstellar cloud has a temperature of 25 K, and there are 10^{12} hydrogen atoms per cubic meter, calculate the Jeans Length. Give your answer in both meters.
2. Convert your answer to parsecs.

Chapter 6

Star Birth

6.1. The Birth of a Star

A newly born star can be thought of as having been born when the core temperature of the protostar reaches about 10 million K. At this temperature, hydrogen fusion can occur efficiently by the so-called *proton-proton chain.*[1] This moment, when ignition of the fusion process occurs, will halt any further gravitational collapse of the protostar. The star's interior structure stabilizes, with the thermal energy created by nuclear fusion maintaining a balance between gravity and pressure. This important balancing act is called *gravitational equilibrium.*[2] It is also sometimes referred to as *hydrostatic equilibrium*. The star is now a hydrogen-burning main sequence star.

The time between the formation of a protostar to the birth of a main sequence star depends on the star's mass. This is an important point to emphasize. A star's mass determines a lot! A handy reference to remember is that *massive stars do everything faster!* A high-mass protostar may collapse in only a million years or less, while a star with a mass of $\approx 1\ M_\odot$ could take around 50 million years. A star with a very small mass, say, an M-type star, could take well over 100 million years to collapse. This means

---

[1] We will discuss the proton-proton chain in much greater detail in the following sections on the Sun and the main sequence.

[2] See the section on the Sun for a full discussion on gravitational equilibrium.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_6

that very massive stars in a young star cluster may be born, live and die before the very smallest stars finish their infant years!

The changes, or transitions, that occur to a protostar's luminosity and surface temperature can be shown on a special H-R diagram. This is known as an *evolutionary track*, or *life track*, for a star.[3] Each point along the star's track represents its luminosity and temperature at some point during its life, and so it shows us how the protostar's appearance changes due to changes in its interior. Figure 6.1 shows the evolutionary tracks for several protostars of different masses, from 0.5 $M_\odot$ to 15 $M_\odot$. (It is important to realize that these evolutionary tracks are theoretical models, and the predictions are only as good as the theory,[4] though they seem to work very well and are being improved all the time.) Recall that protostars are relatively cool, and so the tracks all begin at the right side of the H-R diagram. However, subsequent evolution is very different for stars of differing mass.

As an example of an evolutionary life track for a protostar, we shall look at the life track for a 1 $M_\odot$, rather like the Sun. This period in the star's life has four very distinct phases.

The protostar first forms from a cloud of cold gas and thus is on the far-right side of the H-R diagram. However, its surface area is enormous, so its luminosity can be very large—maybe 100 times the luminosity it will have when it becomes a star.

Due to its large luminosity, the young protostar rapidly loses the energy it generated via gravitational collapse, and so further collapse proceeds at a relatively rapid rate. Its surface temperature increases slightly during the next several million years, but its diminishing size reduces the luminosity. The evolutionary track now progresses almost vertically downward on the H-R diagram.

Now that the core temperature has reached 10 million K, hydrogen nuclei fuse into helium. The rate of nuclear fusion, however, is not sufficient to halt the collapse of the star, although it is slowed down considerably. As the star shrinks, its surface temperature increases. The result of shrinking and heating is a small increase in luminosity over the next 10 million years. The evolutionary track now progresses leftward and slightly upward on the H-R diagram.

---

[3] When astronomers refer to a star's following a specific evolutionary track, or moving on an H-R diagram, what they really mean is the star's luminosity and/or temperature changes. Thus, the position of the star on the H-R diagram will change.

[4] The theoretical calculations were developed by the Japanese astrophysicist C. Hayashi, and the phase a protostar undergoes before it reaches the main sequence is called the *Hayashi phase*.

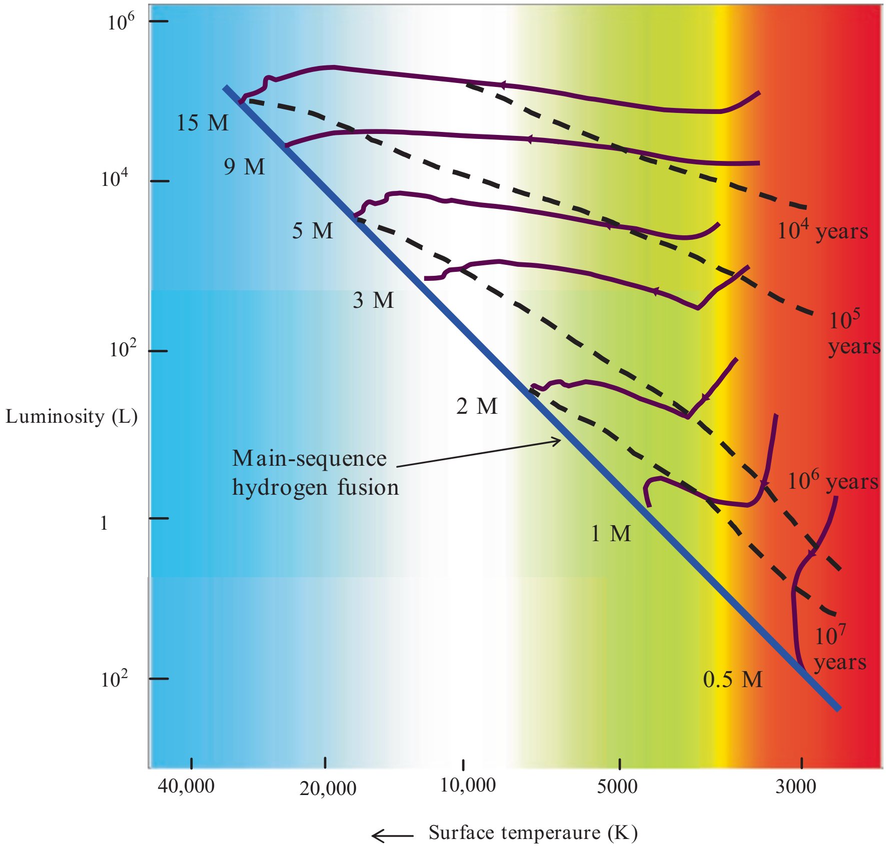


Fig. 6.1. Pre-main sequence life tracks. The evolutionary life tracks of seven protostars are shown. Also identified are the stages reached after an indicated number of years of evolution (*dashed lines*). The mass shown for each protostar is the final mass it has when it becomes a main sequence star. Note that the greater the mass, the higher the temperature and luminosity

Note that both the rate of nuclear fusion and the core temperature increase over the next tens of millions of years. Once the rate of fusion is high enough, gravitational equilibrium is achieved, and fusion becomes self-sustaining. The result is that the star settles onto the hydrogen-burning main sequence (Fig. 6.2).

From the viewpoint of an observer, this fourth stage of stellar evolution doesn’t present itself with many visible objects. Even though the luminosity of such objects is very high, we will never see one. The reason is obvious: they are enshrouded within vast clouds of interstellar dust that, if you recall, are very efficient at blocking out light. The dust in the vicinity of a protostar,

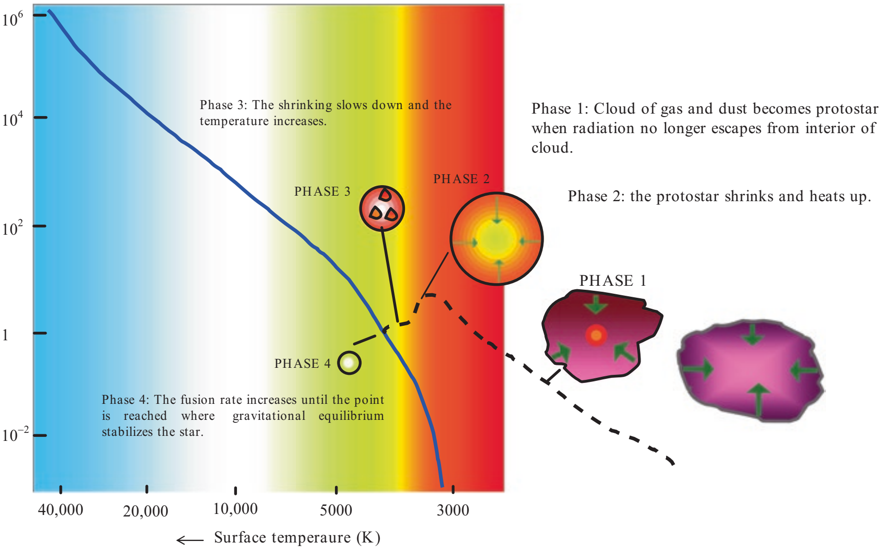


Fig. 6.2. The evolutionary track of a 1 solar mass object

often called a *cocoon nebula*, absorbs the light, and so it is very difficult to observe at visible wavelengths.[5] On the other hand, they can be seen at infrared wavelengths. But this doesn't really help us as visual observational astronomers on Earth.

Thought Question 6.1

An amateur astronomer says he has observed a protostar visually through his telescope. Do you think he has?

6.2. Pre-Main-Sequence Evolution and the Effect of Mass

The previous sections explained how a cloud could contract and become a protostar. In fact, due to the immense amount of material in a molecular cloud, it is believed that rather than an individual protostar being formed, several are formed as a star cluster. However, there is a slight problem with

---

[5] There are a few examples of nebulae in which protostars are currently forming and which are observable in the section on emission nebulae. You will not, however, see protostars, just the region within which they reside.

this scenario; at the time of writing this book, there is no satisfactory explanation for protostars of differing masses actually forming within the same cloud. Just what are the processes that govern the clumping and fragmentation of the cloud into protostars of widely differing masses? Even though we cannot explain the process, we can at least observe its results.

Let's begin this section by looking at how protostars of differing mass are believed to form, and we'll start with a star of 1 $M_\odot$, a star just like our Sun. The outer layers of such a protostar are cool and opaque,[6] which means that any energy released as radiation due to the shrinkage of the inner layers cannot reach the surface. Thus, the only way of moving this energy toward the surface layers must be by the less efficient and slower method of *convection*. The result of this process is that the temperature remains more or less constant as the protostar shrinks, while the luminosity decreases because the radius decreases,[7] and the evolutionary track moves downward on the H-R diagram. This is depicted in Fig. 6.1.

We said previously that the surface temperature remains roughly constant during this phase, but conditions inside the protostar are far from unchanging. The internal temperature starts to increase during this time, and the interior becomes ionized. This reduces the opacity within the protostar and allows the transfer of energy to be by radiation in the interior regions and by convection in the outer layers. This process is the one that is ongoing within the Sun today.

The net result of these changes is that energy can escape much more easily from the protostar, and thus the luminosity increases. This increase in energy transport is represented by the evolutionary track's bending upward (meaning higher luminosity) and to the left (higher temperature). After an interval of a few million years, the temperature within the protostar is high enough—10 million K—for nuclear fusion to begin, and, eventually, enough heat and associated internal pressure are created so as to balance the gravitational contraction of the star. We can say that at this point, hydrostatic equilibrium has been reached, and the protostar has reached the main sequence. It is now a main sequence star.

As to be expected, a more massive protostar will evolve in a different way. Protostars with a mass of about or greater than 4 $M_\odot$ contract and heat up at a more rapid rate, and so the hydrogen-burning phase begins earlier. The net result is that the luminosity stabilizes at approximately its final value, but the surface temperature continues to increase as the protostar

---

[6] We shall see in a later section why the Sun is opaque.

[7] Recall from an earlier section that the luminosity is proportional to the square of the radius and to the fourth power of the surface temperature.

continues to shrink. The evolutionary track of such a high-mass protostar is illustrated on the H-R diagram; the luminosity is nearly horizontal (meaning nearly constant luminosity) from right to left (increasing surface temperature). This is especially so for the stars at mass 9 $M_\odot$ and 15 $M_\odot$.

An increase in mass will result in a corresponding increase in pressure and temperature in the interior of a star. This is very significant because it means that in the very massive stars, there is a much greater temperature difference between the core and its outer layers as compared to, say, the Sun. This allows convection to occur much deeper in the star's interior regions. In contrast, the massive star will have very low-density outer layers, and so energy flow in these regions is more easily performed by radiative methods than by convective methods. Thus, stars on the main sequence with a mass greater than about 4 $M_\odot$ will have convective interiors and radiative outer layers, while stars less than about 4 $M_\odot$ will have radiative interior regions and convective outer layers.

At the very low end of the scale, stars that have a mass less than about 0.8 $M_\odot$ have a very different internal structure. In these objects, the interior temperature of the protostar is insufficient to ionize the inner region and so is too opaque to allow energy transport by radiation. The only possible method to transport the energy to the outer layers in these stars is by convection. Examples of the interior structures of low-mass, high-mass and very low-mass stars are shown in Fig. 6.3.

A very important point to make here is that all the evolutionary tracks shown on Fig. 6.1 end at the main sequence. Thus, *the main sequence represents those stars in which nuclear fusion reactions are producing energy by converting hydrogen to helium*. For the large majority of stars, this is a stable situation, and a mass-luminosity relationship, which is graphically shown in Fig. 6.4, can represent this endpoint on the main sequence. What this diagram implies is that the hot bright blue stars are the most massive, while the cool faint cool stars are the least massive.[8] Thus, the H-R diagram is a progression not only in luminosity and temperature but in mass as well. This can be succinctly summed up as *the greater the mass, the greater the luminosity.*

Now that we have discussed how stars are formed, and how the H-R diagram describes star birth, it is important to emphasize two factors that can cause confusion.

- First, if you look at the evolutionary track of protostars, several of them, especially the high-mass protostars, begin in the upper-right region. But they are not red giant stars! The red giant stars are at a stage in their lives that occurs only *after* being a main sequence star.

---

[8] There is no mass-luminosity relationship for white dwarfs, giant stars and supergiant stars.

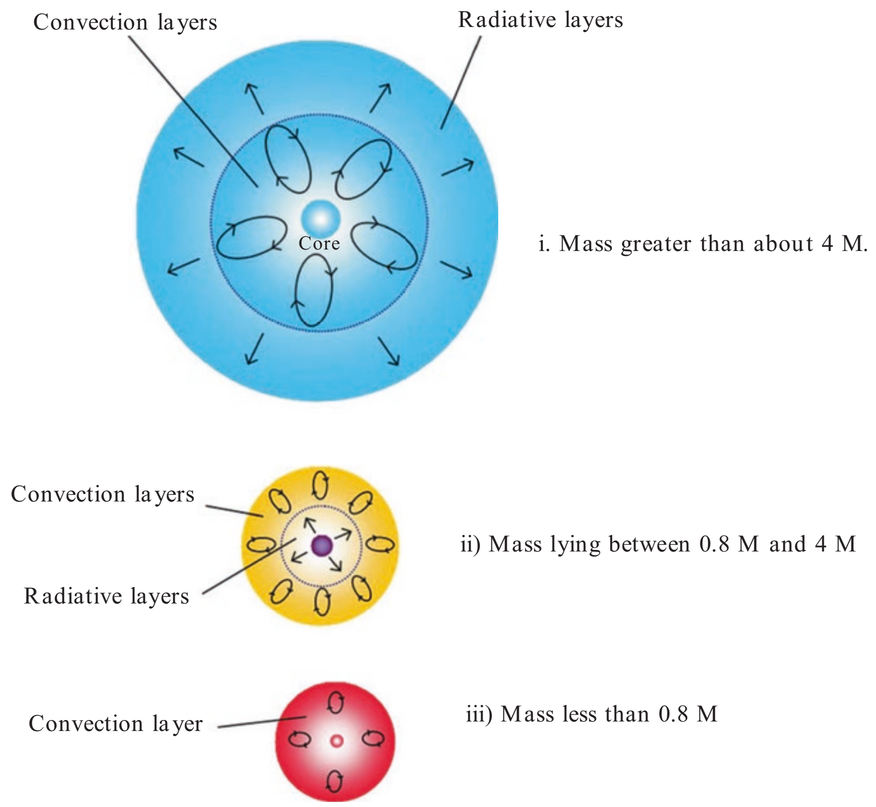


Fig. 6.3. Mass of main sequence stars. (i) Energy flows from the core by convection in the inner regions and by radiation in the outer layers in stars of mass greater than 4 $M_{\odot}$. (ii) Energy flows outward from the core by radiative means in inner regions and by convection in outer layers in stars with a mass of less than 4 $M_{\odot}$ and greater than 0.8 $M_{\odot}$. (iii) Energy flows outward by convection throughout the interior of the stars with a mass of less than 0.8 $M_{\odot}$

- The second point to note is that most stars spend *most* of their lives on the main sequence and only a relatively brief time as protostars. For example, a 1 $M_{\odot}$ protostar takes about 20 million years to become a main sequence star, while a 12 $M_{\odot}$ may only take 20,000 years. In contrast, a star like the Sun has been a main sequence star for nearly 5 billion years and will remain one for another 5 billion!

One final point is that the masses of stars have limits. Using theoretical models, astronomers have deduced that stars above ≈150–200 $M_{\odot}$ cannot form[9]; they generate so much energy that gravity cannot contain their inter-

---

[9] Not surprisingly, there are recent reports of stars with masses in excess of 200 $M_{\odot}$, as large as 250 $M_{\odot}$. How these stars can exist is a matter of much research and fierce debate.

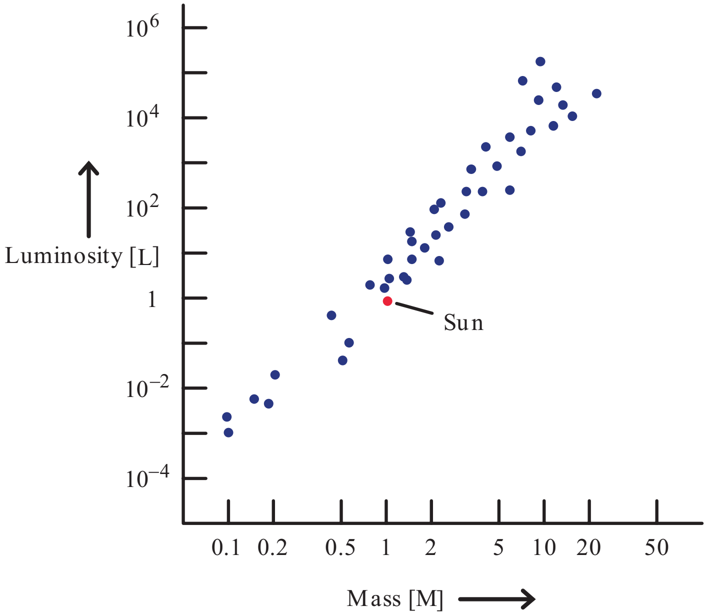


Fig. 6.4. Mass-luminosity relationship. For stars on the main sequence, there is a ratio between the mass and luminosity. Basically, the more massive the star, the greater its luminosity. A star of mass 10 $M_\odot$ has about 3000 $L_\odot$; similarly, a star of mass 0.1 $M_\odot$ has a luminosity of only 0.001 $L_\odot$

nal pressure. These stars literally tear themselves apart. At the other end of the scale, there is also, not surprisingly, a lower limit. Those stars with a mass of less than 0.08 $M_\odot$[10] can never achieve the 10 million K core temperatures necessary to initiate nuclear hydrogen fusion. So, what is actually formed can be thought of as a "failed star," which will slowly radiate away all of its internal energy, gradually cooling with time. These objects have been called *brown dwarfs* and seem to occupy a strange area between what we think of as a planet and a star. Brown dwarfs radiate in the infrared, making them very difficult to detect. The first known detection was in 1995 of Gliese 229B, a 0.05 $M_\odot$ object. Many astronomers believe these small, elusive objects are far more common than previously thought and may in fact be the most common form of ordinary matter[11] in the universe.

---

[10] This figure of 0.08 $M_\odot$ is about 80 times the mass of Jupiter.

[11] By "ordinary" we mean matter composed of atoms, to distinguish it from "dark matter," whatever that may be!

From an observational point of view, this period in a star's life does not present us with many observable opportunities. The protostars are cocooned within vast clouds of gas and dust and are therefore invisible to us. Some objects are, of course, "visible" if infrared telescopes are used. However, it is always worthwhile to look at areas of the night sky where we know such objects exist, even though they cannot be seen. We can always use our imagination as we gaze at them, knowing that hidden deep in these clouds are stars in the process of being formed. One such object is, of course, the Orion Nebula.[12]

Messier 42	NGC 1976	05h 35.4m	−05° 23.5′	December
4.0 m	• 1 to 5	⊕ 85|60′		

Also known as the Orion Nebula. This is the premier emission nebula and one of the most magnificent objects in the sky. It is part of the vast Orion complex, which contains star-forming regions, molecular clouds, and all sorts of nebulae! Visible to the naked eye as a barely resolved pah of light, it shows detail from the smallest aperture upwards. In binoculars, its pearly glow will show structure and detail, and in telescopes of aperture 10 cm, the whole field will be filled. The entire nebulosity is glowing due to the light (and thus energy) provided by the famous Trapezium stars located within it. These stars are stellar powerhouses, pouring forth vast amounts of energy, and they are fairly new stars. What is also readily seen along with the glowing nebula are the dark, apparently empty and starless regions. These are still part of the huge complex of dust and gas but are not glowing by the process of fluorescence; instead, they are vast clouds of obscuring dust, the dark nebulae mentioned in the previous chapter. The emission nebula is one of the few that shows definite color. Many observers report seeing a greenish glow, along with pale gray and blue, but to observe any color besides gray will require excellent observing conditions. Also, amateurs state that with very large apertures of 35 cm, a pinkish glow can be seen. Located within the nebula are the famous Kleinmann-Low sources and the Becklin-Neugebauer object, which are believed to be dust-enshrouded young stars. The whole nebula complex is a vast stellar nursery. M42 is at a distance of 1700 light years from our Sun and about 40 light years in diameter. Spend a long time observing this object—you will benefit from it—consider letting let the nebula drift into the field of view. It is a sight to behold!

---

[12] Yes, we have seen this before, but it is such a magnificent, and important, object that it warrants a second look.

6.3. Mass Loss and Mass Gain

6.3.1 T Tauri Stars

After reading the previous sections, you may have gotten the idea that star formation is simply a matter of material falling inward due to gravity. In fact, most of the material that makes up a cloud is ejected into space as, and after, the star(s) form and may never form a star at all. This ejected material can help sweep away the gas and dust surrounding the young stars and make them visible to us. Several examples of such a process can be seen in the Rosette Nebula, the Trifid Nebula and the Bubble Nebula, mentioned earlier.

There are also examples of individual objects that eject material into space during this event in a star's birth. These are called *T Tauri stars*, which are protostars whose luminosity can change irregularly in a matter of just a few days, and which also have both absorption and emission lines in their spectrum. In addition, due to the conflict between gravitational contraction and hydrogen burning in these first stages of main sequence stability, the element lithium is produced. Spectral lines of lithium are a signature of protostars of the T Tauri type.

The masses of these stars are less than about 3 $M_\odot$, and they seem to be about 1 million years old. If placed on an H-R diagram, they would be on the right-hand side of the main sequence. By analyzing the emission lines, we can see that surrounding these protostars are very thin clouds of very hot gas, which the protostar has ejected into space at a speed of about 80 km/s^{-1} (300,000 km/h^{-1}). A T Tauri star bears a superficial resemblance to the Sun in that it will exhibit a spectral type of F, G or K, with a surface temperature of 4000–8000 K.

Over the period of a year, a typical T Tauri star would have ejected about 10^{-8}–10^{-7} solar masses. You may think that this is a very small amount, but compared with the Sun, which loses about 10^{14} $M_\odot$ a year, it is significant. This phase of a protostar, called the T Tauri phase, can last as long as 10 million years, during which it can eject roughly 1 $M_\odot$ of material. A consequence of this is that the mass of the final main sequence star is very much less than the mass it started with. As these are objects associated with star birth, they are often, if not always, found near or in the Milky Way.

Other young stars with masses greater than 3 $M_\odot$ do not vary in luminosity like T Tauri stars; they do eject mass, however, due to the extremely high radiation pressure at their surfaces. This class of star is called *Ae* or *Be* stars. Stars greater than 10 $M_\odot$ will reach the main sequence before the surrounding dust and gas from which they formed has had a chance to disperse, and so these stars are often detected as highly luminous infrared objects located within molecular clouds.

Fortunately for us as observers, there are several visible examples of T Tauri stars. They are, however, extremely faint, and so only the archetypal one is mentioned below.

T-TAURI	04H 22.0M	+19° 32′	NOVEMBER
8.5 to 13.6$_{v}$m	dGe - K1e		

The prototype T-Tauri star is about 1.8° west and slightly north of ε (epsilon) Tauri, the northernmost bright star in the famous "v" shape of the Hyades star cluster. Discovered in 1852 by J. Hind (who also discovered the associated nebula, Hind's Variable Nebula). The star varies irregularly in several aspects: the brightness varies from about eighth to thirteenth magnitude, the period, with a range from a few weeks to perhaps a few months, and the spectrum varying from F8 to K1. Oddly enough, the variation in spectral type does not necessarily correlate with variability in magnitude. T Tauri and the nebula lie within the Taurus Dark Cloud Complex, within which there are numerous, but faint, variable nebulae and recently formed stars. (Other T Tauri and similar stars are VV Tauri and FU Orionis.[13])

Thought Question 6.2

At what point during a star's birth would you be able to observe it visually?

6.3.2 Discs and Winds

One aspect of protostar formation that came as a surprise to astronomers in the late twentieth century was a curious phenomenon observed in many young stars, including the T Tauri stars mentioned above. It involves a loss of mass, once again, but the loss is directed out from the young star in two jets; these are very narrow, usually flowing out along the rotation axis of the

---

[13] Stars named after the FU Orionis prototype are also worth observing. It is now believed that the activity of FU Orionis (and similar stars) is related to the T Tauri variables. T Tauri variations may result from instabilities within and interactions with the surrounding accretion disk. FU Orionis activity is caused by a dramatic increase in instability due to the dumping of large amounts of material onto an accompanying star. Many astronomers believe that all T Tauri stars probably go through FU Orionis-type behavior at least once in their development.

star and in opposite directions. This jet outflowing is referred to as a *bipolar outflow*. The material is moving with a velocity that can reach several hundred kilometers per second, and it sometimes interacts with the surrounding debris left over from star formation to form clumpy knots of material called *Herbig-Haro objects*. The lifetime of such a phenomenon is relatively short, maybe from 10,000 to 100,000 years. The mechanism that forms these jets is not yet fully understood, although it is believed to involve magnetic fields.

We have discussed mass loss in a protostar, but there exists a mechanism that can add mass to the normal star-formation process. Recall that a protostar is formed from in-falling gas and dust due to gravity. As this cloud of denser material clumps together, the protostar nebula will begin to rotate. This is just a consequence of physics, and it is called the *conservation of angular momentum*. The material will flatten itself out and form a disc, or *protostellar disk*, as it is called. The gas and dust particles within the nebula collide and spin inwards onto the forming protostar, thus adding to its mass. This process is often called *accretion*, and the build-up of material onto the ever faster-rotating disc is called the *circumstellar accretion disc*.

The interactions between the magnetic fields, the jets, and the accretion disc is thought to slow down the protostar's rotation, which would explain why most stars have a much slower spin than protostars of similar mass.

Since the 1990s, the discovery of accretion disks around new stars led astronomers to speculate that these are the precursors to possible planetary formation. Many of these splendid objects were discovered in Orion but are, naturally, unobservable for the amateur astronomer.[14]

Thought Question 6.3

If two protostars are born at the same time in a molecular cloud, with mass 15 $M_\odot$ *and* $2M_\odot$ *respectively, which one will reach the main sequence stage first?*

6.4. Star Formation Triggers

We have seen how stars are formed from clouds of dust and gas, and how these clouds clump together under the force of gravity to form protostars. In addition, the evolution of a protostar to the main sequence depends on the initial mass of the protostar and so determines where it will arrive on there.

---

[14] Eventually an amateur astronomer will image a protostar. It's only a matter of time and money.

The one thing we have not yet addressed in any detail, although it was mentioned at the end of the previous chapter when the Jeans Mass was discussed, is what *causes* a protostar to form in the first place! This is the topic of the final part of this section.

The mechanisms that provide the "triggers" for star formation have three very disparate origins:

6.4.1 The Spiral Arms of a Galaxy

The spiral arms of galaxies are a prime location for star formation because the gas and dust clouds temporarily "pile up" as they orbit around the center of a galaxy.[15] In such a spiral arm, the molecular clouds are compressed as it passes through the region. In the molecular cloud's densest regions, vigorous star formation can then occur.

6.4.2 Expanding HII[16] Regions

Massive stars, such as O-type and B-type stars, emit immense amounts of radiation, usually in the ultraviolet part of the spectrum. This, in turn, causes the surrounding gas to ionize, and an HII region is formed within the larger molecular cloud. The strong stellar winds and ultraviolet radiation that O- and B-type stars possess can carve out a cavity within the molecular cloud into which the HII region expands. The stellar wind is moving at such a high velocity that it is supersonic (i.e., faster than the speed of sound in that particular region). A shock wave associated with this supersonically expanding HII region then collides with the rest of the molecular cloud. In doing so, it compresses the cloud, and so further star formation occurs. The new O- and B-type stars that result from this induce further star formation, but at the same time, the precursor O- and B-type stars that originally started the procedure may well have dispersed by this point. In this manner, an OB association "devours" a molecular cloud, leaving older stars in its wake. The Orion Nebula is one example of such a mechanism, where the four stars of

---

[15] We are talking about spiral galaxies here, not elliptical. Elliptical galaxies are believed to be the result of mergers between spiral galaxies where the rate of star formation is very high initially—a starburst galaxy—and then falls. We shall cover this topic in a later chapter.

[16] Pronounced "aitch 2."

the Trapezium are ionizing the surrounding material. The nebula itself is at the edge of a giant molecular cloud, some 500,000 $M_\odot$.

6.4.3 Supernova

The final mechanism that is believed to induce further star formation is a supernova. As we shall see in later a chapter, a supernova marks the death of a star and results in a catastrophic explosion, usually blowing the star to bits! What is important to us at this stage is that the outer layers of the star are ejected into space at incredible speeds, maybe several thousand kilometers per second! This shock wave, an expanding shell of material, will be moving at supersonic velocities and, in a similar manner as mentioned above, will impact on material in the interstellar medium, and in doing so will compress and heat it. This will stimulate further star formation.

We have now covered the amazing processes involved in star formation, from vast clouds of dust and gas to glowing spheres of nuclear fusion—the birth of a star. However, do not think that we know all there is to know about star birth, because we don't! For instance, a spiral arm that passes through a giant molecular cloud tends to produce giant O- and B-type stars, whereas the stars induced by supernovae shock waves are predominantly A-, F-, G- and K-type stars.

Also, in our home galaxy, there often seems to be a lot of dust associated with star formation that shields the newly born stars from the destructive effects of ultraviolet radiation from other hot stars that are close by. However, in a nearby galaxy—the Large Magellanic Cloud[17] (LMC)—it's been observed that young OB associations have hardly any dust at all! Nevertheless, what we do know is amazing and involves mechanisms from star death to the rotation of galaxies.

Most of the stars that we observe in the night sky all have one thing in common: they are on the main sequence. There are, of course, exceptions, Betelgeuse has left the main sequence and has become a red giant star; the hydrogen-burning at the center of its core has stopped, and now helium is burning by fusion processes. Also, Sirius B has evolved far from the main sequence and has become a white dwarf star, with no nuclear fusion at all occurring within it. But for the large majority, the main sequence is a stable time, with only small changes in mass and luminosity occurring.

---

[17] It isn't a cloud at all; this is just the name ancient astronomers gave the galaxy before they knew what it really was!

However, as a star ages, changes occur in the way energy is formed, and this in turn affects its size and thus its luminosity, and so the star leaves the main sequence to begin the next phase of its life. In the following chapters we will look at these different periods in a star's life, whether it is a small, low-mass and cool star, or a bright, high-mass and hot star.

Thought Question 6.4

If a molecular cloud initially has no star formation triggers, what do you think is the only mechanism that can initiate dust and gas cloud collapse?

Problem

1. Consider Fig. 6.1, approximately how many years longer does it take a 0.5 $M_\odot$ star to form compared with a 14 $M_\odot$ star?

Chapter 7

Galactic Clusters and Stellar Associations

A casual glance at the night sky may lead you into believing that stars are solitary, isolated objects, but in fact no star is born in isolation. As we have discussed in previous chapters, the process of star birth takes place in immense interstellar clouds of dust and gas that, depending upon the cloud's size, can give rise to anything from a few dozen to many thousands of new, and young, stars.

Over time, however, this stellar nursery of young stars will gradually disperse. Theory predicts that massive stars have much shorter life spans than smaller, less massive ones, such that the more massive stars do not live long enough to escape their birthplace, whereas a smaller star, say, of solar mass size, will in most cases easily migrate from its stellar birthplace.

It's important to note, in relation to stars of mass about equal to that of the Sun, that where there may be several thousands of the objects, the combined gravitational attraction of so many stars may slow down the dispersion of the group. It really depends on the star density and mass of the particular cluster. So we can infer that the densest or most closely packed clusters, containing solar mass-sized stars, will be the ones that contain the oldest populations of stars, while the most scattered clusters will have the youngest stellar population.

There are two main types of star clusters, those that are relatively new—*open* or *galactic clusters*, and those that contain very old stars—*globular clusters*. Both are very different in content and appearance, as we shall now see.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_7

However, as open clusters are relatively new objects, and are associated with star birth, while globular clusters are old, and associated with the aging of stars and their inevitable demise, we will only discuss the former here, and leave the latter until the chapter on star death.

7.1. Galactic Star Clusters

The stars that form out of the same cloud of material will not necessarily all have the same mass. Far from it!

The masses will differ and, as a consequence, reach the main sequence at differing times. As mentioned earlier, high-mass stars evolve faster than low-mass stars, and so at a time when these high-mass stars are shining brightly as stars in their own right, the low-mass protostars may still be cocooned within their dusty mantles. Consequently, the intense radiation emitted by the new, hot and bright stars may disturb the normal evolution of the low-mass stars, and so reduce their final mass.

Open clusters, or *galactic clusters*, as they are sometimes called, are collections of young stars containing anywhere from a dozen members to hundreds. A few of them (for example, Messier 11 in Scutum) contain an impressive number of stars, equaling that of globular clusters, while others seem little more than a faint grouping set against the background star field. Such is the variety of open clusters that they come in all shapes and sizes. Several are over a degree in size, and their full impact can only be appreciated by using binoculars, as a telescope has too narrow a field of view. An example of such a large cluster is Messier 44 in Cancer. Then there are tiny clusters, seemingly nothing more than compact multiple stars, as is the case with IC 4996 in Cygnus. In some cases, all the members of the cluster are equally bright, such as Caldwell 71 in Puppis; but there are others that consist of only a few bright members accompanied by several fainter companions, as is the case of Messier 29 in Cygnus. The stars that make up an open cluster are called *Population I* stars, which are metal-rich and are usually to be found in or near the spiral arms of the galaxy.

The size of a cluster can vary from a few dozen light years across, as in the case of NGC 255 in Cassiopeia, to about 70 light years across, as in either component of Caldwell 14, the Perseus Double Cluster.

The reason for the varied and disparate appearances of open clusters lies in where and how they were born. It is the interstellar cloud that determines both the number and type of stars that are born within it. Factors such as the size, density, turbulence, temperature and magnetic field all play a role as the deciding parameters in star birth. In the case of *giant molecular clouds*, or GMCs, the conditions can give rise to both O- and B-type giant stars along with solar-type dwarf stars—whereas in *small molecular clouds* (SMCs), only solar-type stars will be formed, with none of the luminous B-type stars. An example of an SMC is the Taurus Dark Cloud, which lies just beyond the Pleiades.

By observing a star cluster, we can study in detail the process of star formation and interaction between low- and high-mass stars. As an example, look at Fig. 7.1, which shows the H-R diagram for the cluster NGC 2264, located in Monoceros. Note that all the high-mass stars, which are the hottest stars, with a temperature of about 20,000 K, have already reached

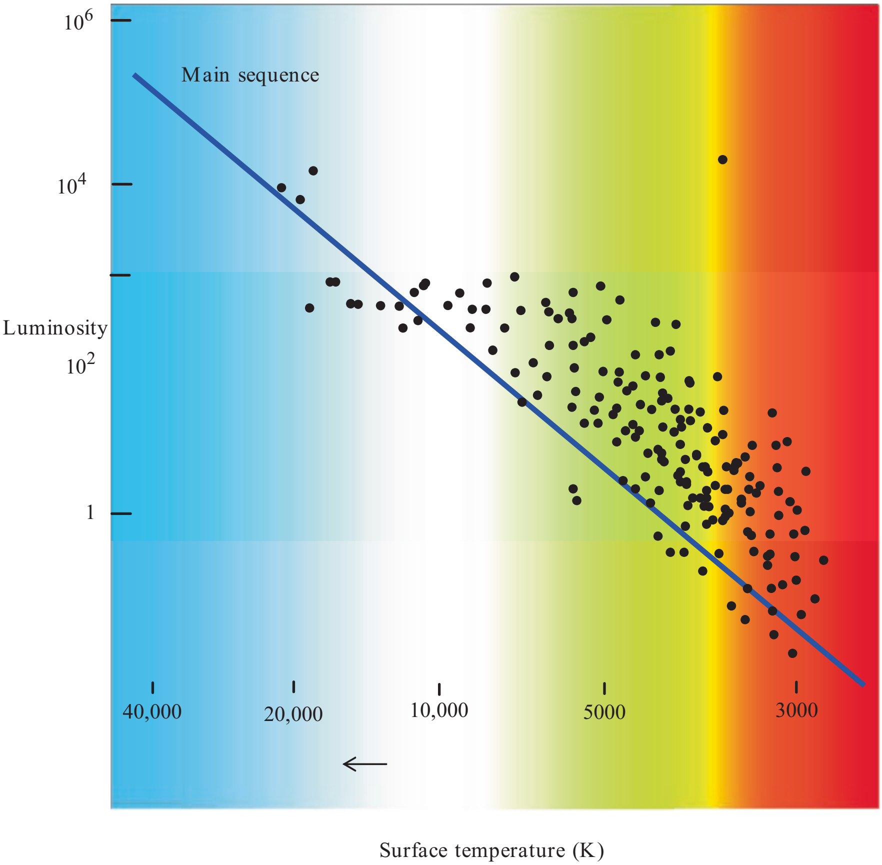


Fig. 7.1. NGC 2264. This young star cluster is about 800 pc from Earth and contains many T Tauri stars. Each *dot* is a star whose temperature and luminosity have been measured

the main sequence, while those with temperatures at about 10,000 K or below have not. These low-mass and cooler stars are in the latter stages of pre-main-sequence star formation, with nuclear fusion just about beginning at their cores. Astronomers can compare this H-R diagram with the theoretical models, and they have deduced that this particular cluster is very young, only 2 million years old.

By comparison, we can look at the Pleiades star cluster—a very famous cluster—on the H-R diagram, Fig. 7.2. We can see straightaway that the cluster must be older than NGC 2264 because most of the stars are already on the main sequence. From studying the Pleiades H-R diagram, astronomers

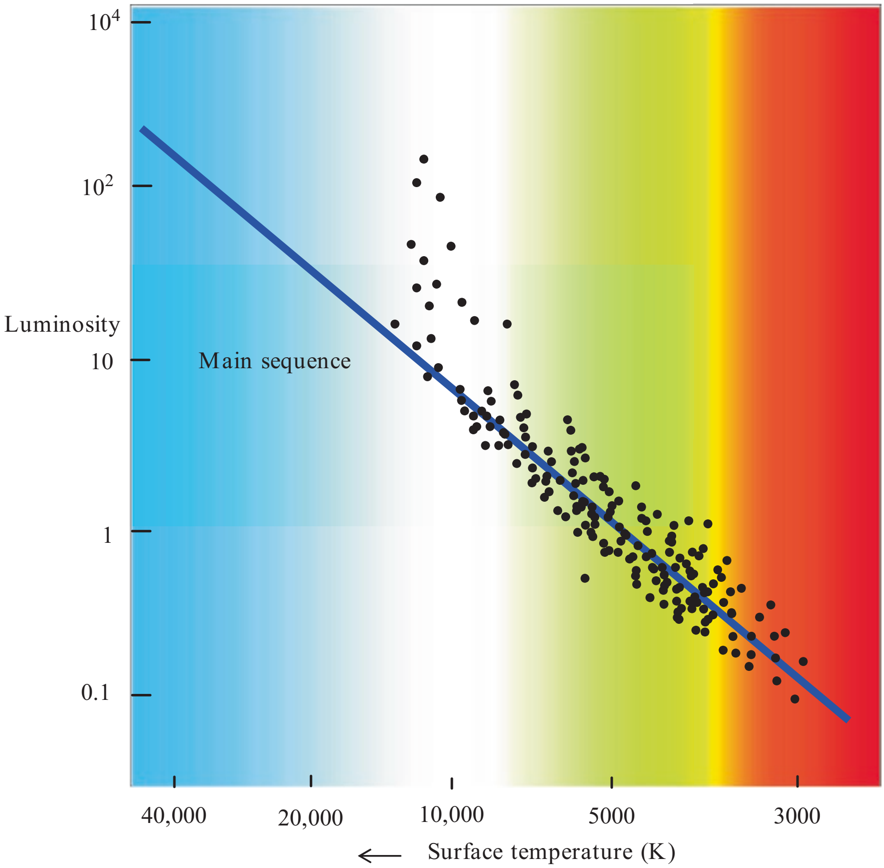


Fig. 7.2. The Pleiades star cluster. This is a much older cluster at 50 million years. Most, if not all, of the low-mass cool stars have reached the main sequence, which implies that hydrogen-burning has started in their cores

believe the cluster to be about 50 million years old. Also look at the area on the H-R diagram with a temperature of about 10,500 K, and luminosities ranging from 10 to 10^2 $L_{\odot}$. You will see a few stars that do not seem to lie on the main sequence. This isn't because they are still in the process of being formed; on the contrary, these massive stars have left the main sequence. They were among the first to be formed and thus are the oldest, and they are now evolving into a different kind of star. As we shall see later, all the hydrogen at the center of these stars has been used up,[1] and helium-burning is now proceeding.

An interesting aspect of open clusters is their distribution in the night sky. You may be forgiven for thinking that they are randomly distributed across the sky, but surveys show that although well over a thousand clusters have been discovered, only a few are observed to be at distances greater than 25° above or below the galactic equator. Some parts of the sky are very rich in clusters—Cassiopeia, and Puppis 4, and this is due to the absence of dust lying along these lines of sight, allowing us to see across the spiral plane of our galaxy. Many of the clusters mentioned here actually lie in different spiral arms, and so as you observe them, you are actually looking at different parts of the spiral structure of our galaxy.

Thought Question 7.1

An amateur astronomer says she has observed an open cluster in the Halo of the Milky Way. Is she correct?

It was mentioned earlier that stars are not born in isolation. Nor are they born simultaneously. Recall that the more massive a star, the faster it contracts and becomes stable, thus joining the main sequence; this results in some clusters having bright young O and B main sequence stars, while at the same time containing low-mass members that may still be in the process of gravitational contraction (for example, the star cluster at the center of the Lagoon Nebula). In a few cases, the star production in a cluster is at a very early stage, with only a few stars visible, the majority still in the process of contraction and hidden within the interstellar cloud.

A perfect example of such a process is the open cluster in Messier 42, the Orion Nebula. The stars within the cluster, the Trapezium, are the brightest, youngest and most massive stars in what will eventually become a large cluster containing many A-, F- and G-type stars. However, the majority are

---

[1] Remember that hydrogen-burning is a characteristic for stars on the main sequence.

blanketed by the dust and gas clouds and are only detectable by their infrared radiation.

As time passes, the dust and gas surrounding a new cluster will be blown away by the radiation from the O-type stars, resulting in the cluster's becoming visible in its entirety, such as in the case of the young cluster Caldwell 76 in Scorpius.

Once a cluster has formed, it will remain more or less unchanged for at least a few million years, but then changes within the cluster may occur. Two processes are responsible for changes within any given cluster.

1. The evolution of open clusters depends on both the initial stellar content of the group and the ever-pervasive pull of gravity. If a cluster contains O-, B- and A-type stars, these stars will eventually become supernovae, leaving the cluster with slower-evolving, less massive and less luminous members of type A and M stars. A famous example of such a cluster is Caldwell 94, the Jewel Box in Crux, which is a highlight of the southern sky, and, alas, unobservable to northern hemisphere observers. However, these too will become supernovae, with the result that the most luminous members of a cluster will, one by one, disappear over time. This doesn't necessarily mean the demise of a cluster, especially clusters that have many tens or hundreds of members. But some, which consist of only a few bright stars, will seem to meld into the background star field.
2. However, even those clusters that have survived the demise of their brighter members will eventually begin to feel the effect of a force that pervades everywhere—the galaxy's gravitational field. As time passes, the cluster will be affected by the influence of other globular clusters and the interstellar matter itself, as well as the tidal force of the galaxy. The cumulative effect of all of these encounters will result in some of the less-massive members of the cluster acquiring enough velocity to escape from it. Thus, given enough time, a cluster will fade and disperse. (Take heart, as this isn't likely to happen in the near future so that you would notice. The Hyades star cluster, even after having lost most of its K- and M-type dwarf stars, is still with us after 600 million years!).

For the amateur, observing open clusters is a very rewarding experience, as they are readily observable, from naked-eye clusters to those visible only in larger telescopes. Happily, many of them are best viewed by binoculars, especially the larger clusters that are of an appreciable angular size. Furthermore, nearly all have double or triple stars within the cluster, and so, regardless of magnification, there is always something of interest to be seen.

Images of some star clusters can be seen at the end of the chapter.

From the preceding chapters, you will know that color in observed stars is best seen when contrasted with a companion(s). Thus, an open cluster presents a perfect opportunity for observing star colors. Many clusters, such as the ever and rightly popular Pleiades, are all a lovely steely blue color. On the other hand, Caldwell 10 in Cassiopeia has contrasting bluish stars along with a nice orange star. Other clusters have a solitary yellowish or ruddy orange star along with fainter white ones, such as Messier 6 in Scorpius. An often-striking characteristic of open clusters is the apparent chains of stars that are seen. Many clusters have stars that arc across apparently empty voids, as in Messier 41 in Canis Major.

Thought Question 7.2

You observe an open cluster that has a lot of blue stars. Is this a recently formed cluster, or one that has been around for a long time?

Because open clusters display such a wealth of characteristics, different parameters are assigned to a cluster that describe its shape and content. For instance, a designation called the *Trumpler* type is often used. It is a three-part designation that describes the cluster's degree of concentration—that is, from a packed cluster to one that is evenly distributed, the range in brightness of the stars within the cluster and finally the richness of the cluster, from poor (fewer than 50 stars) to rich (more than 100). The full classification is:

7.2. Trumpler Classification for Star Clusters

7.2.1 Concentration

I.	Detached—strong concentration of stars towards the center
II.	Detached—weak concentration of stars towards the center
III.	Detached—no concentration of stars towards the center
IV.	Poor detachment from background star field

7.2.2 Range of Brightness

1	Small range
2	Moderate range
3	Large range

7.2.3 Richness of Cluster

p	Poor (with fewer than 50 stars)
m	Moderate (with 50–100 stars)
r	Rich (with more than 100 stars)
n	Cluster within nebulosity

Two further and final points that can often cause problems need to be mentioned: the magnitude and size of the cluster. The quoted magnitude of a cluster may be the result of only a few bright stars, or, on the other hand, may be the result of a large number of faint stars. Also, the diameter of a cluster is often misleading, as in most cases it has been calculated from photographic plates, which, as experienced amateurs will know, bear little resemblance to what is seen at the eyepiece.

Although magnitudes and diameters may be quoted in the text, do treat them with a certain amount of caution.

In the descriptions given below, the first line lists the name, the position, and the approximate midnight transit time; the second line presents the visual magnitude (the combined magnitude of all stars in cluster), object size in arc minutes (⊕), the approximate number of stars in the cluster (bear in mind that the number of stars seen will depend on magnification and aperture, and will increase when large apertures are used, thus the number quoted is an estimate using modest aperture), the Trumpler designation.

Messier 41	NGC 2287	06h 47.0m	−20° 44′	January
4.5 m	⊕ 38′	70	II 3 m	

Nicely resolved in binoculars, it contains blue B-type giant stars as well as several K-type giants. Current research indicates that the cluster is about 100 million years old and occupies a volume of space 80 light years in diameter.

Caldwell 64	NGC 2362	07h 18.8m	−24° 57′	January
4.1 m	⊕ 8′	60	I 3 p n	

Using small binoculars the glare from t CMa trends to overwhelm the majority of stars, although it itself is a nice star, with two bluish companion stars (recent research indicates that the star is a quadruple system). It is believed to be very young—only a couple of million years old—and thus has the distinction of being the youngest cluster in our galaxy. Contains O- and B-type giant stars.

Messier 48	NGC 2548	08h 13.8m	−05° 48′	January
5.8 m	⊕ 55′	80	I 3 r	

Located in a rather empty part of the constellation Hydra, this is believed to be the missing Messier object. Many amateurs often find the cluster difficult to locate for the reason mentioned above, but also for the fact that within a few degrees of M48 is another nameless, but brighter, cluster of stars which is often mistakenly identified as M48. Some observers claim that this nameless group of stars is in fact the correct missing Messier object, and not the one that now bears the name.

Messier 44	NGC 2632	08 h 40.1 m	+19° 59′	January
3.1 m	⊕ 95′	60	II 2 m	

This is one of the largest and brightest open clusters from the viewpoint of an observer, it is a famous cluster, called Praesepe (the Manger) or the Beehive. An old cluster, about 700 million years, distance 500 light years, with the same space motion and velocity as the Hyades, which suggests a common origin for the two clusters. A nice triple star, Burnham 584, is located within M44, located just south of the cluster's center. A unique Messier object in that it is brighter than the stars of the constellation within which it resides.

Caldwell 54	NGC 2506	08h 00.2m	−10° 47′	January
7.6 m	⊕ 7′	100	I 2 m	

A nice rich and concentrated cluster best seen with a telescope, but one that is often overlooked owing to its faintness even though it is just visible in binoculars. Includes many eleventh- and twelfth-magnitude stars. It is a very old cluster, around 2 billion years, and contains several *blue stragglers*. These are old stars that nevertheless have the spectrum signatures of young stars. This paradox was solved when research indicated that the young-looking stars are the result of a merger of two old stars.

Messier 67	NGC 2682	08h 50.4m	+11° 49′	February
6.9 m	⊕ 30′	200	II 2 m	

Often overlooked owing to its proximity to M44, it is nevertheless very pleasing. However, the stars it is composed of are faint ones, so in binoculars it will be unresolved, and seen as a faint misty glow. At a distance of 2500 light years, it is believed to be very old, possibly 3.2 billion years, and thus has had time to move from the galactic plane, the usual abode of open clusters, to a distance of about 1600 light years off the plane.

Caldwell 76	NGC 6231	16h 54.0m	−41° 48′	June
2.6 m	⊕ 14′	100	I 3 m	

A truly superb cluster located in an awe-inspiring region of the sky. Brighter by 2.5 magnitudes than its northern cousins, the double cluster in Perseus. The cluster is full of spectacular stars: very hot and luminous O-type and B0-type giants and supergiants, a couple of Wolf-Rayet stars and ξ^{-1} Scorpii, which is a B1.5 Ia extreme supergiant star with a luminosity nearly 280,000 times that of the Sun! The cluster is thought to be a member of the stellar association Sco OB1, with an estimated age of three million years. A wonderful object in binoculars and telescopes, the cluster contains many blue, orange and yellow stars. It lies between μ^{1+2} Scorpii and ξ^{-1} Scorpii, an area rich in spectacular views.

Trumpler 24	Harvard 12	16h 57.0m	−40° 40′	June
8.6 m	⊕ 60′	100	IV 2 p n	

A loose and scattered cluster, set against the backdrop of the Milky Way. It is, along with nearby Collinder 316, the core of the Scorpius OB1 stellar association.[2]

Messier 7	NGC 6475	17h 53.9m	−34° 49′	June
3.3 m	⊕ 80′	80	I 3 p	

This is an enormous and spectacular cluster that presents a fine spectacle in binoculars and telescopes, containing over 80 blue-white and pale yellow stars. It is only just over 800 light years away but is over 200 million years old. Many of the stars are around sixth and seventh magnitude, and thus should be resolvable with the naked eye.

Messier 24	–	18h 16.5m	−18° 50′	June
2.5 m	⊕ 95′x 35′		–	

This is the Small Sagittarius Star Cloud, visible to the naked eye on clear nights, and nearly four times the angular size of the Moon. The cluster is in fact part of the Norma spiral arm of our Galaxy, located about 15,000 light years from us. The faint background glow from innumerable unresolved stars is a backdrop to a breathtaking display of sixth- to tenth-magnitude stars. It also includes several dark nebulae, which adds to the three-dimensional impression. Many regard the cluster as a showpiece of the sky.

Messier 16	NGC 6611	18h 18.8m	−13° 47′	June
6.0 m	⊕ 22′	50	II 3 m n	

---

[2] Stellar associations are discussed in the next section.

It is about 7000 light years away, located in the Sagittarius-Carina spiral arm of the Milky way. The hot O-type stars provide the energy for the Eagle Nebula, within which the cluster is embedded. A very young cluster of only 800,000 years, with a few at 50,000 years old.

Messier 25	IC 4725	18h 31.6m	−19° 15′	June
4.6 m	⊕ 32′	40	I 3 m	

Unique for two reasons. It is the only Messier object referenced in the *Index Catalogue* (IC), and it is one of the few clusters to contain a Cepheid-type variable star—U Sagittarii. The star displays a magnitude change from 6.3 to 7.1 over a period of 6 days and 18 h.

Messier 11	NGC 6705	18h 51.1m	−06° 16′	July
5.8 m	⊕ 13′	200	I 2 r	

Also known as the Wild Duck Cluster, this is a gem of an object. Although it is visible with binoculars as a small, tightly compact group reminiscent of a globular cluster, they do not do it justice. With telescopes, its full majesty becomes apparent. Containing many hundreds of stars, it is a very impressive cluster. It takes high magnification well, where many more of its 700 members become visible. At the top of the cluster is a glorious pale yellow-tinted star.

–	IC 1396	21h 39.1m	+57° 30′	August
3.7 m	⊕ 50′	40	II m n	

Although a telescope of at least 20 cm is needed to truly appreciate this cluster, it is nevertheless worth searching out. It lies south of Herschel's Garnet Star and is rich but compressed. What makes this so special, however, is that it is cocooned within a very large and bright nebula.

Caldwell 13	NGC 457	01h 19.1m	+58° 20	October
6.4 m	⊕ 13′	80	I 3 r	

This cluster can be considered one of the finest in Cassiopeia. Easily seen in binoculars as two southward-arcing chains of stars, surrounded by many fainter components. The gorgeous blue and yellow double φ Cass and a lovely red star, HD 7902, lie within the cluster. Located at a distance of about 8000 light years, this young cluster is located within the Perseus spiral arm of our galaxy.

Caldwell 14	NGC 869	02h 19.0m	+57° 09′	October
5.3 m	⊕ 29′	200	I 3 r	

	NGC 884	02h 22.4m	+57° 07′	
6.1 m	⊕ 29′	115	II 2 p	

The famous Double Cluster in Perseus should be on every amateur's observing schedule and is a highlight of the northern hemisphere winter sky. Strangely, never cataloged by Messier even though it visible to the naked eye, but it is best seen using a low-power, wide-field optical system. But whatever system is used, the views are marvelous. NGC 869 has around 200 members, while NGC 884 has about 150. Both are composed of A-type and B-type supergiant stars with many nice red giant stars. However, the systems are dissimilar; NGC 869 is 5.6 million years old (at a distance of 7200 light years), whereas NGC 884 is younger at 3.2 million (at a distance of 7500 light years). But be advised that in astrophysics, especially distance and age determination, there are very large errors! Also, it was found that nearly half the stars are variables of the type Be, indicating that they are young stars with possible circumstellar discs of dust. Both are part of the Perseus OB1 Association[3] from; spend a long time observing both of them and the background star fields.

Messier 45	Melotte 22	03h 47.0 m	+24° 07′	November
1.2 m	⊕ 110′	100	I 3 r	

Without a doubt the sky's premier star cluster. The Seven Sisters, or Pleiades, is beautiful however you observe it—naked-eye, through binoculars or with a telescope. To see all the members at one go will require binoculars or a rich-field telescope. Consisting of over 100 stars, spanning an area four times that of the full Moon, it will never cease to amaze. It is often stated that from an urban location 6–7 stars may be glimpsed with the naked eye. However, it may come as a surprise to many of you that it has 10 stars brighter than sixth magnitude, and that seasoned amateurs with perfect conditions have reported 18 being visible with the naked eye. It lies at a distance of 410 light years, is about 20 million years old (although some report it as 70 million) and is the fourth-nearest cluster to us. It contains many stunning blue and white B-type giants.

Caldwell 41	Melotte 25	04h 27.0m	+16° 00′	November
0.5 m	⊕ 330′	40	II 3 m	

Also known as the Hyades. The nearest cluster after the Ursa Major Moving Stream, lying at a distance of 151 light years, with an age of about

---

[3] See later in this chapter for a discussion on stellar associations.

625 million years. Even though the cluster is widely dispersed both in space and over the sky, it nevertheless is gravitationally bound, with the more massive stars lying at its center. Best seen with binoculars due to the large extent of the cluster—over 5½°. Hundreds of stars are visible, including the fine orange giant stars γ, δ, ε and θ^{-1} Tauri. Aldebaran, the lovely orange K-type giant star, is not a true member of the cluster, but it is a foreground star only 70 light years away. Visible even from light-polluted urban areas—a rarity!

Collinder 69	–	05h 35.1m	+09° 56′	December
2.8 m	⊕ 65′	20	II 3 p n	

This cluster surrounds the third-magnitude stars λ Orionis and includes φ^{-1} and φ^{-2} Orionis, both fourth-magnitude. Encircling the cluster is the very faint emission nebula Sharpless 2-264, only visible using averted vision and an OIII filter with extremely dark skies. Perfect for binoculars.

Messier 37	NGC 2099	05h 52.4m	+32° 33′	December
5.6 m	⊕ 20′	150	II 1 r	

The finest open cluster in Auriga can be likened to a sprinkling of stardust, and some observers liken it to a scattering of gold dust. Contains many A-type stars and several red giants. In small telescopes using a low magnification it can appear as a globular cluster. The central star is colored a lovely deep red, although several observers report it as a much paler red, which may indicate that it is a variable star.

Collinder 81	NGC 2158	06h 07.5m	+24° 06′	December
8.6 m	⊕ 5′	70	II 3 r	

Lying at a distance of 16,000 light years, this is one of the most distant clusters visible using small telescopes, and lies at the edge of the galaxy. It needs a 20-cm telescope to be resolved, and even then only a few stars will be visible against a background glow. It is a very tight, compact grouping of stars and something of an astronomical problem. Some astronomers class it as intermediate object lying between an open cluster and a globular cluster, and it is believed to be about 800 million years old, making it very old as open clusters go.

Thought Question 7.3

An astronomer says he has observed a small compact open cluster that he believes is several billion years old. Is he correct in his assumption?

7.3. Stellar Associations and Streams

There exists another grouping of stars that is much more ephemeral and spread over a very large region of the sky, and although not strictly associated with star formation, they are, however, an integral part of star evolution. Furthermore, as this is a book dealing with both the evolution *and* observational properties of stars, it is probably wise to mention it here.

A stellar association is a loosely bound group of very young stars that may still be swathed in the dust and gas cloud within which they formed, and star formation may still be occurring within the cloud. Where they differ from open clusters is in their enormity, covering both a sizable angular area of the night sky and at the same time encompassing a comparably large volume in space. As an illustration of this huge size, the Scorpius-Centaurus Association is about 700 by 760 light years in extent, and it covers about 80°. There are three types of stellar associations:

1. *OB associations*, containing very luminous O- and B-type main-sequence, giant and supergiant stars.
2. *B associations*, containing only B-type main sequence and giant stars but with an absence of O-type stars. These associations are just older versions of the OB association, and thus the faster-evolving O-type stars have been lost to the group as supernovae.
3. *T associations*, which are groupings of T Tauri-type stars. These are irregular variable stars that are still contracting and evolving toward being A-, F- and G-type main sequence stars. As they are still in their infancy, more often than not they will be shrouded in dark dust clouds, and those that are visible will be embedded in small reflection and emission nebulae (see Chap. 5 —The ISM and Protostars).

The OB associations are truly enormous objects, often covering many hundreds of light years. This is a consequence of the fact that massive O- and B-type stars can only be formed within the huge giant molecular clouds that are themselves hundreds of light years across. On the other hand, the T associations are much smaller affairs, perhaps only a few light years in diameter. In some cases, the T association is itself located within or near an OB association.

The lifetime of an association is comparatively short. The very luminous O-type stars are soon lost to the group as supernovae, and, as usual, the ever-pervasive gravitational effects of the galaxy soon disrupt the association. The coherence and identity of the group can only exist for as long as the brighter components stay in the same general area of a spiral arm, as well as having a similar space motion through the galaxy.

As time passes, the B-type stars will disappear through stellar evolution, and the remaining A-type and later stars will now be spread over an enormous volume of space, with the only common factor among them their motion through space. At that point, the association is called a *stellar stream*.

An example of such a stream and one that often surprises the amateur (it did the author!) is the Ursa Major Stream. This is an enormous group of stars, with the five central stars of Ursa Major (The Plough) being its most concentrated and brightest members. The stream is also known as The Sirius Supercluster, after its brightest member. The Sun actually lies within this stream (more information about this fascinating stream can be found below). Here now are some brightest solar associations and streams.

The Orion Association	1600 light years

This association includes most of the stars in the constellation down to 3.5 magnitude, except for Orionis and 3 Orionis. Also included are several fourth, fifth and sixth magnitude stars. The wonderful nebula M42 is also part of this spectacular association. Several other nebulae[4] (including dark, reflection and emission nebulae) are all located within a vast Giant Molecular Cloud, which is the birthplace of all the O-and B-type supergiant, giant and main sequence stars in Orion. The association is believed to be 800 light years across and 1000 light years deep. By looking at this association, you are in fact looking deep into our own spiral arm, that, incidentally is called the Cygnus-Carina arm.

The Scorpius-Centaurus Association	550 light years

A much older but closer association than the Orion association. It includes most of the stars of first, second and third magnitude in Scorpius down through Lupus and Centaurus to Crux. Classed as a B-type association because it lacks O-type stars, its angular size on the sky is around 80°. It is estimated to be 750 × 300 light years in size, and 400 light years deep, with the center of the association midway between α Lupi and ζ Centauri. Its elongated shape is thought to be the result of rotational stresses induced by its rotation around the galactic center. Bright stars in this association include θ Ophiuchi, β, ν, δ and σ Scorpii, α, γ Lupi, ε, δ, μ and Centauri and β Crucis.

The Zeta Persei Association	1300 light years

---

[4] These nebulae were described in Chap. 5.

Also known as Per OB2, this association includes ζ and ξ Persei, as well as 40, 42 and o Persei. The California Nebula, NGC 1499, is also within this association.

The Ursa Major Stream	75 light years

Mentioned earlier in this section, this stream includes the five central stars of the Plough. It is spread over a vast area of the sky, approximately 24°, and is around 20 × 30 light years in extent. It includes as members, the Sun, Sirius [α Canis Majoris], α Coronae Borealis, δ Leonis, β Eridani, δ Aquarii, and β Serpentis. Due to the predominance of A1 and A0 stars within the association, its age has been estimated at 300 million years.

The Hyades Stream	

There is some evidence (although it is not fully agreed upon), that the Ursa Major stream is itself located within a much older and larger stream. This older component includes M44, Praesepe in Cancer, and the Hyades in Taurus, with these two open clusters being the core of a very large, but loose grouping of stars. Included within this are Capella [α Aurigae], α Canum Venaticorum,[5] δ Cassiopeiae and λ Ursae Majoris. The stream extends to over 200 light years beyond the Hyades star cluster, and 300 light years behind the Sun, and thus, the Sun is believed to lie within this stream.[6]

The Alpha Persei Stream	540 light years

Also known as Melotte 20, this is a group of about 100 stars including, α Persei, ψ Persei, 29, and 34 Persei. The stars δ and ε Persei are believed to be amongst its most outlying members, as they also share the same space motion as the main groups of stars. The inner region of the stream is measured to be over 33 light years in length: the distance between 29 to ψ Persei.

Thought Question 7.4

Are you likely to a find a very old OB association?

---

[5] Capella and a Canum Venaticorum are also thought to be members of the even larger Taurus Stream, which has a motion through space similar to the Hyades, and thus may be related.

[6] The bright stars that extend from Perseus, Taurus and Orion, and down to Centaurus and Scorpius, including the Orion and Scorpius-Centaurus associations, lie at an angle of about 1 °5 to the Milky way, and thus to the equatorial plane of the Galaxy. This group or band of stars is often called *Gould's Belt*.

We will leave our discussion of Globular Clusters till a later chapter, on the evolution of stars, as it makes more sense doing this, once the topic of stellar evolution has been discussed.

Before we look at the many types of stars on the main sequence, it will be helpful (and indeed necessary) for us to look at the nearest star to us—the Sun. After all, astronomers have been studying the closest star to us for a long time now, and so we have a good idea of what's going on.[7]

In looking at the Sun in detail, we will be able to see how energy is produced in the core, and how this energy is transported to the surface, and then to us on the Earth! We can then look at other stars and compare and contrast them with what we know about the Sun. Onward then to the next chapter on the Sun!

Problems

1. Look at Fig. 7.3. Determine the approximate difference in luminosity of the coolest star near the main sequence star on, to that of the most luminous star.
2. Look at Fig. 7.4. Determine the approximate difference in luminosity of the hottest star on the main sequence, to that of the red giant star.

---

[7] This is, of course, an exaggeration as it is only recently that astronomers have solved the problem of the solar neutrino.

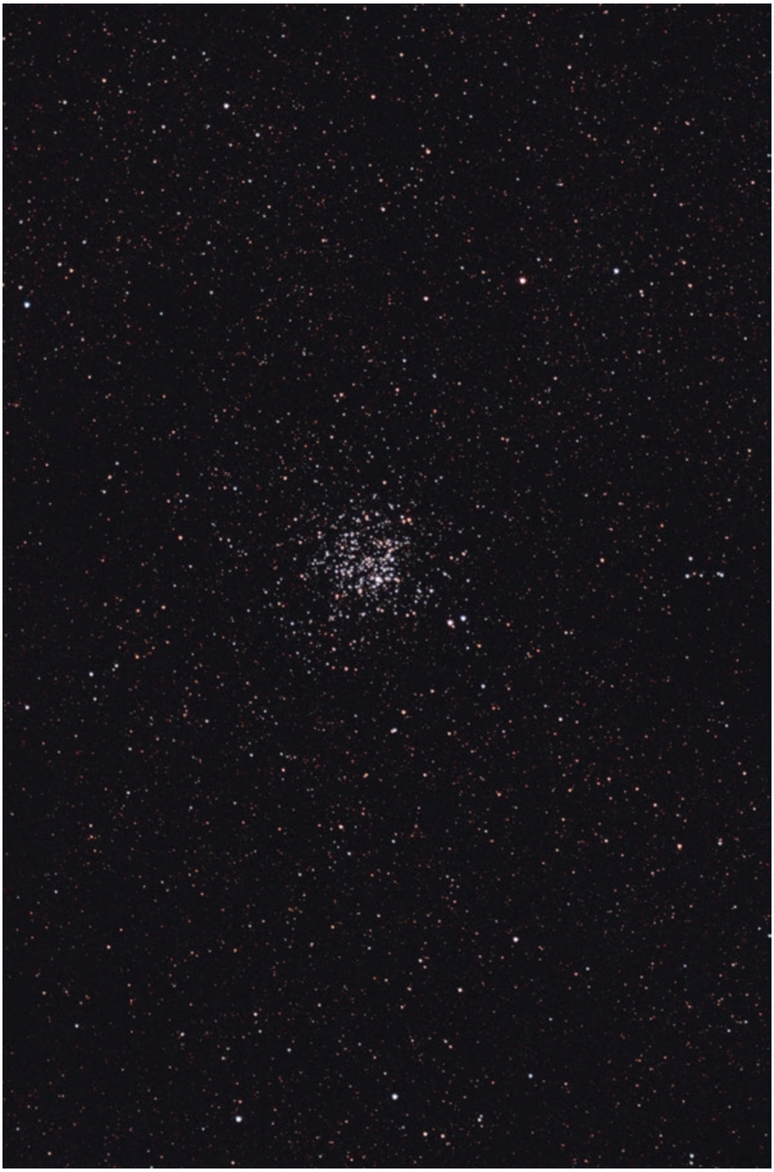


Fig. 7.3. The Wild Duck cluster, Messier 11

Fig. 7.4. Messier 52 & Bubble Nebula

Chapter 8

The Sun, Our Nearest Star

Due to the advances not only in astronomy but in computing as well, astronomers have been able to determine the conditions inside the Sun by solving several equations that describe how the temperature, mass, luminosity and pressure change with distance from the center of the Sun. To solve them, we need to know the mechanisms by which energy is transported throughout the Sun, either with radiation or convection, as well as the chemical composition of the Sun and the rate of energy production at any specific distance from its center.

Now, although the equations are simple to solve, computers are needed, so we will just say that the results seem to match the observations, which is always a good test for any theory.

8.1. From the Core to the Surface

The Sun's structure is shown in Fig. 8.1. The visible surface of the Sun, called the *photosphere*, has a temperature of about 5800 K, and although it may look like a well-defined surface from Earth, the gas there is less dense than Earth's atmosphere. Both the density and temperature increase steadily as we progress from the surface to the core.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_8

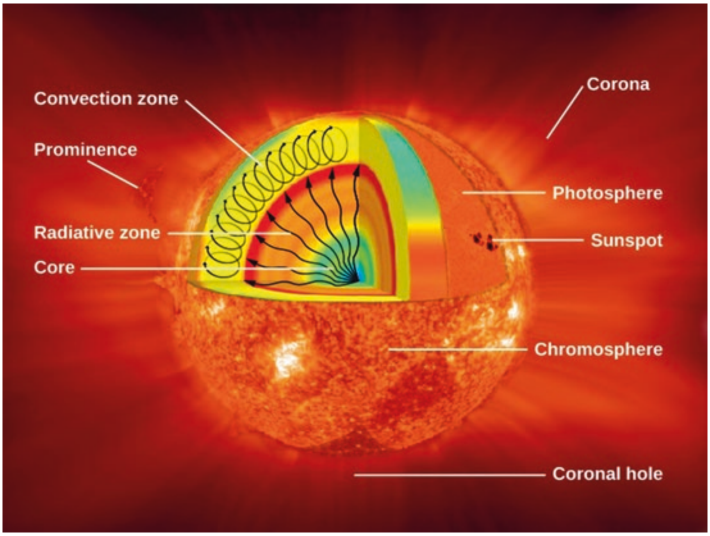


Fig. 8.1. The structure of the Sun

Beneath the photosphere is a very turbulent area called the *convection zone*, where energy generated in the core travels upward, transported by rising columns of hot gas and the falling of cool gas. This process is called *convection*. So the photosphere is in fact the top of the convection zone. Descending deeper through the convection zone, the pressure and density increase quite substantially, along with the temperature. The density there can be far greater than that of water, but remember that we are still talking about a gas, albeit one in a very strange state. A gas under these extreme conditions of temperature and/or pressure is usually called *plasma*.[1] The temperature in this region is about 2 million K, and the solar plasma absorbs the photons.

About one-third of the way down to the center, the very turbulent convection zone gives way to the more stable plasma of the *radiation zone*. Here the energy is transported outward primarily by photons of X-ray radiation. The temperature in this region is now about 10,000,000 K. At the central region, the core of the Sun, the temperature is now 15,000,000 K, and it is here that hydrogen is being transformed to helium. The pressure in this region is nearly 200 billion times the surface pressure found on Earth. The central temperature and pressure are both impressive, with the core compressed to a density of about 150,000 kg m^3, which is about 150 times the density of water. See Math Box 8.1.

---

[1] Plasma is a collection of positively charged ions and free electrons.

Math Box 8.1 The Density of the Sun

With a very simple calculation we can calculate the average density of the Sun.

The mass of the Sun is 2.0×10^{30} g, and its radius is 7×10^{10} cm (7×10^{5} km).

Thus its average density is given by

$$\rho = \frac{M_{sun}}{\left(\frac{4\pi}{3}\right) R_{sun}^{3}}$$

Using the values given above

$$\rho = \frac{2 \times 10^{33}\, g}{\left(4\pi/3\right) \times \left(7 \times 10^{10}\, cm\right)^{3}}$$

$$\rho = 1.4 \text{ g cm}^{-3}$$

For comparison, the density of water is 1 g cm^{-3}.

It may come as a surprise to some people that essentially all of the Sun's energy is produced in the inner 25% of its radius, which corresponds to about 1.5% of its volume. This is a consequence of the very acute temperature sensitivity of nuclear reactions. If we were to actually go to a point about 1/4 of the distance from the center of the Sun to its core, the temperature would have fallen to about 8000,000 K, and at this lower temperature nuclear fusion energy production would have fallen to practically zero. So, virtually no energy is produced beyond the inner 25% of the solar radius.

At the surface of the Sun, each kilogram of gas contains about 71% hydrogen, whereas in the core, the percentage of hydrogen will be much lower, around 34%. The reason for this is obvious—hydrogen has been the fuel for nuclear fusion for the past 4.6 billion years.

The total power output of the Sun, which is its luminosity, is a staggering 3.8×10^{26} J per second. This may not mean much to most of us, but if we could somehow capture all of this energy for even 1 s, it would be sufficient to meet all current energy demands for the human race for the next 50,000 years! But remember, only a tiny fraction of this reaches Earth, as it is all dispersed in all directions into space.

The current model of energy production in the Sun is that in which nuclear fusion is the generator. It is a source so efficient that the Sun will shine for 10 billion years, and as it is only 4.6 billion years old at the moment, it has a long way to go! This current model of solar energy generation means that the Sun's size will generally be stable, maintained by a balance between the competing forces of gravity pulling inward and pressure pushing outward. This balance between forces is called *hydrostatic equilibrium* (or sometimes *gravitational equilibrium*). What this means is that at any given point within the Sun, the underlying pressure supports the weight of the overlying material.

You may think that this is a simple concept, and so it is, but it maintains the integrity of the Sun as well as most stars in the universe. When one or the other of the forces gains the upper hand, however, the consequences are spectacular, as we shall see in a later chapter. Hydrostatic equilibrium in the Sun means the pressure increases with the depth; this makes the Sun extremely hot and dense in its core.

The efficiency with which the energy is transported outward by radiation is strongly influenced by the opacity of the gas through which the photons flow. The opacity describes the ability of a substance to stop the flow of photons. For instance, when the opacity is low (think of it as a clear day), photons are able to travel much greater distances between emission and re-absorption than when the opacity is high (a foggy, hazy day). If opacity is low, the transport of energy by photons is very efficient. But when the opacity is high, the efficiency is reduced, which leads to an inefficient flow of energy and a higher rate of temperature decline.

Thought Question 8.1

If the stability of the Sun relies on the balance between the competing forces of gravity pulling inward and pressure pushing outward, what do think will happen when the balance is disrupted?

8.2. The Proton-Proton Chain

To explain the Sun's energy, we need a process that involves the most abundant element in the Sun, hydrogen. The British astronomer A. S. Eddington first proposed the fusion of hydrogen into helium in 1920, although the details were not fully understood until 1940.

Hydrogen, which is the lightest element, has a nucleus consisting of just one proton. The nucleus of helium, however, has four nuclear particles—two protons and two neutrons. So four hydrogen nuclei are needed to make one helium nucleus. But we cannot expect four protons to collide and instantly make a helium nucleus. This is so unlikely to happen that it has probably never happened before—not even once—in the history of the universe. What happens instead is a series of reactions involving two reactions at a time. This series of reactions is called the *proton-proton chain*.[2]

The reactions begin with an interaction between two protons that must come within 10^{-15} m of each other for a nuclear reaction to occur. There is a slight problem, however, as protons are positively charged and so, just like magnets, repel each other. The result of this mutual repulsion is that most collisions between protons do not result in any reaction; instead, the two protons deflect each other and move apart. At room temperature, there is absolutely no possibility that two protons would collide with enough energy to get close enough to instigate a reaction.

So, for any reactions to occur we need conditions that will allow protons to move at very high velocities; these conditions exist in the center of the Sun (and, of course, stars). At the core of the Sun, the temperature is 15 million K, and a typical proton will be traveling at about 1 million kilometers per hour. But even at this fantastic speed, the likelihood of a reaction occurring is very small. If we could watch a single proton to see how long it would take before it eventually reacted with another proton in nuclear fusion, we would be waiting about 5 billion years! The important point here is that there are so many protons in the Sun's core; every second 10^{34} of them can undergo a reaction.

The sequence of steps in the proton-proton chain is shown in Fig. 8.2.

Step 1: Two protons fuse to form a nucleus consisting of one proton and one neutron. This is the isotope of hydrogen called *deuterium* (^{2}H). The other products formed are a positively charged electron, called a *positron* (β^+), and a *neutrino* (ν), a minuscule particle with a tiny mass. The positron doesn't last long, however; it soon meets up with an ordinary electron, and the result is the creation of two gamma rays that are rapidly absorbed by the surrounding gas, which consequently heats up. What happens to the neutrino? We shall discuss that later.

Step 2: The deuterium now fuses with a proton, producing a helium nucleus (^{3}He) and gamma rays. The ^{3}He nucleus consists of two protons and one neutron, whereas an ordinary helium nucleus has two protons and two neutrons. This step of producing ^{3}He from a deuteron occurs very rapidly, so that a typical deuteron in the core of the Sun will survive for only 4 s before reacting with a proton.

---

[2] In stars that are more massive than the Sun, the fusion of helium occurs via a different series of reactions called the CNO cycle. We shall discuss this later.

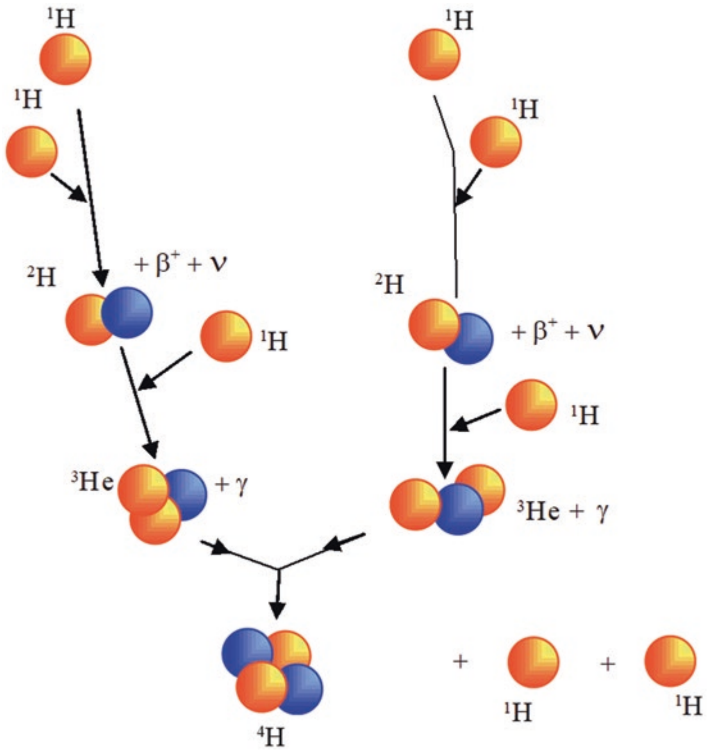


Fig. 8.2. The reactions of the proton-proton chain. The symbol ν indicates a neutrino, the β^+ indicates positrons, and γ indicates photons. The *orange sphere* represents protons, and *blue* indicates neutrons

Step 3: Usually, the final reaction in the proton-proton chain requires the addition of another neutron to the ^{3}He nuclei, thereby making normal ^{4}He. This final step can proceed in several ways, with the most common involving a collision of two ^{3}He nuclei. Each of these ^{3}He nuclei resulted from a prior, separate occurrence of Step 2 somewhere else in the core. The final result is a normal ^{4}He nucleus and two protons. On average, a ^{3}He nucleus must wait for 4 million years before it participates in this reaction.

The net result of the chain of reaction is:

$$6\,^1\mathrm{H} \rightarrow {}^4\,\mathrm{He} + 2\,^1\mathrm{H} + 26.72\ MeV$$

Although it takes six protons to make one helium nucleus, there is a net loss of only four protons because two are regenerated in the final step. Because the six protons are more massive than the two protons and a helium nucleus, mass is lost in the proton-proton chain and converted to energy.

Each resulting ^{4}He nucleus has a mass that is slightly less than the combined mass of the four protons that created it (by about 0.7%). The energy produced by a single proton-proton chain reaction is 26.27 MeV, and although the units are unfamiliar to you, this converts to about one ten-millionth of the amount of energy needed to lift a drop of water. As you can see, this is not a lot of energy; overall, however, the Sun converts about 600 million tons of hydrogen into 596 million tons of helium every second. See Math Box 8.2.

The remaining 4 million tons of matter have not just disappeared. They are converted to energy in accordance with the famous equation formulated by Einstein: $E = mc^2$. The neutrinos carry off about 2% of this energy and rarely interact with matter and so pass straight out into space. The remaining energy emerges as kinetic energy of the nuclei and as radiative energy of the gamma rays.

Math Box 8.2 Mass and Energy Conversion in the Sun

It is relatively simple to calculate how much mass the Sun loses through nuclear fusion. First let's look at the input and output masses of the proton-proton chain A single proton has a mass of 1.6726×10^{-27} kg, so four protons have a mass of 6.693×10^{-27} kg.

A single ^{4}He nucleus has a mass of only 6.645×10^{-27} kg, which is slightly less than the mass of four protons. In other words:

$$6.69 \times 10^{-27}\ \text{kg} - 6.643 \times 10^{-27}\ \text{kg} = 4.8 \times 10^{-29}\ \text{kg}$$

This is only 0.7% or 0.007 of the original mass, so if 1 kg of hydrogen fuses, the resulting helium weighs 993 g, and 7 g of mass are turned into energy. Math Box 8.2 (continued). To calculate the amount of energy for a single reaction, let's use Einstein's famous equation, E = mc^2.

$$E = mc^2 = \left(4.8 \times 10^{-29}\ \text{kg}\right)\left(3.0 \times 10^{8}\ \text{ms}^{-1}\right)^2 = 4.3 \times 10^{-12}\ \text{joules}$$

This is the tiny amount of energy created by the formation of one helium atom. It is so small, it would only power a 10-W light bulb for half a trillionth of a second. Remember that about 1×10^{-11} J is carried away by neutrinos.

Let's now see how much total energy is produced when hydrogen is converted to helium every second. We know that only 0.7% of the mass of hydrogen is fused for every kilogram:

$$E = mc^2 = (0.7\ \text{kg})\left(3.0\times10^8\ \text{ms}^{-1}\right)^2 = 6.3\times10^{14}\ \text{joules}$$

However, the Sun's luminosity is 3.9 × 10^{26} J per second, thus hydrogen must be consumed at a rate:

$$\frac{\left(3.9\times10^{26}\,\text{joules per second}\right)}{\left(6.3\times10^{14}\,\text{joules per second}\right)}$$
$$= 6\times10^{11}\,\text{kilograms per second}$$

So the Sun fuses 600 million metric tons of hydrogen each second, with 596 tons fused into helium and the remaining 4 million becoming energy.

8.3. Energy Transport from the Core to the Surface

Energy produced in the central region of the Sun flows outward towards the surface. If the Sun were transparent, the photons, or gamma rays, emitted by the extremely hot gases in the core would travel straight out at the speed of light 2 s after being emitted. The Sun's gases, however, are not very transparent, and so a typical photon only travels about 10^{-6} m before it is re-absorbed. In being absorbed, it heats up the surrounding gases, which in turn emit photons, which are then subsequently re-absorbed. The emitted photon will not necessarily be emitted in an outward direction, but rather a totally random direction, which means that at least 10^{25} absorptions and reemissions occur before energy reaches the surface. This slow, outward migration of photons is often called a *random walk*.

Let's now look in detail how energy produced in the central region of the Sun flows outward towards the surface, using two methods. See Fig. 8.3.

The first process occurs in the *Radiative Zone*. This layer, or zone surrounds the core. If the Sun were transparent, the photons, or gamma rays, emitted by the extremely hot gases in the core would travel straight out at the speed of light 2 seconds after being emitted. The Sun's gases, however, are not very transparent, and so a typical photon only travels about 10^{-6} m before it is re-absorbed. In being absorbed, it heats up the surrounding gases, which in turn emit photons, which are then subsequently re-absorbed. The emitted photon will not necessarily be emitted in an outward direction, but rather a totally random direction, which means that at least 10^{25} absorptions and reemissions occur before energy reaches the next layer. This slow, outward migration of photons is often called a *random walk*. The process itself is called *radiative energy transport*, and thus gives the layer its name. Furthermore, even though

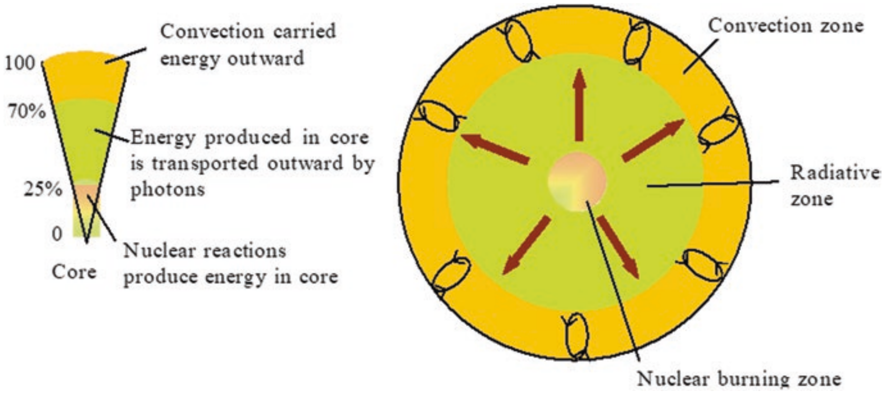


Fig. 8.3. Energy transport in the Sun

the initial photon created in the proton-proton chain is a gamma ray, the interactions themselves the cause the photons to lose energy. For instance, the gamma ray could be changed into two lower energy X-rays, and these X-rays themselves could be converted into many ultraviolet photons, and so it continues. One calculation suggests that a single gamma ray formed in the pp-chain, can end up as over 2000 lower energy photons.

The next layer is the *Convective Zone*. This can be thought of as the penultimate layer[3] of the solar interior. It is a relatively thick layer, approximately 200,000 kilometres that transports energy from the edge of the radiative zone to the surface through giant convection cells, similar to a pot of boiling water. The plasma at the top of the radiative zone/bottom of the convective zone is extremely hot, and it bubbles to the surface where it loses its heat to space. The plasma cools and sinks back to the bottom of the convective zone, where the process repeats itself.

The final layer in our discussion of energy transport is called the *Photospher*e. When one looks at the Sun (but do not ever do this without the proper safety precautions), what we are seeing is the surface of the Sun, and this is the photosphere. Solar astronomers have established that the solar atmosphere changes from being almost perfectly transparent to completely opaque in a distance of just over 400 kilometres, this thin region is the photosphere.[4] When astronomers speak of the "diameter" of the Sun, they mean the size of the region surrounded by the photosphere.[5]

---

[3] To be accurate, the photosphere is really just the top part of the convective zone, but some astronomers regard the photosphere as a completely separate layer.

[4] From the Greek for "light sphere."

[5] Note that the photosphere of the Sun is not very dense compared to the air you are breathing now. In the photosphere, the pressure is less than 10% of Earth's pressure, and the density is about one ten-thousandth of Earth's atmospheric density, both at sea level.

You may think that the surface, or photosphere, would have a smooth appearance, but this is not the case. Just as a pot of boiling water has as bubbling and seething appearance, so does the photosphere. There is a uniformly mottled network of bright patches surrounded by thin, darker outlines called granules, and this patter is named *Solar Granulation*. The lighter areas are the tops of the columns of hot gas rising from the convective zone, and the dark outlines are the cooler gases sinking back down.[6] This is represented in Fig. 8.1. So in fact, the whole the surface of the sun is ever changing.

The overall energy transport process means that there is a considerable time delay before energy produced at the core reaches the surface. On average, about 170,000 years will pass before energy created at the core eventually reaches the surface.[7] Furthermore, the energy produced in 1 second does not erupt from the surface all in one go; it appears that it is radiated from the surface over a period of more than 100,000 years. Some energy appears in about 120,000 years, whilst other energy takes 220,000 years. But the bulk of it is emitted after 170,000 years.

This tells us two things about the Sun. First, when we observe the light emitted by the Sun, we learn nothing about what is going on in the core *at that moment*. All we can say is that energy was created in the core many thousands of years ago, and the photons we see from the sun are not the original photons created in the core.

The second point is that if energy generation were to suddenly cease in the core for, say, a day, or even a hundred years, we would not notice it because by the time the energy flowed to the surface, it would have been averaged out over more than 100,000 years. This implies that the brightness of the Sun is very insensitive to changes in the energy production rate.

Thought Question 8.2

If the Sun was to suddenly stop making energy, and by some unknown mechanism go completely dark, what observational consequence would be apparent to the amateur astronomer? Ignore the fact that you cannot observe the Sun at all.

We will now look at features that occur on the surface, or photosphere of the Sun, such as *Sunspots*, *Prominences*, *Flares*, and *Coronal Holes*.

---

[6] Using the correct equipment, even modest telescopes will show the granulation as a subtle mottling. But always use the appropriate safety measures. Always.

[7] This means that averaged over the distance from the core to the surface, a "photon" travels about 0.5 m per hour, or about 20 times slower than a snail.

8.4. Sunspots

Sunspots have been telescopically observed for about 400 years,[8] or since we have had telescopes. They are temporary, dark, small patches surrounded by a lighter area, often occurring in pairs, each spot of a pair having opposite magnetic polarity; one will be a magnetic south pole and the other a magnetic north pole. As they are located on the surface or photosphere, by measuring how long it takes a sunspot to make a complete rotation of the Sun allows us to determine the rotation period of the Sun.[9] When this is done, it gives a surprising answer, as the rotation period is different for differing latitudes. For instance, the rotation period is about 25 days at the equator, around 28 days at latitude 40°, and about 36 days at latitude 80°. This is called *differential rotation.*[10] The cause of this phenomena is that the Sun is a gas and behaves like a fluid in that it does not have to rotate rigidly, the way a solid body like a terrestrial planet does.

Returning to our discussion of sunspots, it was a long time before anyone came up with a reasonable explanation to what they are. The answer is simple, as are most things when finally understood. They are just areas of a lower temperature than that of the surrounding photosphere. The cause of the lower temperature is the convection process below the photosphere that we discussed earlier. The temperature of the central darker area, known as the umbra, is roughly 3000–4500 K (2,700–4200 °C), in contrast to the lighter peripheral area, known as the penumbra, at about 5780 K (5,500 °C), this difference in temperature leaves sunspots clearly visible as dark spots, that, on some occasions are visible even to the naked eye. This occurs because the luminance (essentially "brightness" in visible light) of a heated black body (approximated by the photosphere) at these temperatures varies greatly with temperature. If we could take a single sunspot from the photosphere, it would shine brighter than the full moon, with a crimson-orange color.

How sunspots are made is still a topic of active research, but it appears that sunspots are the visible equivalents of magnetic fields deep in the convective zone that get "wound up" by differential rotation. When the stress on the mag-

---

[8] The earliest record of sunspots is found in the Chinese *I Ching*, completed before 800 BC. The earliest record of a deliberate sunspot observation also comes from China, and dates to 364 BC, based on comments by the astronomer Gan De. By 28 BC, Chinese astronomers were regularly recording sunspots. The first clear mention of a sunspot in Western literature is dated around 300 BC, by the ancient Greek scholar Theophrastus, student of Plato and Aristotle. The first drawings of sunspots were made by the English monk John of Worcester in December 1128.

[9] In 1612, Galileo showed that the Sun rotates on its axis with a rotation period of approximately 1 month. He also lost vision in one eye from observing the Sun without due precautions.

[10] Jupiter and Saturn also show differential rotation.

netic field reaches a certain limit, a loop of the field may project through the photosphere.[11] At this location, convection is inhibited and the energy from the Sun's interior is decreased, and thus, the surface temperature, causing the surface area through which the magnetic field passes to look dark against the bright background of the photosphere. As the loop exits and enters the photosphere at two points, we therefore get two sunspots.

The lifetime of a sunspot can last from a few days to a few months, depending on its size, although groups of sunspots and their active regions tend to last longer, weeks or months, but they all eventually decay and disappear from view. They have surface areas in the range of 16 km (10 mi) to 160,000 km (100,000 mi), and are, contrary to popular belief, depressions in the surface.[12]

8.4.1 The Solar Cycle

Over the many years that sunspots had been observed it was noticed that the number of sunspots visible on the surface would increase, and then over a period, the number would decrease. This phenomenon would repeat on a more-or-less regular manner. In this manner the Solar Cycle was discovered. To go from the start of a cycle, when there are none, or hardly any sunspots on the surface, to a time when there are a lot, and then back to a time when there are few or none, takes about, on average 11 years. Observations over the past 400 years have shown this to be a regular activity (see Fig. 8.2), with a few exceptions that are discussed later. The time when there are a lot of sunspots is referred to as *Solar Maximum*, and when there are few, *Solar Minimum* (Fig. 8.4).

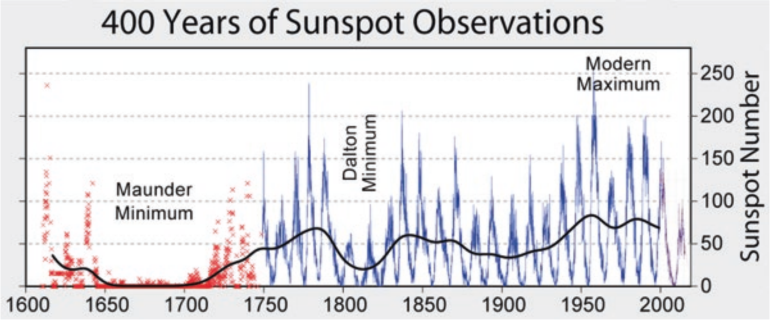


Fig. 8.4. The solar cycle. (Courtesy of NASA/ARC)

---

[11] Just like the shape of a horseshoe magnet.

[12] This was first noticed by the Scottish astronomer Alexander Wilson during Solar Cycle 2 in 1769. He noticed that the shape of a sunspot noticeably flattened as it approached the Sun's limb due to solar rotation. However, doubts have been raised about this explanation. Thus, research is still ongoing.

Note that when we talk about the number of sunspots, we are referring to the number of new sunspots, as, if you recall, the lifetime of a sunspot can be measured in day, weeks, and months, so more and more (or fewer and fewer) new sunspots are being seen.

Also, sunspots do not appear randomly distributed at all latitudes on the surface of the Sun. If we start at the beginning of the cycle, then new sunspots will appear at latitudes of around ±40°. Then as the cycle progress the new sunspots appear closer to the equator,[13] until they average 15° at solar maximum. The average latitude of sunspots then continues to drift lower, down to around 7°. Then while the old sunspot cycle diminishes, the new cycle starts with sunspots appearing at high latitudes.[14]

So, if someone was to show you a sunspot group near the equator then you could say with confidence that this was toward the end of a cycle, approaching solar minimum. If another photo showed sunspots at higher latitudes, say, ± 20–25°, you would know that the cycle was progressing toward solar maximum. This positional sunspot relationship is shown in Fig. 8.5, and is known as the Butterfly diagram, because of its obvious shape.

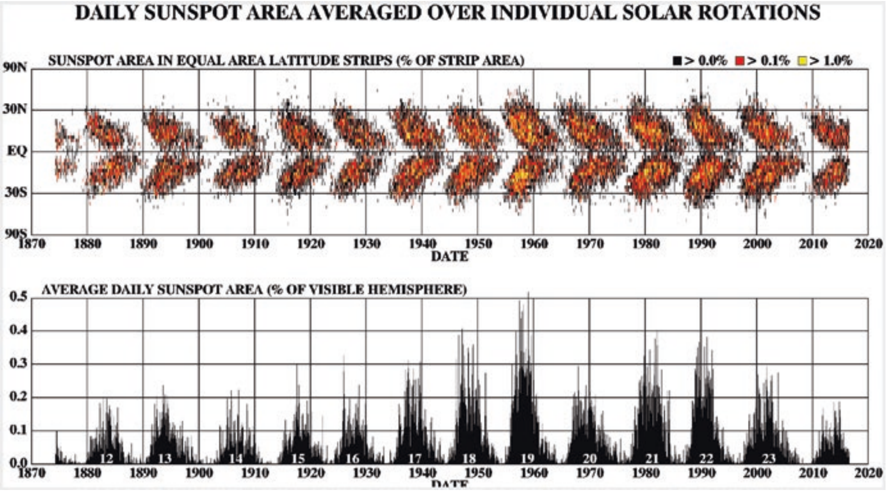


Fig. 8.5. The butterfly diagram. (Courtesy of NASA/ARC)

---

[13] Do not make the mistake of thinking that a new sunspot may appear at a high latitude, and then during its short life, moves to lower latitudes. It doesn't. It is the newer sunspot that appears at lower latitudes during the cycle.

[14] This is known as Spörer's law. It was discovered by the English astronomer Richard Christopher Carrington around 1861 and refined by the German astronomer Gustav Spörer.

Looking at Fig. 8.4, you will see the peeks of several solar maxima over hundreds of years. A few things will stand out. Firstly, the heights of the peaks differ over many cycles, meaning that at certain solar maxima, there were a lot of sunspots, whilst at other maxima, there were fewer. Also, there is a definite reduction over a period of around 30 years called the Dalton Minimum, and a striking, almost zero, absence of sunspots during a period called the Maunder Minimum.

It appears that along with the familiar 11-year cycle, there may be even longer cyclic activity. Referring to the figure note that during the period known as the modern maximum from around 1900 to 1958, the solar maxima trend of sunspot count was upwards; for the following 60 years the trend was mostly downwards. Also, the Dalton Minimum coincided with a period of lower-than-average global temperatures. During that period, there was a variation of temperature of about 1 °C in Germany. At the moment, the cause of the lower-than-average temperatures and their possible relation to the low sunspot count are not well understood, but recent research suggests that a rise in volcanism may have been responsible for the trend in cooling.[15]

Thought Question 8.3

You are shown a photo of the Sun, that shows a few sunspots at high latitudes. At what stage on the solar cycle do you think the photo was taken?

8.4.2 The Maunder Minimum

The Maunder Minimum, also known as the "prolonged sunspot minimum", is the name used for the period around 1645 to 1715 during which sunspots became exceedingly rare, as was then noted by solar observers. (Refer to Fig. 8.4).

---

[15] It is thought that the prime reason for Dalton Minimum cool temperatures was the highly explosive 1815 eruption of Mount Tambora in Indonesia, which was one of the two largest eruptions in the past 2000 years. Amazingly, there is a recent idea that volcanism may have been triggered by low levels of solar output as there is a weak but statistically significant link between decreased solar output and an increase in volcanism. Here's what the research suggests "Time series analysis of the volcanogenic acidities in a deep ice core from Greenland, covering the years 553–1972, reveals several very long periods ranging from ~80 to ~350 years and are similar to the very slow solar cycles previously detected in auroral and carbon 14 records. Solar flares are believed to cause changes in atmospheric circulation patterns that abruptly alter the earth's spin. The resulting jolt probably triggers small earthquakes which may temporarily relieve some of the stress in volcanic magma chambers, thereby weakening, postponing, or even aborting imminent large eruptions." Stothers, R.B., 1989: Volcanic eruptions and solar activity. J. Geophys. Res., 94, 17371–17381, https://doi.org/10.1029/JB094iB12p17371.

It occurred during a long period of lower-than-average known as the Little Ice Age, during which Europe and North America experienced colder than average temperatures which are likely to have been primarily caused by volcanic activity similar to that for the Dalton Minimum. However, the start of the Little Ice Age occurred well before the beginning of the Maunder Minimum, and northern-hemisphere temperatures during this time were not significantly different from the previous 80 years. This has led many scientists to believe that a decline in solar activity was not the main causal driver of the Little Ice Age.

From analysis of other data sources had to the idea that the Sun actually has not just an 11-year cycle, but many different cycles.[16] Overall, the Sun was last as active as the modern maximum over 8000 years ago.

8.4.3 The Solar Magnetic Cycle

Before leaving sunspots, there is one other topic worth mentioning. In addition to the 11-year sunspot cycle, there is also a 22-year solar magnetic cycle. It goes something like this.

Consider two sunspots formed at the beginning of the 11-year cycle, located in the northern solar hemisphere, and two similar sunspots located in the southern solar hemisphere. In the northern hemisphere, the trailing sunspot (the source of a rising magnetic field), in the direction of the sun's rotation, can have, say, a north polarity, whilst the leading sunspot (the magnetic field dives back down) will have a south polarity. In the southern solar hemisphere, the reverse situation occurs, with the trailing sunspot now having a south polarity and the leading sunspot, a north polarity. For the whole of a solar cycle, that is to say, for the 11 years, the same situation will occur in the northern hemisphere, with new sunspots having the same polarity setup, leading sunspot south polarity, and the southern hemisphere have the opposite setup, leading sunspot north polarity.

At the beginning of the next solar cycle, the situation is changes with the polarity between leading and trailing sunspots reversed. What was south now becomes north and vice-versa. Furthermore, it is not just sunspots that

---

[16] Various historical sunspot minima have been detected either directly or by the analysis of the cosmogenic isotopes; these include the Spörer Minimum (1450–1540), and less markedly the Dalton Minimum (1790–1820). In a 2012 study, sunspot minima have been detected by analysis of carbon-14 in lake sediments. The analysis suggests that there have been 18 periods of sunspot minima in the last 8000 years. Furthermore, studies indicate that the Sun currently spends up to a quarter of its time in these minima.

have this flipped polarity, but the magnetic field of the sun itself is reversed with the solar north magnetic pole becoming the south pole, and the solar south magnetic pole becoming the north pole. This large-scale reorientation of the magnetic field system is called the **Solar Magnetic Cycle**, and thus it takes two sunspot cycles, or 22 years, for the Sun to go through one complete magnetic cycle, returning to the polarity orientation we started with.

Thought Question 8.4

Can we predict solar activity, including sunspot numbers, for a future solar cycle?

8.5. Prominences, Flares, Coronal Mass Ejections, and the Solar Wind

In this final section on the Sun, we are going to cover those topics that deal with surface phenomena that are not sunspots, but before we do this, we need a small detour to talk about the solar regions that lie above the photosphere, namely the Chromosphere and the Corona.

Astronomers regard the photosphere as the surface of the Sun, even though it is made of gas and hence is not solid. Nevertheless, it is accepted as the surface, so this means that any layers of the Sun that lie above the surface, can be referred to as the solar atmosphere. It has two components, namely the Chromosphere, the layer above the photosphere, and the Corona, that lies above the Chromosphere.

The chromosphere, a relatively thin layer of about 2000 km ranges in temperature from around 4000 K at the photosphere-chromosphere boundary to around 35,000 K at its outermost limits, and with a gas density 10^{-8} times that of the atmosphere of Earth at sea level. This low density makes the chromosphere normally invisible, and it is only during a total solar eclipse, that its reddish colour can be seen. It also has a strange appearance,[17] consisting of features called spicules, that are long thin tendrils of luminous gas which have been compared to a vast field of flaming wheat

---

[17] There are also features known as a plage, found in regions near sunspots, and looks like is a bright region in the chromosphere. The term itself is poetically taken from the French word for "beach". The plage regions map closely to the bright areas (faculae) in the photosphere below, but the latter have a much smaller spatial scale, and thus accordingly, plages occurs most visibly near a sunspot region.

rising upwards from the photosphere below. The spicules climb to the top of the chromosphere before sinking back down again over about 10 minutes. Likewise, there are horizontal wisps of gas called fibrils, which last about twice as long as the spicules.

What can be considered the final layer of the Sun is called the Corona.[18] This is an area of plasma surrounding the Sun and incidentally, other stars. The corona extends millions of kilometres into space and is most famously seen during a total solar eclipse. It has an extremely high temperature of around 1–3 million K, with a gas density 10^{-12} times as dense as the photosphere, and so yields about one-millionth as much visible light. The origin of this very high temperature is somewhat of a problem as usually things tend to get cooler the further one is from a heat source, or in this case the sun's core. Contemporary ideas suggest that induction by the Sun's magnetic field and magnetohydrodynamic waves from below maybe the source of the energy responsible for heating the corona, whilst a relatively newer idea is that a new class of spicules (TYPE II) discovered in 2007, which travel faster (up to 100 km/s) and have shorter lifespans, could account for the coronal temperature as these spicules inject heated plasma into the corona. One final point is that the corona is not like, say, the Earth's atmosphere which covers the whole planet. Rather, there are patches where the corona is absent, and these are called coronal holes.

Now, we can begin our discussion of solar surface features, and discuss *Prominences*.

A prominence, also referred to as a filament when viewed from the solar disk (and seen as a darker sinuous shape against the brighter photosphere), is a large, bright, gaseous feature that extends outward from the Sun's surface, more often than not in a loop shape. Prominences are attached to the Sun's surface in the photosphere and extend outward into the solar corona. As mentioned earlier, the corona consists of plasma, that does not emit much visible light, while prominences contain much cooler plasma, similar in composition to that of the chromosphere. The prominence plasma is typically a hundred times more luminous and denser than the coronal plasma. Prominences vary considerably in size, shape, and motion and are of two main types, active and quiescent.[19] The Active prominence erupts rapidly and has a lifetime lasting of anything from several minutes to a few hours. They are associated with sunspot groups and, like these, are correlated in

---

[18] Some astronomers and textbooks assert that the Solar Wind is the final layer. I will leave it to you to decide.

[19] From an observational viewpoint, the classifications can be further subdivided. For instance, the quiescent and long-lived Class I prominence can be one of the following: hedgerow, curtain, flame, fan, arch, platform arch, cap, irregular arch, fragment, detached, Whilst the active, moving, or transient Class II, prominence can be eruptive, surge, spray flare loop. This is quite a list.

numbers and activity with the solar cycle. A quiescent prominence on the other hand tends to emerge effortlessly and wane much more slowly, with some being visible for several months. From an amateur astronomer point of view, prominences are one of the few astronomical objects where changes of say, the shape and size, can be seen over just a few hours.[20]

Moving now onto *Flares*.

A solar flare is an abrupt flash of increased brightness on the Sun, usually seen near the surface and near a sunspot group. The most powerful flares are often, but not always, accompanied by a coronal mass ejection (see next section). They were first observed by Richard Christopher Carrington and independently by Richard Hodgson in 1859 as a localized visible brightening of small areas within a sunspot group.

A typical flare will affect all the upper layers of the sun, namely the photosphere, chromosphere, and corona. The plasma is heated to tens of millions of kelvins, and electrons, protons, and helium nuclei, as well as heavier ions are accelerated to near the speed of light. The electromagnetic radiation produced ranges from radio waves to gamma rays, but most of the energy is spread outside of the visual range and so most flares are invisible to the naked eye. They occur in active regions around sunspots, where intense magnetic fields penetrate the photosphere to link the corona to the solar interior and are believed to originate from the sudden (timescales of minutes to tens of minutes) release of magnetic energy stored in the corona. This energy release may also cause a coronal mass ejection (CMEs), but it must be admitted that the actual relationship between CMEs and flares is still not well understood at this time.

For instance, it's not clear how magnetic energy is transformed into the kinetic energy of the particles, nor is it understood how some particles can be accelerated to the high energies seen. Furthermore, inconsistencies exist regarding the total number of accelerated particles, which sometimes appear to be greater than the total number in the coronal loop. What all this means is that scientists are currently unable to forecast when flares may occur, and this is a problem, as flares can have dire consequences. These are

---

[20] This cannot be emphasised enough, but reckless observing of the sun can be very dangerous. I can do no better that quote Richard Hill in his book "Observe and Understand the Sun" where he states "Observing the sun is the only inherently dangerous observing an amateur astronomer can do. Be aware of this at all times and take all necessary precautions. If you do not know a filter or procedure is safe, then do not use it! Always err on the side of safety. An eye once damaged is forever damaged. Filters that let too much INFRARED light through can burn an eye if used visually. There is NO PAIN when this happens. Burned retinas cannot be repaired. Excessive ULTRAVIOLET light has been shown to cause cataracts. So be very careful." Wise words indeed, and also remember to cover or remove your finder scope and never use one of those "solar filters" that screws into an eyepiece.

- Influence the local space weather in the vicinity of the Earth, whereby particles can impact the Earth's magnetosphere, presenting a radiation hazard to spacecraft and astronauts.
- Extreme solar flares are accompanied by coronal mass ejections (CMEs) that produce geomagnetic storms. These can disable satellites and knock out terrestrial electric power grids for extended periods of time.
- The soft X-rays emitted by X class flares can increase the ionization of the upper atmosphere, this can interfere with short-wave radio communication and can also warm the outer atmosphere thus increasing the drag on low orbiting satellites, leading to orbital decay.
- Large amounts of hard x-rays can damage to spacecraft electronics.
- One major concern in current deliberations of human missions to Mars, the Moon, or anywhere in the solar system is the radiation risk posed by solar flares. Energetic protons from flares can pass through the human body, causing biochemical damage presenting a hazard to astronauts during interplanetary travel. To minimize or even eradicate this risk, physical or magnetic shielding will be required to protect the astronauts.[21] Here's an example of this; a typical proton storm can take at least two hours from the time of visual detection to reach the Earth's orbit, and on January 20, 2005 a solar flare released the highest concentration of protons ever directly measured, which would have given astronauts on the moon little time to reach shelter.

Thought Question 8.5

A flare produces light in the X-ray, UV and visible parts of the spectrum, but only the visible light can be observed from the surface of the Earth. Why?

Finally, we come to the penultimate discussion in the chapter, where we discuss *Coronal Mass Ejections*, or CME's.[22]

Basically, a coronal mass ejection is a substantial release of plasma from the solar corona. They often follow solar flares and are typically present during a solar prominence.

---

[21] This even presents a problem to the astronauts in a space station. How could they protect themselves?

[22] Recently, a small number of CMEs have been observed on other stars, all of which as of 2016 have been found on red dwarfs.

Coronal mass ejections are often related with other forms of solar activity, but as with many things dealing with the Sun, a theoretical understanding of any relationship has not yet been established. However, what is known that CMEs most often originate from active regions on the Sun's surface, such as groupings of sunspots associated with frequent flares. Also, it has been noted that near solar maxima, the Sun produces about three CMEs every day, whereas near solar minima, there is about one CME every five days.

As in our discussion of flares, if a CME is directed towards Earth and reaches it as an interplanetary CME (ICME), the shock wave of mass causes a geomagnetic storm that may disrupt Earth's magnetosphere. Furthermore, coronal mass ejections, along with solar flares of other origin, can disrupt radio transmissions, and cause damage to satellites and electrical transmission line facilities, that may result in possibly enormous and long-lasting power outages.

However, it isn't all bad news as the high energy particles can cause particularly strong aurorae in large regions around Earth's magnetic poles. More commonly referred to as the Northern Lights (aurora borealis) in the northern hemisphere, and the Southern Lights (aurora australis) in the southern hemisphere.

Finally, we can now talk about the *Solar Wind*.

The solar wind is a stream of charged particles such as electrons, protons and alpha particles but with trace amounts of heavy ions and atomic nuclei such as C, N, O, Ne, Mg, Si, S, and Fe, emitted from the corona. The solar wind differs from the more dramatic flares and CME's as it is a constant stream of matter, and not random or sporadic. The amount of matter that is ejected is also impressive – about 1.3–1.9 million tonnes per second[23] with speeds ranging from 200 to 500 kms^{-1}. All this amounts a lot of material and energy.

Basically, the temperature of the corona is so high that the Sun's gravity cannot hold on to it, Although, it must be said that the details about how and where the coronal gases are accelerated to these high velocities is not clearly understood, although it must be something to do with magnetic fields. It varies in density, temperature and speed over time and over solar latitude and longitude as well as having two fundamental states, termed the slow solar wind and the fast solar wind. The origin of the slow solar wind is located around the Sun's equatorial belt, whereas the fast solar wind originates from the coronal holes, which are funnel-like regions of open field lines in the Sun's magnetic field and are particularly prevalent around the Sun's magnetic poles.

---

[23] Other stars, such as Sub-giants, Red Giants, and Supergiants, have much stronger stellar winds that result in a significantly higher mass-loss rate.

Besides having an effect on the Earth, the solar wind has effects on a much larger scale, that of the solar system itself.

- It is responsible for comets' ion tails and contributes to fluctuations in celestial radio waves observed on the Earth, through an effect called interplanetary scintillation.
- Also, those planets with a weak or non-existent magnetosphere are subject to a process called atmospheric stripping, whereby the atmosphere is actually ripped from the planet by the solar wind.
- It actually impacts the surfaces of the Moon and Mercury as they have very weak magnetic fields.
- The solar wind can be thought of as "blowing a bubble" in the interstellar medium that permeates the Milky Way. The location where the solar wind's strength is strong enough to push back the interstellar medium is known as the *heliopause* and is often considered as the outer border of the Solar System. This distance is not precisely known but probably depends on the current strength of the solar wind and the local density of the interstellar medium, but it believed to be far outside Pluto's orbit

The above processes occur in some form or other in many of the stars on the main sequence. As we shall see, more massive stars carry their energy outward in a different manner, and the energy is created in a slightly different way.

In the next few chapters we shall look at other stars and how they are placed on the main sequence.

Observing the Sun is a very popular pastime for amateur astronomers, but let me say now:

NEVER Observe, or Even Look at, the Sun with the Naked Eye or Through a Telescope.[24]

It is exceedingly dangerous, and you must have specially made equipment to do so. So don't do it. Instead, project the Sun onto a card. There are several excellent books on solar observing, several of which are mentioned in the appendices.

---

[24] There are of course filters and screens that one can buy to use with the telescope, but never, under any circumstances, use a telescope to look at the Sun, if not using specialized equipment. EVER!

Problems

1. Assume that the solar wind is travelling at ~400 kms^{-1}, and that the distance to the Sun is 1.5×10^8 km. Calculate the time it takes in days, for the solar wind particles to reach the Earth.
2. If the amount of mass converted to energy in the center of the Sun was 2.5 times greater, what now would be the energy released in a single reaction?
3. If the Sun fuses 600 million metric tons of hydrogen each second, and the mass of the sun is ~2×10^{30} kg, how long would it take to convert all the available hydrogen to helium? (Ignore stellar evolution).

Chapter 9

Binary Stars and Stellar Mass

9.1. Binary Stars

Binary stars, or, as they are sometimes called, double stars, are stars that may appear to the naked eye to be just one star, but on observation with either binoculars or telescopes resolve themselves into two stars. Indeed, some apparently single stars turn out to be several stars! Many appear as double stars due to their lying in the same line of sight as seen from Earth, and these are called *optical doubles*. It may well be that the two stars are separated in space by a vast distance.

Others, however, are actually gravitationally bound and may orbit around each other over a period of days, or even years. These systems are the ones we will discuss here.[1] The classification of some binary stars is quite complex, but we can broadly place them into one of three groups.

- Many that cannot be resolved by even the largest telescopes and are called *spectroscopic binaries*, the double component only being fully understood when the spectra are analyzed.
- Others are *eclipsing binaries*, such as Algol (β Persei), where one star moves during its orbit in front of its companion, thus alternately brightening and dimming the light observed.

---

[1] For a very detailed listing of beautiful double and multiple stars, see *Field Guide to Deep Sky Objects* 2nd Edition by the author.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_9

- A third type is the *astrometric binary*, such as Sirius (α Canis Majoris), where the companion star may only be detected by its influence on the motion of the primary star.

As this book is concerned with objects that can be observed visually, we will concentrate on binary stars that are physically associated and can be split with either the naked eye or with the use of some sort of optical equipment. What makes binary stars so important to astronomers is that by observing their motion,[2] as they dance around each other, it is possible to determine their mass. This is, of course, vitally important in determining the evolutionary processes in stars.

Terminology must now be introduced that is specific to visual binary star observation. The brighter of the two stars is usually called the *primary star*, while the fainter is called the *secondary star* (in some texts it may be called the companion; both terms will be used throughout this book). This terminology is employed regardless of how massive either star is or whether the brighter is in fact the less luminous of the two but appears brighter simply because it is closer to us.

Perhaps the most important terms used in visual binary star work are the *separation* and *position angle* (PA). The separation is the angular distance between the two stars, usually in seconds of arc, and is measured from the brighter star to the fainter. The position angle is a somewhat more difficult concept to understand. It is the relative position of one star, usually the secondary, with respect to the primary, and it is measured in degrees, with 0° at due north, 90° at due east, 180° due south, 270° at due west and back to 0°.

This system is best described by an example. Using Fig. 9.1, the double star γ Virginis, with components of magnitude 3.5 and 3.5, has a separation of 1.42″ (arc seconds) at a PA of 24° (epoch 2012.0).

Note that the secondary star is the one always placed somewhere on the orbit, the primary star is at the center of the perpendicular lines and that the separation and PA of any double star are constantly changing and should be quoted for the year observed. Some stars (where the period is very long) will have no appreciable change in PA for several years; others, however, will change from year to year and can be recorded by amateur astronomers.

It is worth mentioning again that although your optical equipment, including your eyes, should in theory be able to resolve many of the binaries

---

[2] Regardless of whether the binary star system is a visual, spectroscopic, eclipsing or astrometric binary star system.

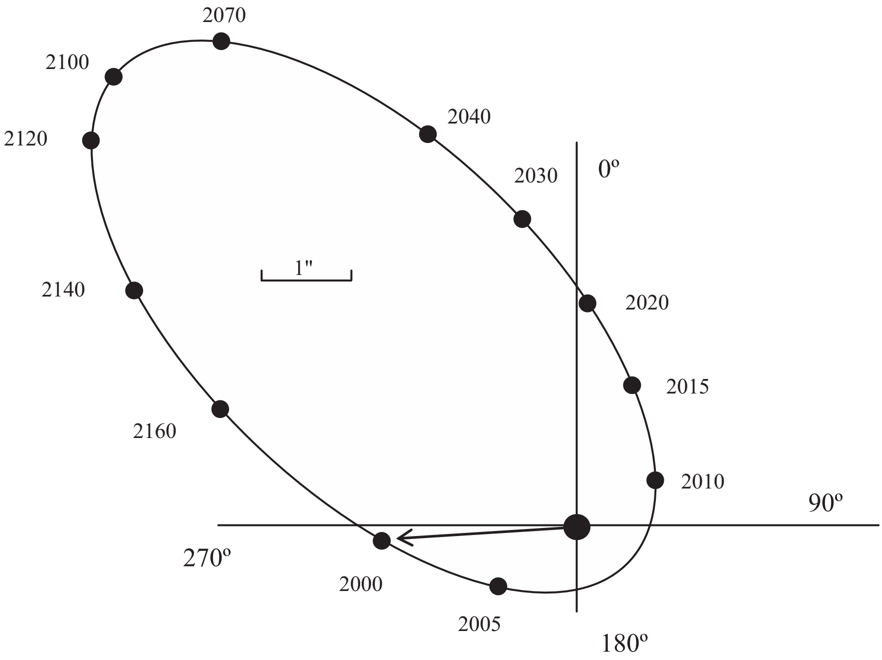


Fig. 9.1. γ Virginis

listed here,[3] there are several factors that will constrain the resolution (e.g., the seeing conditions, light pollution, dark adaption, your temperament, etc.). Thus, if you cannot initially resolve a double star, do not despair, but move onto another and return to the one in question at another date. Also recall that the colors ascribed to a star will not necessarily be the colors you see; they are just indicators of the general color, and in fact, as you will see from the text, many observers disagree on several stars' colors.

Thought Question 9.1

Can you think of a reason why visual binaries usually have longer periods than spectroscopic binaries?

---

[3] There are literally thousands of double, triple, and multiple star systems in the sky, all within reach of the amateur astronomer. The list that follows is but a taste of what awaits the observer. If your favorite star is omitted, I apologize, but to include them all would have been impossible.

Presented below is a brief list of visual binaries that will, as an initial indicator, help you to determine the resolution of both you and your binoculars/telescope.[4] All of the positions quoted are for the primary star.

μ Canis Majoris	ADS 5605	06h 56.1m	−14° 03′	January
5.3, 8.6 m	P.A. 340°; Sep. 3.0″		Spec. G5	

Two stars of differing brightness that nevertheless present a glorious double of orange and blue.

ξ Ursae Majoris	ADS 8119	11h 18.2m	+31° 32′	March
4.3, 4.8 m	P.A. 273°; Sep. 1.8″		Spec. G0	

Discovered by William Herschel in 1780, this is a close pair of pale yellow stars. It also has the distinction of being the first binary system to have its orbit calculated by Félix Savary in 1828. Both components are also spectroscopic binaries.

ζ Ursae Majoris	ADS 8891	13h 23.9m	+54° 56′	April
2.3, 4.0 m	P.A. 152°; Sep. 14.4″		Spec. A2 A2	

The famous double Mizar and Alcor (80 UMa). Visible to the naked eye, it is nice in binoculars. A small telescope will resolve Mizar's fourth-magnitude companion. Mizar also has several distinctions: the first double to be discovered by telescope (by Riccioli in 1650), the first to be photographed (by Bond in 1857), and the first spectroscopic binary to be detected (by Pickering in 1889). It used to be thought that the system was not a true binary; however, in 2009, research indicated that in fact Alcor is actually two stars, gravitationally bound with Mizar, and as Mizar is itself a quadruple system, we now have an amazing sextuplet system of stars!

α CVn	ADS 8706	12h 56.0m	+38° 19′	April
2.9, 5.5 m	P.A. 229°; Sep. 19.4″		Spec. A0	

Also known as Cor Caroli, the stars of this system are separated by a distance equivalent to five Solar System widths—770 astronomical units! The two stars are yellowish in small instruments; however, with large aperture, subtle tints become apparent and have been called flushed white and pale lilac or pale yellow and fawn!

ε Boötis	ADS 9372	13h 45.0m	+27° 04′	April
2.5, 4.9 m	P.A. 339°; Sep. 2.8″		Spec. K0 A0	

---

[4] Note that the position angle and separation are quoted for epoch 2000.0. With double stars that have small periods, these figures will change appreciably.

Also known as Mirak. This star shows a wonderful contrast of gold and green stars, although some observers report them to be yellow and blue. It can be difficult with apertures of about 7.5 cm, and even a challenge for beginners with apertures of 15 cm. With small telescopes, a high power is needed to resolve them.

β Lyrae	ADS 11745	18h 50.1m	+33° 22′	July
3.4_{v}, 8.6 m	P.A. 149°; Sep. 45.7″		Spec. B9	

This pair of white stars is a challenging double for binoculars. β^1 is also an eclipsing binary. A fascinating situation occurs due to the gravitational effects of the components of β^1—the stars are distorted from their spherical shapes into ellipsoids.

β Cygni	ADS 12540	19h 30.7m	+27° 58′	July
3.1, 5.1 m	P.A. 54°; Sep. 34.4″		Spec. K3 B8	

Thought by many to be the finest double in the skies, Albireo is a golden-yellow primary and has a lovely blue secondary against the backdrop of the myriad of fainter stars of the Milky Way. Easy to locate at the foot of the Northern Cross, and if you slightly defocus the image at the eyepiece, the colors become more vivid!

$\varepsilon^{1,2}$ Lyrae	ADS 11635	18h 44.3m	+39° 40′	July
4.67 m (4.7, 6.2 m)	P.A. 357°; Sep. 2.6″		Spec. F1V	
4.59 m (5.1, 5.5 m)	P.A. 94°; Sep. 2.3″		Spec. A8V	

This is the justly famous system known as the Double-Double, easily split, but to resolve the components of each star, ε^1 (magnitude 4.67) and ε^2 (magnitude 4.59), requires a high power. The stars themselves are at a P.A. 172°, separated by 208″, which is near the naked-eye limit, and some keen-eyed observers report being able to resolve them under perfect seeing conditions. However, there is fierce debate among amateurs—some say the double is difficult to resolve, others the opposite. All stars are white or cream-white in color.

61 Cygni	ADS 14636	21h 06.9m	+38° 45′	August
5.2, 6.0 m	P.A. 150°; Sep. 30.3″		Spec. K5 K7	

Famous for being the first star to have its distance measured by the technique of parallax, the German astronomer Friedrich Bessel determined its distance to be 10.3 light years; modern measurements give a figure of 11.36. It also has an unseen third component, which has the mass of eight Jupiters. The system has a very large proper motion across the sky. Best seen with binoculars (but sometimes a challenge if conditions are poor), which seem to emphasize the vibrant colors of these stars, both orange-red.

α Ursae Minoris	ADS 1477	02h 31.8m	+89° 16′	October
2.0, 8.7 m	P.A. 218°; Sep. 18.4″		Spec. F7Ib F6V	

This is possibly the most famous star in the sky. Polaris, or the Pole Star, is located less than a degree from the celestial pole and is a nice double consisting of a yellowish primary and a faint whitish-blue secondary. The primary is also a Population II Cepheid[5] variable and a spectroscopic binary. Recent research suggests that the distance to Polaris is about 434 l. y., while others claim it may be 30% closer, so that, if correct, this will make Polaris the closest Cepheid variable to Earth. Although claims have been made to the effect that the system can be resolved in an aperture as small as 4 cm, at least 6 cm will be required to split it clearly.

o Eradini	ADS 3093	04h 15.2m	−07° 39′	November
4.4, 9.5, 11.1 m	P.A. 104°; Sep. 83″		Spec. K1 DA4 M4.5e	

Omicron (o) Eridani, also known as 40 Eridani, is a star system at a distance of about 16.5 light years. 40 Eridani A is a main sequence star, spectral type K1, its two companion stars, 40 Eridani B and 40 Eridani C, being a ninth-magnitude white dwarf and an eleventh-magnitude red dwarf flare star, respectively.[6] What makes this system so interesting is that the secondary is the brightest white dwarf star visible from Earth and was the first white dwarf to be discovered, by William Herschel in 1783. It is relatively easy to observe but can be a challenge in binoculars.

Thought Question 9.2

How would you determine that a binary system is a true binary and not a visual double?

9.2. The Masses of Orbiting Stars

It may come as no surprise to you that the mass of a star can be determined. However, the question that needs to be asked is, “How?” Well, usually, we need to use binary stars, as well as the laws of Kepler and Newton. Both

---

[5] Cepheid variables are covered in a later chapter.

[6] Yes, I know that this is a triple star system, but as it contains a white dwarf, I’ve included it in the li dst.

men, and their contributions, were covered in the chapter on the Solar System, so it may be wise to re-read that chapter again to get an idea of their laws and how we can use them in an analysis of binary stars.

Kepler's law, which demonstrates how the time required for a planet orbiting the Sun is related to its distance from the Sun, can be modified to describe the motion of any two bodies that orbit around each other. The person who did this was the great Isaac Newton. To find the mass of the stars in a visual binary, we must first determine their orbits by observing them over several years. It may be that this will take a few years, or even tens of years, but eventually we can determine the time needed for one star to completely orbit the other.

This period of time is called P. By using a plot of the orbit, and forearmed with knowledge of the system's distance from the Sun, we can then measure the distance of the semi-major axis,[7] a, of one star to the other. An important point to note here is that this method only gives us the combined mass of the stars, not their individual masses. To achieve that, we need to go one step further.

From the above description, we can easily determine the combined masses of two stars orbiting each other. To determine individual stellar masses, however, we determine how much one star moves relative to the other. For instance, if one star is much more massive than the other, then it will hardly move at all relative to the less-massive star; the less-massive star will seem to do all of the orbiting in the system in a manner reminiscent of planets orbiting the Sun. To be accurate, we should really say that the stars (and, incidentally, the planets and Sun) orbit about their common center of mass, or center of gravity. In fact, they "wobble."[8]

The masses of stars usually orbit the center of mass that is more or less equidistant from both of them. This center of mass is found along a line joining the two stars at a position that depends on the stars' masses. Think of it as the balance point on a child's seesaw. If one star is four times as massive as the other, the balance point will be four times closer to the more massive star.

If the two stars have mass M_A and M_B, their distances from the center of mass (the balance point) are a_A and a_B,[9] with the larger mass having the smaller distance from the center of mass. So, for example, if the two stars have equal mass, or $M_A = M_B$, then $a_A = a_B$, and they orbit a point that is

---

[7] The semi-major axis is a term used in elliptical orbits (as nearly all orbits are). It is defined as half of the longer axis of an ellipse, and it is the average separation of the two stars.

[8] It is this wobble that is used to detect unseen planets around stars!

[9] Note that $a_A + a_B$ add up to the semi-major axis a, as used in Kepler's Law.

exactly halfway between them. On the other hand, if star *A* is four times more massive than star *B* ($M_B = ¼ \times M_A$), star *B* orbits four times farther from the center of mass than star A ($a_B = 4 \times a_A$). Again, using the image of a seesaw, an adult weighing four times as much as a child must sit four times closer to the pivot as a child.

In this manner, stellar masses can be determined, and using sophisticated techniques, the masses of double-star systems that cannot be optically resolved can also be measured. Using these and other techniques, we have determined that stellar mass ranges from about 0.08 $M_\odot$ to 50 $M_\odot$. Of course, there are larger stars, but these are few and far between. See Math Box 9.1.

Math Box 9.1 Determining Stellar Mass

Consider the orbits of the double-star system of Sirius A and Sirius B.

The two stars have an orbital period, *P*, of 50.1 years and an average semi-major axis, *a*, of 19.8 astronomical units. To determine their combined mass, $M_A + M_B$, we use the modified form of Kepler's Law:

$$M_A + M_B = \frac{a^3}{P^2}$$

Inserting the measure values for *P* and *a*, we get:

$$M_A + M_B = \frac{19.8^3}{50.1^2} \approx \frac{7762}{2510}$$

$$\cong 3.1\ M_\odot$$

Thus, the combined mass of Sirius A and Sirius B is approximately 3.1 $M_\odot$.

In reality, the orbit of the system is very eccentric, or elliptical, and so the distance between them varies from 31.5 to 8.1 AU and back again.

Using this information, and data gained from spectroscopic studies, we can determine that the mass of Sirius A is 2.12 $M_\odot$, and that of Sirius B 1.03 $M_\odot$.

Thought Question 9.3

At what point would it be more appropriate to use a spectroscopic method and not a visual method to determine the characteristic of a binary system?

Problems

1. A binary system has an orbital period of 75 years, along with an average semimajor axis of 27.5 AU. Make an estimate of their combined mass in solar units.
2. Another binary system has the following characteristics: a perihelion distance of 15 AU, and an aphelion distance of 5 AU, all measured form the mutual center of gravity. Along with an orbital period of 10 years. Make an estimate of their combined mass in solar units.
3. A binary system has a combined mass of $100M_{\odot}$, along with a period of 3.16 years. Determine its average semimajor axis.

Chapter 10

Life on the Main Sequence

10.1. Lifetimes of Main Sequence Stars

The stars that are on the main sequence are fundamentally alike in their cores because it is here that stars convert hydrogen to helium. This process is called *core hydrogen burning*. The *main sequence lifetime* is the amount of time a star spends consuming hydrogen in its core, and so the main sequence lifetime will depend on the star's internal structure and evolution. A newly born star is often referred to as a *zero-age main sequence star*, *ZAMS* for short. There is a subtle but important difference between a ZAMS star and a main sequence star. During its long life on the main sequence, a star will undergo changes to its radius, surface temperature, and luminosity due to the core hydrogen burning. The nuclear reactions alter the percentage of elements within the core. Initially, it would have had, say, in the case of the Sun, about 74% hydrogen, 25%, helium and 1% metals,[1] but now, after a period of 4.6 years, the core has a much greater mass of helium than hydrogen at its core.

Due to the hydrogen burning at the core, the total number of atomic nuclei decreases with time, and so with fewer particles in the core to provide

---

[1] Remember that all elements other that hydrogen and helium are called metals by astronomers.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_10

the internal pressure, the core will shrink very slightly under the weight of the star's outer layers. This has an effect on the star's appearance. The outer layers expand and become brighter. This may seem odd to you; if the core shrinks, why doesn't the star shrink? The explanation is very simple: the core shrinkage increases its density and temperature, which causes the hydrogen nuclei to collide with each other much more often, which in turn increases the rate of hydrogen burning.

The resulting increase of core pressure causes the star's outer layers to expand slightly, and as luminosity is related to the surface area of a star, the increase of the star's size will result in an increase of luminosity. In addition, the surface temperature will increase. In the case of the Sun, astronomers have calculated that its luminosity has increased by 40%, its radius by 6% and its surface temperature by 300 K, all during the past 4.6 billion years.

As a star ages on the main sequence, the increase of energy flowing from its core will also heat the surrounding area, and this will cause hydrogen burning to begin in this surrounding layer. As this can be thought of as "new" fuel for the star, its lifetime can be lengthened by a few million years for a main sequence star.

The one factor that determines how long a star will remain on the main sequence is its mass. Basically, it can be summed up in a few words: *Low-mass stars have much longer lifetimes than high-mass stars*. Figure 10.1 illustrates this nicely.

High-mass stars are extremely bright, and their lifetimes are very short. This means that they are using up their reserve of hydrogen in the core at a very high rate. Thus, even though an O- or B-type star is much more massive, and thus contains more hydrogen than, say, a less massive M-type star, it will use up its hydrogen much sooner. It may only take a few million years for O- or B-type stars to use up their supply of hydrogen, whereas for low-mass M-type stars, it may take hundreds of billions of years. Think about that for a second. The lifetime of an M-type star may be *longer than the present age of the universe!*[2] Table 10.1 shows how the mass of a star relates to temperature and spectral class.

Looking at star clusters one can easily see the differing lifetimes of stars. Massive stars have shorter lifetimes than less massive stars, and so a star cluster's H-R diagram will give information on the evolution of the stars in the cluster. Such a diagram will show a main sequence that lacks O-type

---

[2] We shall discuss this remarkable fact in later sections.

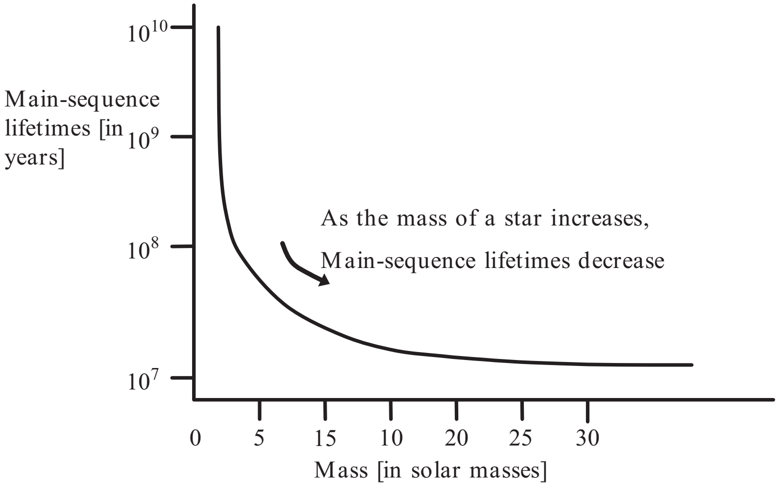


Fig. 10.1. Main sequence lifetimes for stars of different masses

Table 10.1 Mass, spectral class, and main sequence lifetimes

Mass ($M_\odot$)	Temperature (K)	Spectral class	Luminosity ($L_\odot$)	Main sequence lifetime (10^6 years)
25	35,000	O	80,000	3
15	30,000	B	10,000	15
3	11,000	A	60	500
1.5	7000	F	5	3000
1	6000	G	1	10,000
0.75	5000	K	0.5	15,000
0.50	4000	M	0.03	200,000

stars, which are the most massive, then A-type, and so on and so forth as the cluster ages. Figure 10.2 shows this erosion of main sequence stars by comparing the H-R diagrams of several different star clusters. In every case, some stars will have left the main sequence to become red giants,[3] or they are already red giants.

The temperature and spectral type of the very hot stars that remain on the main sequence are used to determine the age of a star cluster. Let's suppose the hottest star on the main sequence is an A0-type star, with the much hot-

---

[3] See the next section for a discussion on red giant stars.

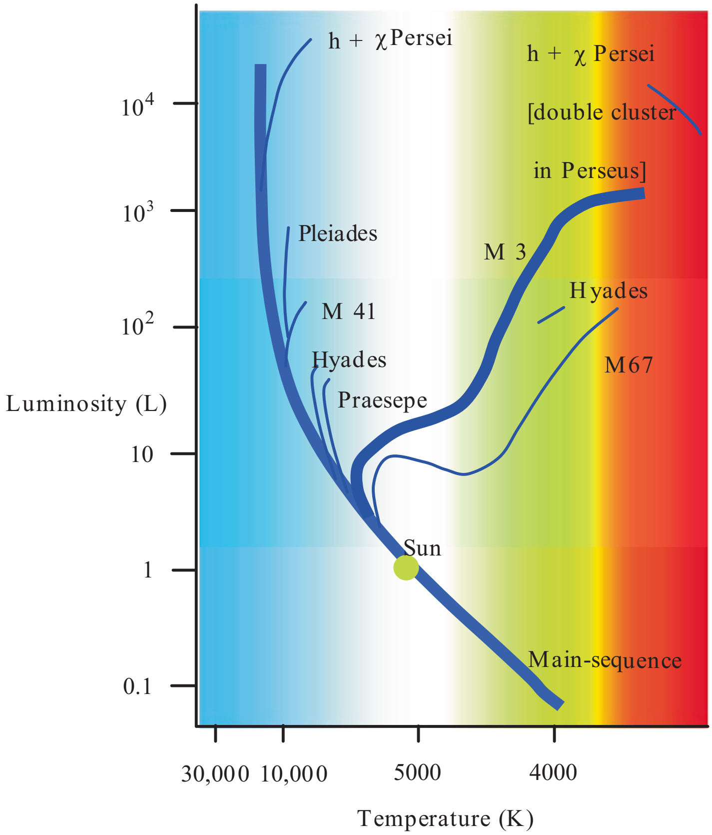


Fig. 10.2. H-R diagrams for several star clusters of different ages

ter and more massive stars already evolved to red giants. We know that A0 stars have a main sequence lifetime of about 100 million years, so we can say with some confidence that the star cluster is about 100 million years old.

The temperature and spectral type of the very hot stars that remain on the main sequence are used to determine the age of a star cluster. Let’s suppose the hottest star on the main sequence is an A0-type star, with the much hotter and more massive stars already evolved to red giants. We know that A0 stars have a main sequence lifetime of about 100 million years, so we can say with some confidence that the star cluster is about 100 million years old.

Thought Question 10.1

How many M class, main sequence red dwarf corpses have been observed?

Generally, the more massive the star, the faster it goes through all of its phases, so we are fortunate to be able to observe stars in the main sequence phase, as they remain on it for such a long time. It is also very easy to estimate a star's lifetime if we know its mass (see Math Box 10.1).

There are many stars on the main sequence that can be observed. The brightest of these have already been mentioned in previous sections. They include: Regulas, Vega, Sirius A, Procyon A, the Sun and Barnard's Star, to name but a few.

Math Box 10.1 Main Sequence Lifetimes

The length of time a star remains on the main sequence is very easy to calculate. There is an approximate relationship between the mass of a star and its lifetime:

$$t = \frac{1}{M^{2.5}} = \frac{1}{M^2\sqrt{M}}$$

Astronomers usually relate the main-sequence lifetime to the Sun (a typical 1 $M_\odot$ star), which is believed to be 10^{10} years, or 10 billion years.

For example, the main sequence lifetime of Sirius, a 2.12 $M_\odot$, star will be:

$$\frac{1}{2.12^{2.5}} = \frac{1}{2.12^2\sqrt{2.12}} = \frac{1}{6.54} \quad \text{solar lifetimes}$$

Thus Sirius will burn hydrogen in its core for about $1/6.54 \times 10^{10}$ years, or about 1.5 billion years.

On the other hand, a main sequence star with a mass of 0.5 $M_\odot$ will have a lifetime of:

$$\frac{1}{0.5^{2.5}} = \frac{1}{0.5^2\sqrt{0.5}} = \frac{1}{0.177} = 5.66 \quad \text{solar lifetimes}$$

This is about 56 billion years.

10.2. Red Giant Stars

Although the amount of hydrogen in a star's core is vast, it is not infinite, and so, after a very long time, the production of energy will cease when the central supply of hydrogen is used up. Throughout the length of time that nuclear fusion has been taking place, the hydrogen has been transformed into helium, by way of the proton-proton chain, and without this source of energy, the star uses gravitational contraction to supply its energy needs. Thus, the core will start to cool down, which means that the pressure also will decrease, with the result that the outer layers of the star begin to weigh down on the core and compress it. This has the effect of causing the temperature within the core to rise again, and for heat to flow outward from the core. Note that although a tremendous amount of heat is formed now, it is not due to nuclear reactions but to gravitational energy being converted into thermal energy.

In a relatively short time, astronomically speaking, the region around the star's hydrogen-depleted core will become hot enough to begin nuclear fusion of hydrogen into helium in a thin shell around the core in a process called *shell hydrogen burning*. This is shown in Fig. 10.3.

For a star like the Sun, this hydrogen-consuming shell develops almost immediately from the moment nuclear fusion stops in the core, and so the supply of energy is more or less constant. For massive stars, there can be an interval of perhaps a few thousand years to a few million years from the end of the core nuclear-fusion phase to the beginning of the shell hydrogen-burning phase.

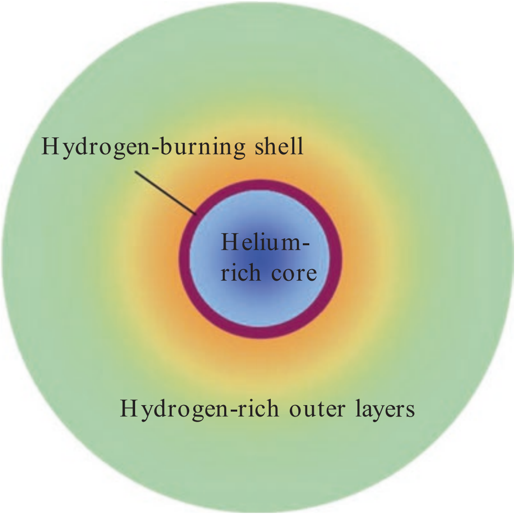


Fig. 10.3. Star with shell hydrogen burning. The core will consist of helium, but the outer layers are hydrogen rich. The shell, where energy production occurs, is relatively thin. [Not to scale]

The new supply of energy, and thus heat, has the effect of causing the rate of shell hydrogen burning to increase, and so it begins to eat further into the surrounding hydrogen. The helium that is the byproduct of the hydrogen fusion in the shell falls to the center of the star, where, along with the helium already there, heats up as the core continues to contract and increase its mass. In the case of, say, a 1 $M_\odot$ star, the core will be compressed to as much as one-third of its original size. The result of this core compression is an increase of the temperature, from about 15 million K to nearly 100 million K.

Now, most of what has happened in this stage of a star's life has occurred inside of it and so was invisible to our eyes. Nevertheless, it does have effects on a star's structure that drastically alter its appearance. The star's outer layers expand as the core contracts. With the increased flow of heat from the contracting core, and the ever-expanding shell of hydrogen burning, the star's luminosity increases quite substantially. This causes the star's internal pressure to increase and makes the outer layers of the star expand to many times their original radius. The tremendous expansion actually causes the outer layers to cool, even though the inner core temperature has risen dramatically. The new, much-expanded and cooler outer layers can reach temperatures as low as 3500 K and will glow with a very distinctive reddish tint, as can be explained by Wien's law (mentioned earlier). The star has now become a *red giant* star.

So, we can now see that red giant stars are former main sequence stars that have evolved into a new phase of their lives.

Due to the large diameter and thus weaker surface gravity of the red giant, quite a substantial amount of *mass loss* can occur. This means that gases can escape from the surface of a red giant star. Such an effect is relatively easy to observe by looking at the absorption lines produced in the star's spectrum. Calculations and measurements have shown that a typical red giant star can lose 10^{-7} $M_\odot$ each year. Compare this with the much-smaller 10^{-14} $M_\odot$ that the Sun loses each year. From this, we can see that as a star evolves from the main sequence to the red giant stage, it can lose quite a lot of its mass. The evolutionary track from the main sequence to the red giant phase for stars of differing mass is shown in Fig. 10.4.

There are many wonderful red giant stars that are observable in the night sky. We have already mentioned a few of these in earlier sections—Capella A, Arcturus, Aldebaran, Pollux, Mirach, R Leporis and Mira. But there are several other, lesser-known red giant stars that are also worth observing.

RS CYG	HD 192443	20^h 13.3^m	+38° 44′	JULY
8.1VM	B-V:3.3	C5		

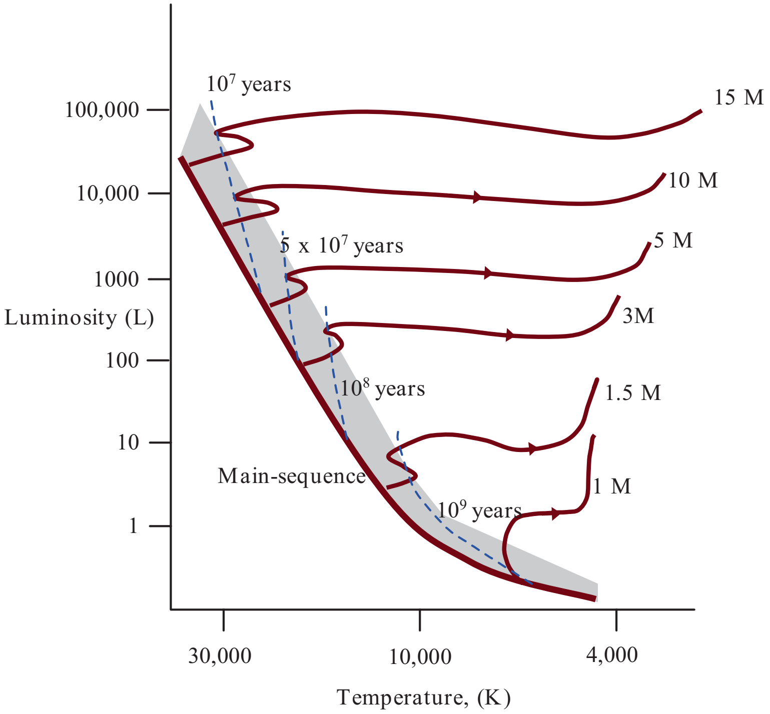


Fig. 10.4. Evolutionary tracks from the main sequence to the red giant phase for stars of different mass. The *dotted blue lines* indicate time scales of 10 million, 50 million, 100 million and 1 billion years. You can see that a star of about 15 solar masses leaves the main sequence (the *shaded area*) about 100 times earlier than a star of 1.5 solar masses

This is a red giant star with a persistent periodicity, class SRA, with a period of 417.39 days and a magnitude range of 6.5–9.5 m. A strange star where the light curve can vary appreciably, with the maxima sometimes doubling. A lovely deep red-colored star.

R Aqr	HD 222800	$23^h\ 43.8^m$	−15° 17′	August
5.8vm	B-V:1.5	M4 p e	2500 K	

This is a symbiotic double star and is classed as a Z Andromedae-type star. R Aqr is a nice red giant, which incidentally, has a small, blue (thus, very hot) companion star. Due to its variable nature, its magnitude can fall to 11.5, and so it can be somewhat difficult to locate. It is believed to lie at a distance of about 640 light years.

R Cas	HD 224490	$23^h\ 58.4^m$	+51° 23′	October
5.5_v m	B-V:1.5	M7 IIIe	2000 K	

This is a Mira-type variable star with quite a large magnitude range, 5.5–13.0 m. It is estimated to lie at a distance of 350 l. y. Its surface temperature of only 2000 K is still a matter of speculation.

R Leo	HD 84748	$09^h\ 47.6^m$	+11° 26′	February
6.02vm	B-V:1.5	M8 IIIe	2000 K	

Here we have a very bright Mira-type variable star that has become a favorite among amateur astronomers. Again, like so many other red giants, its low temperature of 2000 K is in some doubt. It is a very deep red color. This star is often cited as being a perfect introductory star for those who wish to observe a variable star. It is also an AGB[4] star.

Note that as a star ages and moves from the main sequence to the red giant stage, its spectral type will also change. For instance, the Sun, at present a G-type star, will gradually change its spectral class to K, and then to a warm M-type star. It may even become an M2- or M3-type, with the temperature falling to about 3200 K. Similarly, stars of different masses will also change their spectral type.

Finally, for this section, it is worth noting in passing that there are two red giant phases. Which one a star will follow depends, as you may have guessed by now, on its mass, and can lead to the formation of *supergiant* stars. We will discuss these in a later chapter.

Thought Question 10.2

What percentage of red stars visible to the naked eye are red giants or supergiants?

10.3. Helium Burning and the Helium Flash

All of the stars with a mass greater than, or equal to, the Sun's will eventually become red giants. But how energy is produced in the star after it has reached the red giant phase depends on its mass. We shall look at these two stages, beginning with how the helium in its core produces energy.

Helium Burning

Helium can be thought of as the "ash" left over from the hydrogen-burning reactions, and can in fact be used as the fuel for another nuclear fusion reaction that, this time, uses helium. This is the *helium-burning phase*. As a star

---

[4] Asymptotic Giant Branch.

approaches and becomes a red giant, its core temperature is too low to initiate helium burning. But the hydrogen-burning shell that surrounds the dormant helium core adds mass to the core, with the result that it contracts further, becomes denser, and the temperature increases substantially.[5] Something else happens as the temperature increases—the electrons in the gas become degenerate. Electron degeneracy is a very important process and is explained in greater detail in the appendices in this book. When the electrons become degenerate, they in effect resist any further contraction of the core, and the internal temperature of the core will no longer affect the internal pressure.

As the hydrogen shell continues to burn, the degenerate core grows even hotter, and when it reaches 100 million K and has a mass of about 0.6 $M_\odot$ (i.e., the inner 60% of the hydrogen in the star has been converted to helium), *core helium burning* begins, converting helium into carbon and producing nuclear energy. During this stage of a star's life, it can be nearly 1 AU in radius and almost 1000 times as luminous as the Sun. The by-now old star has once again obtained a central energy source for the first time since it left the main sequence.

The helium burning in the core fuses three helium nuclei to form a carbon nuclei in the so-called *triple α process*. This occurs in two steps. In the first step, two helium nuclei combine to form an isotope of beryllium:

$$^{4}\mathrm{He} + {}^{4}\mathrm{He} \rightarrow {}^{8}\mathrm{Be}$$

This isotope of beryllium is very unstable and very quickly breaks into two helium nuclei. But in the extreme conditions in the core, a third helium nucleus may strike the ^{8}Be nucleus before it has had a chance to break up. If this happens, a stable isotope of carbon is formed and energy is released as a gamma-ray photon (γ):

$$^{8}\mathrm{Be} + {}^{4}\mathrm{He} \rightarrow {}^{12}\mathrm{C} + \gamma$$

The origin of the phrase "triple α" comes about because helium nuclei are also called alpha particles.[6] The carbon nuclei formed in this process can also fuse with additional helium nuclei, producing a stable isotope of oxygen and supply additional energy:

---

[5] Recall that for hydrogen burning to start, the temperature has to reach about 10 million K, whereas for helium burning, the temperature has to achieve a staggering 100 million K.

[6] If you are a particle physicist!

$$^{12}\mathrm{C} + ^{4}\mathrm{He} \rightarrow ^{16}\mathrm{O} + \gamma$$

So the “ash” of helium burning is carbon and oxygen. This process is very interesting, as you will note that both of these isotopes of oxygen and carbon are the most abundant forms, and in fact make up the majority of carbon atoms in our bodies, as well as the atoms in the oxygen we breathe. We will explore this fascinating piece of information in greater depth later in the book.

The formation of carbon and oxygen not only provides more energy but also re-establishes thermal equilibrium in the core of the star. This prevents the core from any further contraction due to gravity. The length of time a red giant will spend burning helium in its core is about 20% as long as the time it spent burning hydrogen on the main sequence. The Sun, for example, will only spend 2 billion years in the helium-burning phase.

The Helium Flash

As mentioned earlier, the mass of a star will direct how helium burning begins in a red giant star. In a high-mass star (that is, with a mass greater than 2–3 $M_{\odot}$), the helium burning begins gradually as the temperature in the core approaches 100 million K. The triple α process is initiated, but it occurs before the electrons become degenerate. However, in low-mass stars (that is, with a mass less than 2–3 $M_{\odot}$), the helium-burning stage can begin very suddenly, in a process called the *helium flash*. This stage, the helium flash, occurs due to the most unusual conditions found in the core of a low-mass star as it becomes a red giant.

The energy produced by helium burning heats up the core of the star and raises its temperature. Now, in normal circumstances, this would result in an increase of pressure that would lead to an expansion and subsequent cooling of the core. This explains why nuclear reactions do not usually cause a rapid increase in the central temperature of a star. But we must remember that the gas in the core of a 1 $M_{\odot}$ red giant is far from normal; it is a gas of degenerate electrons. This means that any temperature increase that the helium burning produces does not increase the internal pressure. What the rise in temperature does do, however, is strongly affect the rate at which the triple α process occurs. A doubling of the temperature will increase the triple α production rate by about 1 billion times.

The energy that is produced by the triple α process heats up the core, and its temperature begins to rise even more. This increase and the subsequent rise in energy production can cause the temperature to reach an amazing 300 million K. Due to the rapid heating of the core, a nearly explosive consumption of helium occurs, and this is the helium flash mentioned earlier. At the peak of the helium flash, the core of the star has, very briefly, an energy output that is some 10^{11}–10^{14} times the solar luminosity. This converts to a rate of energy output that is about 100 times greater than the entire Milky Way.

Eventually, however, the high temperature becomes so high that the electrons in the core can no longer remain degenerate. They then behave normally for an electron in a gas, with the result that the star's core expands, with the corresponding end to the helium flash. These events occur very quickly, so the helium flash is over in a matter of seconds, and the star's core settles down to a steady rate of helium burning.

An important point to make here is that whether the helium flash occurs or doesn't occur, the start of helium burning actually reduces the star's luminosity. Here's what happens: the superheated core expands, and this core expansion pushes the hydrogen-burning shell outward, lowering its temperature and burning rate. The result is that even though the star has both helium fusion in its core and a shell of hydrogen burning taking place simultaneously, the total energy production falls from its peak during the red giant phase. This reduced total energy output of the star therefore reduces the luminosity and allows its outer layers to contract from their peak size during the red giant phase. As the outer layers contract, the star's surface temperature will increase slightly.

The helium burning in the core lasts for a relatively short time, however, and from calculations we can make an estimate of this time. For, say, a 1 $M_{\odot}$ star like the Sun, the period after the helium flash will only last about 100 million years, which is 1% of its main sequence lifetime.

Thought Question 10.3

The star Antares is a red supergiant, and has a lower surface temperature than the Sun, yet, has a much higher luminosity of about 85,000 $L_{\odot}$. *How?*

10.4. Globular Star Clusters and the H-R Diagram

At this point in our story of stellar evolution, it is a good idea to take stock of what we have learned so far. We have discussed how stars are formed before moving onto the main sequence. Their lifetime on the main sequence depends on their mass; massive stars have shorter lives. The red giant phase is next, along with a change in the hydrogen and helium burning within the star's core. To put all of this together in one coherent picture is useful, as we can see how a star develops from the moment of its birth; so we shall do just that by looking at the H-R diagram for stars that have just started their main sequence lifetimes and those that are in the red giant phase.

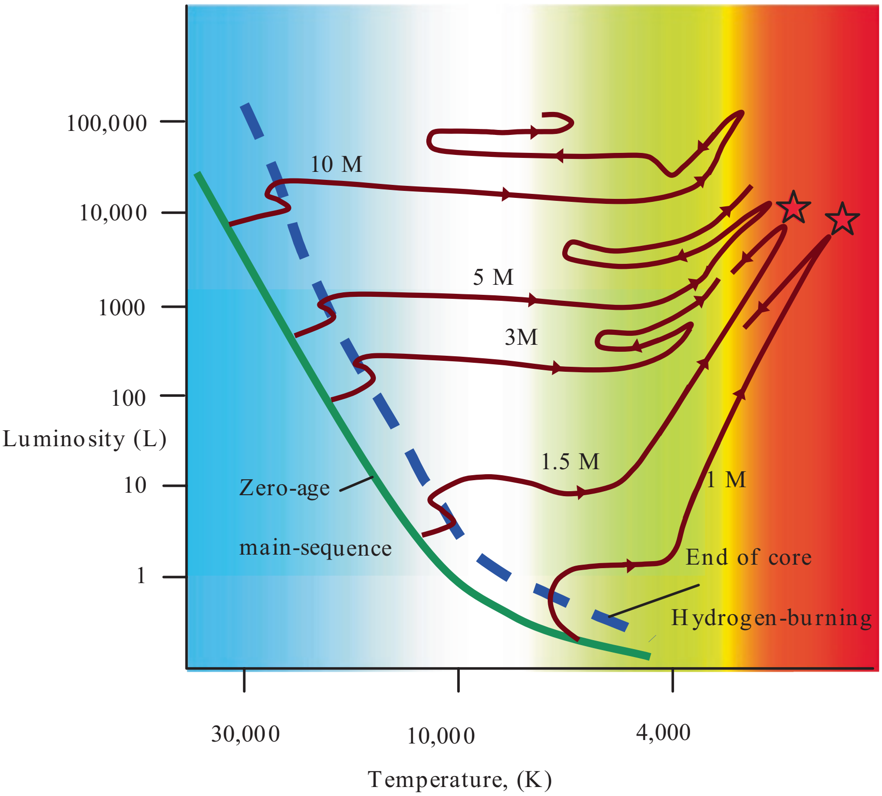


Fig. 10.5. Post-main sequence evolutionary track for several stars of different masses

Stars that have just emerged from the protostar stage and are about to join the main sequence are burning hydrogen steadily and have attained hydrostatic equilibrium. These stars are often referred to as zero-age main sequence stars and lie along a line on the H-R diagram called the zero-age main sequence, or ZAMS. This is shown on the H-R diagram in Fig. 10.5 as a green line. Over time, which can be relatively short or exceptionally long, depending on the star's mass, the hydrogen in the core is converted to helium, and the luminosity increases. This is accompanied by an increase in the star's diameter, and so the star moves on the H-R diagram away from the ZAMS. This explains why the main sequence is actually more of a broad band, rather than, as often portrayed, a thin line.

The blue-dashed line in Fig. 10.5 represents those stars where the hydrogen has been used up in the core and so nuclear fusion has ceased. As you can see, high-mass stars, 3 $M_\odot$, 5 $M_\odot$ and 10 $M_\odot$, then move rapidly from left (high temperature) to right (low temperature) across the H-R diagram. What is happening here is a decrease in surface temperature, but the surface area is increasing, so its overall luminosity remains fairly constant (i.e., an approximately horizontal line). In this phase, the core is contracting and outer layers expanding as energy flows from the hydrogen-burning shell.

High-mass stars with core helium burning exhibit sharp downward turns in the red giant region of the H-R diagram. Low-mass stars have a helium flash at their cores (red stars).

The evolutionary track of the high-mass stars then makes a turn to the upper-right section of the H-R diagram. This occurs just before the onset of core helium burning. After the start of helium burning, the core expands, the outer layers contract and the evolutionary track of the star falls from these high, albeit temporary, luminosities. Notice how the tracks then just wander back and forth on the H-R diagram. This represents the stars adjusting to their new energy supplies.

The low-mass stars, 1 $M_{\odot}$ and 1.5 $M_{\odot}$, behave in a somewhat different manner. The start of helium burning is marked by the helium flash, indicated by the red stars in the diagram. The star shrinks and becomes less luminous after the helium flash, although the surface temperatures rise. This occurs because the reduction in luminosity is proportionally less than the reduction in size. So now the evolutionary tracks move down (lower luminosity) the H-R diagram and to the left (a hotter) region.

We can observe the evolution of stars from birth to helium burning by looking at young star clusters and comparing the actual observations with theoretical calculations. But there exists another astronomical grouping of stars that contain many, maybe millions, of very old post-main sequence stars—globular clusters. These are the subjects of our next section.

10.5. Post-Main Sequence Star Clusters—The Globular Clusters

In the night sky are many compact and spherical collections of stars. These stars clusters are called *globular clusters*. These are metal-poor stars and are usually to be found in a spherical distribution around the galactic centre at a radius of about 200 light years.[7] Furthermore, the number of globular

---

[7] Recently, astronomers discovered a completely new type of star cluster in the Andromeda Galaxy (M31), that appear very similar to globular clusters They also contain hundreds of thousands of stars, just like that found in globular clusters, along with other characteristics with globular clusters such as stellar populations and metallicity. However, what distinguishes them from the globular clusters is that they are much larger – several hundred light-years across – and hundreds of times less dense. The distances between the stars are, therefore, much greater. It has been suggested that these clusters lie somewhere between a globular cluster and a dwarf spheroidal galaxy. Why the Andromeda galaxy has such clusters, while the Milky Way does not, is still unresolved, but it’s unlikely that Andromeda is the only galaxy with these extended clusters.

clusters increases significantly the closer one gets to the galactic center. This means that particular constellations that are located in a direction towards the galactic bulge have a high concentration of globular clusters within them, such as Sagittarius and Scorpius.

The origin and evolution of a globular cluster is very different from that of an open cluster. All the stars in a globular cluster are very old, with the result that any star earlier than a G- or F-type star will have already left the main sequence and be moving toward the red giant stage of its life.[8] In fact, new star formation no longer takes place within any globular clusters in our galaxy, and they are believed to be our galaxy's oldest structures. In fact, the youngest of the globular clusters is still far older than the oldest open cluster.[9]

The origin of globular clusters is a scene of fierce debate and research, with some suggesting that they may have been formed within the proto-galaxy clouds that came to make up our galaxy. However, in recent years two other origins for the globular clusters have come to light.

- Firstly, several seem to have been literally ripped from nearby smaller galaxies by the gravitational attraction of the Milky Way, and now orbit the core of our galaxy, for example, Palomer 12.
- Secondly, it is known that our galaxy has destroyed several smaller galaxies by a process called galactic cannibalism. The remnants of these devoured galaxies are, in some instances, believed to be some of the globular clusters we see.[10] Such a globular cluster is the deservedly famous *Omega Centauri.*[11]

---

[8] Current research suggests that globular clusters do in fact have multiple populations of stars, for example the clusters in the Large Magellanic Cloud are believed to have encountered a large molecular cloud which triggered a new burst of star formation.

[9] New research conducted, at the Department of Physics of The University of Hong Kong using data collected by the Hubble Space Telescope, suggests that some globular clusters are much younger. Looking at globular clusters surrounding the giant galaxy at the centre of the Perseus galaxy cluster it appears a few thousand of the clusters seem to have formed over at least the past 1 billion years.

[10] It is believed that up to quarter of the globular cluster population in the Milky Way may have originally been dwarf galaxies. In addition, more than 60% of the globular clusters in the outer halo of the M31, the Andromeda Galaxy, were likely accreted in a similar way.

[11] For many years, it had been suggested that Omega Centauri is the core of a dwarf galaxy that was tidally disrupted and assimilated by the Milky Way. Furthermore, Kapteyn's Star, only 13 light years away from Earth, is thought to originate from Omega Centauri. The chemistry and motion of the cluster support this idea. In addition, the cluster exhibits a range of metallicities and stellar ages that suggests that it did not all form at once as globular clusters are thought to form and may in fact be the remainder of the core of a smaller galaxy long since incorporated into the Milky Way. The globular cluster Mayall II also has similar characteristics.

Over the past several years, new ideas have come forth about the clusters, especially concerning clusters that consist of more than just old stars. Due to the fact that they have very high star densities, then it is not too far-fetched to assume that interactions and even near-collisions may, from time-to-time, occur. This could result in such exotic objects such as blue stragglers,[12] millisecond pulsars[13] and low mass X-ray binaries.[14]

Searching for black holes in a globular cluster presents its own set of problems, however. A telescope with a very high resolution is needed, and to date only a few observations have been made using the Hubble Space Telescope, that have resulted in a few controversial detections of possible black holes. Research suggested that in the globular cluster M15, there was an intermediate-mass black hole with a mass of 4000 $M_{\odot}$. Furthermore, a 20,000 $M_{\odot}$ black hole in the Mayall II cluster located in the Andromeda galaxy was also detected, and both x-ray and radio emissions from Mayall II seemed to be confirm this, but observations[15] made in 2018 now cast doubt on this result, although it has been suggested that it could be in fact much smaller, with a black hole in the range 500–1000 $M_{\odot}$.

As previously mentioned, globular clusters are old, as they contain no high-mass main sequence stars, and this can be shown on a special kind of H-R diagram called a *color magnitude diagram*. On a color magnitude diagram, the apparent brightness is plotted against the color ratio for many of the stars in a cluster (see Fig. 10.6). The color ratio of a star can tell you the surface temperature, and if we assume that all the stars in a cluster lie at the same distance from us, their relative brightnesses can tell us their relative luminosities.

Even a cursory glance at such a color magnitude diagram will tell you something strange has happened. In fact, you will see that the upper half of the main sequence has disappeared. This means that all of the high-mass stars in a globular cluster evolved into red giants a long time ago. What remain are the low-mass main sequence stars that are very slowly turning into red giants.

---

[12] A blue straggler is formed from the merger of two stars, maybe as a result of an encounter with a binary system. As a consequence, the star will have a higher temperature than comparable stars in the globular cluster with the same luminosity, and thus will be different from the main sequence stars formed at the birth of the cluster.

[13] A millisecond pulsar (MSP) is a pulsar (see the section of Neutron stars) with a rotational period smaller than about 10 milliseconds.

[14] A low-mass X-ray binary (LMXB) is a binary star system where one of the components is either a black hole or neutron star.

[15] These same observations from 2018 concluded that there was no evidence for an intermediate-mass black hole in any globular cluster.

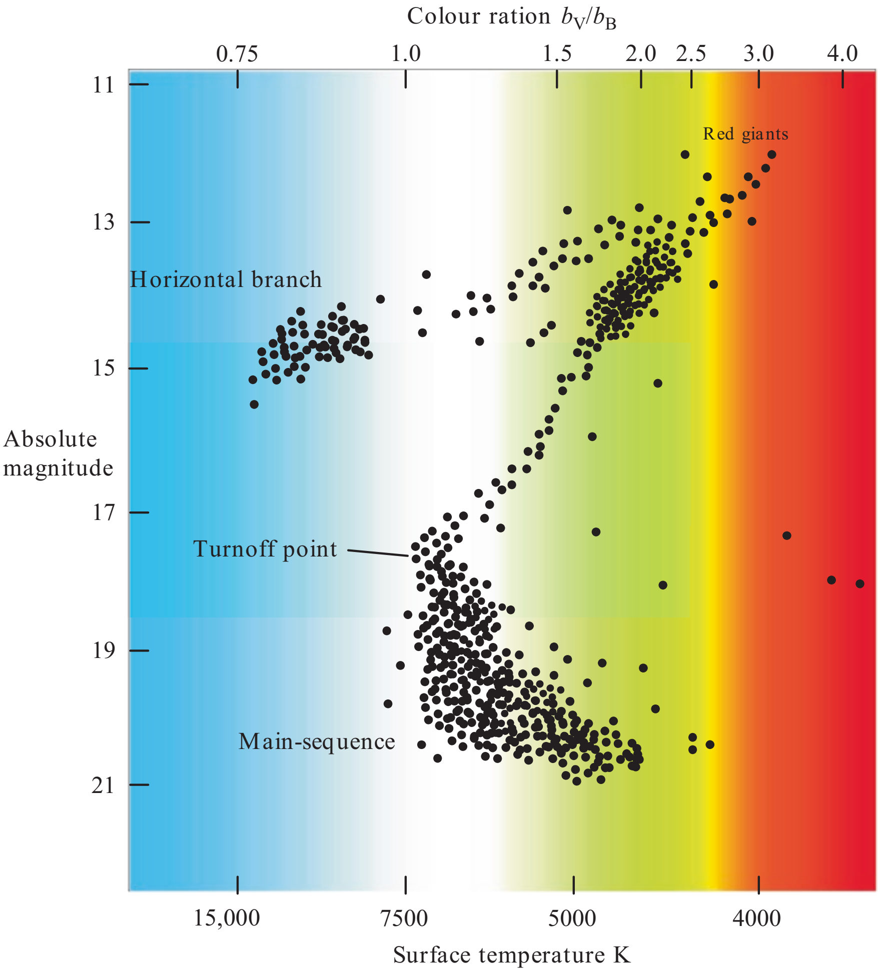


Fig. 10.6. Color magnitude diagram for the globular cluster M3

One thing that is very apparent on the diagram is a grouping of stars that lie on a horizontal band towards the center-left of the diagram. This is called the *horizontal branch*, while the stars in this area are the *horizontal branch stars*. These stars are low-mass, post-helium flash stars of about 50 $L_\odot$, in which there are both core helium burning and shell hydrogen burning. In the future, these stars will move back toward the red giant region as the fuel is devoured.

A star that has had its surface temperature and visual magnitude determined is represented by a black dot. All the stars in M3 lie at approximately the same distance from us (32,000 light years), so their magnitudes are a direct measurement of their luminosities. The asymptotic giant branch is described in a later section.

One very practical use of the H-R diagram is to estimate the age of a star cluster. With a very young star cluster, most, if not all, of the stars are on or near the main sequence. As they age, however, the stars will move away from the main sequence, with the high-mass, high-luminosity stars being the first to become red giant stars. As time passes, the main sequence will get increasingly shorter. The top of the main sequence, which remains after this time, can be used to determine the cluster's age and is called the *turnoff point*. The stars that are at the turnoff point are those that are just exhausting the hydrogen in their cores, so the main sequence lifetime is in fact the age of the star cluster. Figure 10.7 illustrates an example of the H-R diagram for open star clusters showing their turnoff points.

The time for the turnoff point is shown as a purple line. For example, the cluster M41 has a turnoff point near the 10^8 year point; thus the cluster is about 100,000,000 years old.

From an observational point of view, globular clusters can be a challenge. Many are visible in optical instruments, from binoculars to telescopes, and a few are even visible to the naked eye. There are about 200 globular clusters, ranging in size from 60 to 150 light years in diameter.[16] They all lie at vast distances from the Sun and are about 60,000 light years from the galactic plane. The nearest globular clusters (for example, Caldwell 86 in Ara) lie at a distance of over 6000 light years, and thus the clusters are difficult objects for small telescopes. That is not to say they can't be seen; rather, it means that any structure within the cluster will be difficult to observe. Even the brightest and biggest globular will require apertures of at least 15 cm for individual stars to be resolved. However, if large-aperture telescopes are used, these objects are magnificent. Some globular clusters have dense concentrations towards their center, while others may appear as rather compact open clusters. In some cases, it is difficult to say where the globular cluster peters out and the background stars begin.

An image of a globular cluster can be seen at the end of the chapter. Photo 10.1

As in the case of open clusters, there exists a classification system, the *Shapley-Sawyer Concentration Class*, where Class I globular clusters are

---

[16] When is a globular cluster not a globular cluster? For example, BH 176 in the southern part of the Milky Way exhibits the properties of both an open and a globular cluster. Research shows it to be a transition-type object, a metal-rich globular, or an open cluster, with distance estimates in the range of 15–85 kpc, indicating that the true nature of this cluster is rather uncertain. I would suggest that more work needs to be done on this object.

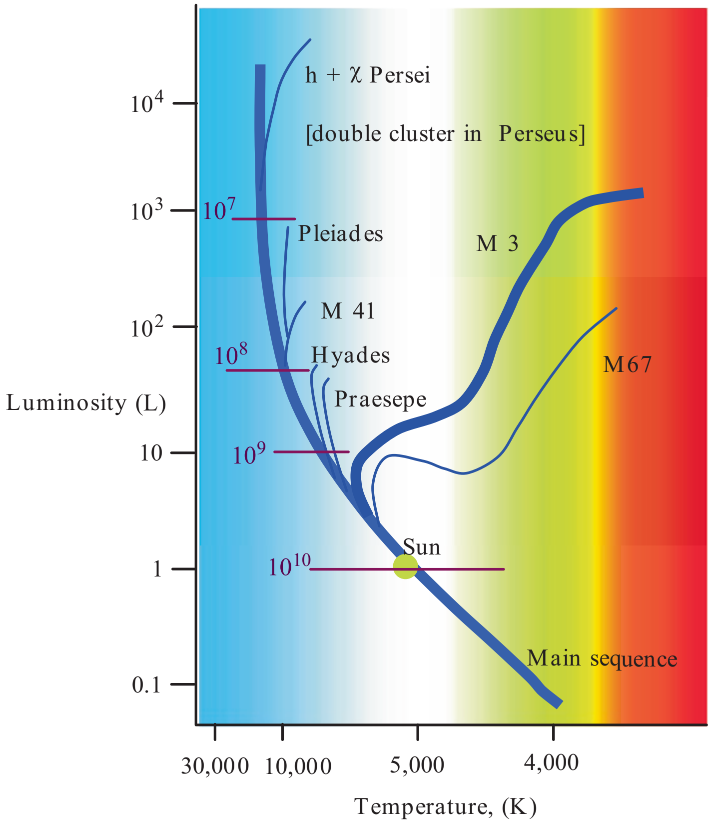


Fig. 10.7. H-R diagram for star clusters showing the turnoff point

the most star-dense and Class XII the least.[17] The ability of an amateur to resolve the stars in a globular actually depends on how condensed the cluster is, and so the scheme will be used in the descriptions, but it is really only useful for those amateurs who have large-aperture instruments. Nevertheless, the observation of these clusters, which are amongst the oldest objects visible to amateurs, can provide you with breath taking, almost three-dimensional aspects.

---

[17] In 2015, a new type of globular cluster was proposed on the basis of observational data: dark globular clusters. These have an unusually high mass for the number of stars within it. So, it has been suggested that they contain objects with significant dark matter components, as well as massive black holes at their centre. Needless to say, these are not visually observable, however, several have been located around the galaxy Centaurus A, so could be a challenge for the experienced astrophotographer. Maybe**.**

Photo 10.1. The Hercules Cluster, Messier 13

The many globular clusters listed below are just a few of the literally hundreds that can be observed and are meant to be just a representative sample. The ⊕ indicates the approximate size of the cluster.

Messier 68	NGC 4590	12h 39.5m	−26° 45′	April
7.3 m	⊕ 11′	X		

Appearing only as a small, hazy patch in binoculars, this is a nice cluster in telescopes, with an uneven core and faint halo. A definite challenge to naked-eye observers, where perfect seeing conditions will be needed. Use averted vision and make sure that your eyes are well and truly dark-adapted.

Messier 3	NGC 5272	13h 42.2m	+28° 23′	April
6.2 m	⊕ 18.6′	VI		

What we have here is a splendid globular cluster, easily seen in binoculars and a good test for the naked eye. If using giant binoculars with perfect seeing, some stars may be resolved. A beautiful and stunning cluster in telescopes, it easily rivals Messier 13 in Hercules. It definitely shows pale colored tints, and reported colors include, yellow, blue and even green; in fact, it is often quoted as the most colorful globular in the northern sky. Full of structure and detail, including several dark and mysterious tiny dark patches. Many of the stars in the cluster are also variable. This is one of the three brightest clusters in the northern hemisphere, and deserves to be on every observers list. The cluster is the prototype for Oosterhoff Type I objects, which are considered "metal-rich." That is, for a typical (is there such an object?) globular cluster, Messier 3 has a relatively high abundance of heavier elements. Located at a distance of about 34,000 light years it is believed to be about 8 billion years old.

Messier 5	NGC 5904	15h 18.6m	+02° 05′	May
5.7 m	⊕ 23.0′	V		

Easily seen as a disc with binoculars and with large telescopes, the view is breathtaking—presenting an almost three-dimensional vista. One of the few colored globulars, with a faint, pale yellow outer region surrounding a blue-tinted interior. It gets even better with higher magnification, as more detail and stars become apparent. Possibly containing over half a million stars, this is one of the finest clusters in the northern hemisphere; many say it is *the* finest.

Messier 4	NGC 6121	16h 23.6m	−26° 32′	May
5.8 m	⊕ 26.3′	IX		

Another superb object, presenting a spectacle in all optical instruments. It does however lie very close to the star, α Scorpii, or Antares, so that the glare may prove a problem in its detection with small-aperture telescopes. High-power binoculars will even resolve several stars. Telescopes of all apertures show detail and structure within the cluster, and the use of high magnification will prove beneficial; but what is more noticeable is the bright lane of stars that runs through the cluster's center. Thought to be the closest globular to Earth at 7200 light years (although NGC 6397 in Ara may be closer), and about 12.2 billion years old. Research has discovered many white dwarf stars in the cluster, believed to be among the oldest known, and one in particular is a binary star with a pulsar companion, PSR B1620–26, and a planet orbiting it with a mass of 2.5 times that of Jupiter. Note that there are a few, unsubstantiated reports that the cluster has been glimpsed with the naked eye, but this author has never managed to see it.

Messier 13	NGC 6205	16h 41.7m	+36° 28′	June
5.8 m	⊕ 20.0′	V		

A splendid object and the premier cluster of the northern hemisphere. Visible to the naked eye, it has a hazy appearance in binoculars; with telescopes, however, it is magnificent, with a dense core surrounded by a sphere of a diamond dust-like array of stars. In larger telescopes, several dark bands can be seen bisecting the cluster. It appears bright because it is close to us, at only 23,000 light years, and also because it is inherently bright, shining at a luminosity equivalent to more than 250,000 Suns. At only 140 light years in diameter, the stars must be very crowded, with several stars per cubic light year, a density some 500 times that of our vicinity. Also known as the Hercules Cluster. See Photo 10.1

Messier 10	NGC 6254	16h 57.1m	–04° 06′	June
6.4 m	⊕ 20′	VII		

This globular is similar to Messier 12, but slightly brighter and more concentrated. It lies close to the orange star 30 Ophiuchi (spectral type K4, magnitude 5), and so if you locate this star, then by using averted vision M10 should be easily seen. Under medium aperture and magnification, several colored components have been reported—a pale blue-tinted outer region surrounding a very faint pink area, with a yellow star at the cluster's center.

Messier 19	NGC 6273	17h 02.6m	–26° 16′	June
6.8 m	⊕ 17′	VIII		

A splendid, albeit faint, cluster when viewed through a telescope. Although a challenge to resolve, it is nevertheless a colorful object, reported as having both faint orange and blue stars, while the overall color of the cluster is a creamy white.

Messier 9	NGC 6333	17h 19.2m	−18° 31′	June
7.7 m	⊕ 12.0′	VII		

Visible in binoculars, this is a small cluster with a brighter core. The cluster is one of the nearest to the center of our galaxy and is in a region conspicuous for its dark nebulae, including Barnard 64; it may be that the entire region is swathed in interstellar dust, which gives rise to the cluster's dim appearance. It lies about 19,000 light years away.

Messier 22	NGC 6656	18h 36.4m	−23° 54′	June
5.1 m	⊕ 32′	VII		

A truly spectacular globular cluster, visible under perfect conditions to the naked eye. Low-power eyepieces will show a hazy spot of light, while high power will resolve a few stars. Often passed over by northern hemisphere observers due to its low declination. Only 10,000 light years away, nearly twice as close as M13.

Messier 92	NGC 6341	17h 17.1m	+43° 08′	June
6.3 m	⊕ 14.4′	IV		

This is a beautiful cluster often overshadowed by its more illustrious neighbor, Messier 13. It is a somewhat difficult object to locate, but once found is truly spectacular. It can be glimpsed with the naked eye. In binoculars it will appear as a hazy small patch, but in 20-cm telescopes its true beauty becomes apparent with a bright, strongly concentrated core. It also has several very distinct dark lanes running across the face of the cluster. It is a very old cluster, possibly 13 billion years, 25,000 light years distant and is an Oosterhoff Type II (OoII) globular cluster, signifying that it belongs to the group of metal-poor clusters with longer period RR Lyrae variable stars.

Messier 54	NGC 6715	18h 55.1m	−30° 29′	July
7.6 m	⊕ 9.1′	III		

This has a colorful aspect—a pale blue outer region and pale yellow inner core. Recent research has found that the cluster was originally related to the Sagittarius Dwarf Galaxy, but the gravitational attraction of our Galaxy has pulled the globular from its parent. Among the globular clusters in the Messier catalog, it is one of the densest, as well as the most distant.

Messier 15	NGC 7078	21h 30.0m	+12° 10′	August
6.4 m	⊕ 12′	IV		

Messier 15 is an impressive globular cluster and can be glimpsed with the naked eye under perfect conditions. Using binoculars it will appear as a hazy object with no resolvable stars. However, under medium magnification and aperture, it becomes much more impressive and will show considerable detail such as dark lanes, arcs of stars and a noticeable asymmetry. It is one of only four globular clusters that have a planetary nebula located within it—Pease-1. To see this nebula an aperture of 30 cm at least will be needed, along with a detailed star map of the field, the judicious use of an appropriate filter, along with a lot of patience. The cluster has undergone "core collapse," resulting in an extremely high number of stars at its center that may be home to a black hole. The cluster is also an X-ray source.

Thought Question 10.4

Look at the above list of clusters, especially the given dates when the clusters are highest in the sky. Can you think of a reason why most are in the summer/autumn sky?

10.6. Pulsating Stars

We saw earlier that there are stars far more massive than the Sun that contract and move horizontally across the H-R diagram, while at the same time they get hotter but remain at a constant luminosity. As they move across the H-R diagram they can also become unstable and vary in size. Some stars change their size quite considerably, alternatively shrinking and expanding, as their surface moves in and out. As the stars vary in size, so does their brightness. These stars are the *pulsating variable stars*. There exist several classes of pulsating variable stars, but we will just discuss the main types: the *long-period variables*, the *Cepheid variables* and the *RR Lyrae* stars. Figure 10.8 shows where on the H-R diagram these pulsating stars reside.

10.7. Why Do Stars Pulsate?

You may think that the pulsations of a star are caused by variations in the rate of energy production deep in its core. You would be wrong, however, as the rate of nuclear fusion remains constant in a pulsating star. Astronomers have realized that the variations are caused by changes in the rate at which energy can *escape* from the star. The explanation is surprisingly simple, but somewhat involved, so the various stages will be presented in some detail.

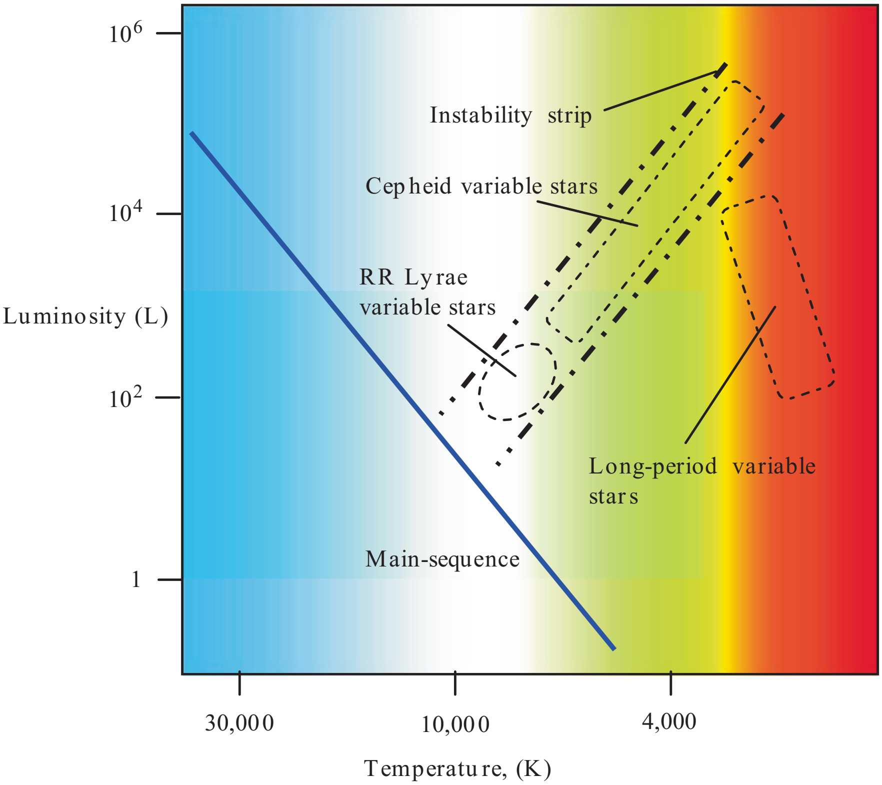


Fig. 10.8. Variable stars on the H-R diagram

Imagine a normal star, where there is a balance between the downward-pulling force of gravity and the upward force of pressure (i.e., the star is in hydrostatic equilibrium). Now picture a star where the pressure in the outer layers *exceeds* the force of gravity in those layers. In such a scenario, the star's outer layers would begin to expand. (See Fig. 10.9 for a schematic of this process.) As the star expands, its gravity will naturally fall, but the pressure force will fall at a faster rate. A time would then come when the star will have expanded to a larger size where, once again, hydrostatic equilibrium would reign.

However, this doesn't necessarily mean that the star would stop expanding. The inertia of the outward-moving outer layers will carry the expansion past the balance point. By the time gravity would have brought everything to a stop, the pressure would now be too small to balance the gravity, and so the outer layers would begin to fall inward. At this point gravity would

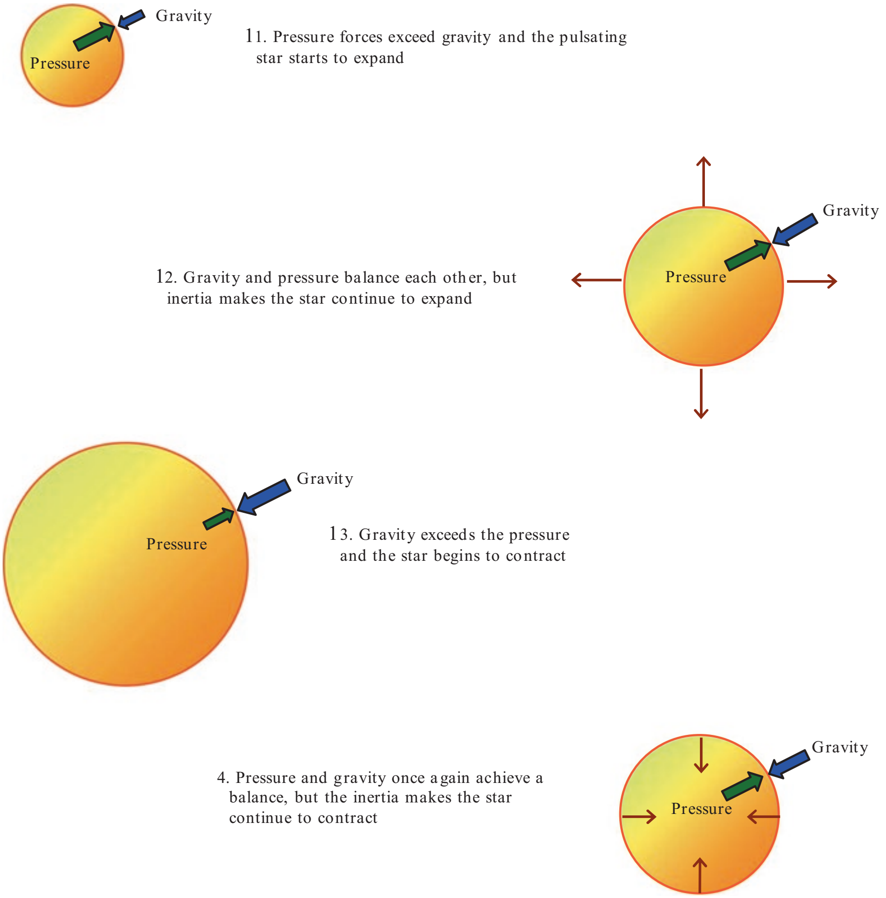


Fig. 10.9. Gravity and pressure during the pulsation cycle of a pulsating star

rise again, but less than the pressure. The outer layers would fall past the balance point until eventually the force of pressure would prevent any further fall and so would come to a stop. And here is where we came in—the pulsations would start all over again.

You can think of a pulsating star behaving just like a spring with a heavy weight attached to it. If you pull down on the weight and then let it go, the spring will oscillate around the point at which the tension in the spring and the force of gravity are in balance. After a while, however, friction in the spring will dampen down the oscillations unless the spring is given a little push upwards each time it reaches the bottom of an oscillation. In a pulsating star, for the pulsations to continue and not die out, the star also needs an outward push each time it contracts to its minimum size. Discovering what causes that extra push was a challenge to astronomers of the twentieth century.

The first person to develop an idea of what was happening was the British astronomer Arthur Eddington in 1914. He suggested that a star (in this case, a Cepheid variable) pulsated because its opacity increases more when the gas is compressed than when it is expanded. Heat is trapped in the outer layers if a star is compressed, which increases the internal pressure; this, in turn, pushes upward the outer layers. As the star expands the heat will escape, and so the internal pressure falls and the star's surface drops inward.

Then in 1960, the American astronomer John Cox further developed the idea and proved that helium is the key to a Cepheid's pulsations. When a star contracts, the gas beneath its surface gets hotter, but the extra heat doesn't raise the temperature; instead, it ionizes the helium. This ionized helium is very good at absorbing radiation. In other words, it becomes more opaque and absorbs the radiant energy flowing outward through it, towards the surface. This trapped heat makes the star expand. This, then, provides the "push" that propels the surface layers of the star back outward. As the star expands, electrons and helium ions recombine, and this will cause the gas to become more transparent (i.e., its opacity falls, and so the stored energy escapes).

For a star to be susceptible to pulsations, it must have a layer beneath its surface in which the helium is partially ionized. The existence of such a layer will depend not only on the size and mass of a star but also on its surface temperature that, in most cases, will be in the range of 5000–8000 K. There is a region on the H-R diagram where such an area exists, and it is in the location of the pulsating stars. The region is called the *instability strip*. In this region are found the Cepheid variable and RR Lyre stars.

10.8. Cepheid Variables and the Period-Luminosity Relation

Cepheid variables are named after δ Cephei, which was the first star of its type to be discovered. It is a yellow giant star that varies by a factor of two in brightness over 5.5 days.[18] Figure 10.10 shows the variations of δ Cephei in luminosity, size and temperature.

---

[18] The time for a star to complete one cycle in its brightness variation is called its period. Thus, for δ Cephei, its period is 5.5 days.

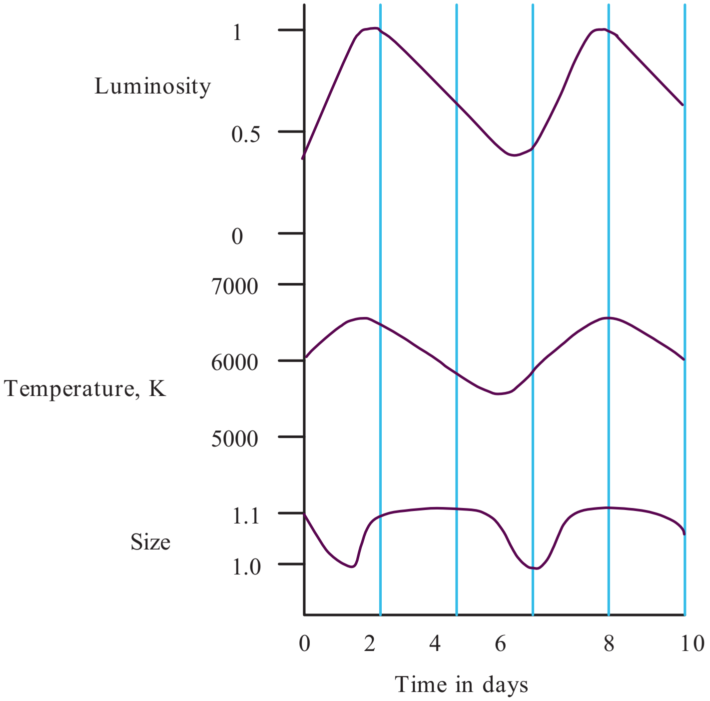


Fig. 10.10. The size, temperature and luminosity of δ Cephei during one period

You will notice immediately that its luminosity and temperature have a maximum value when its size has a minimum value, and vice-versa; its size is at maximum when its luminosity and temperature are at minimum. Cepheids are very important for astronomers for two reasons. They can be seen at extreme distances, perhaps as great as a few million parsecs. This is because they are very luminous, with a range from a few hundred to a few tens of thousands of solar luminosities (i.e., 100 $L_\odot$ to 10,000 $L_\odot$). Second, there exists a relationship between the period of a Cepheid and its average luminosity. The very faintest Cepheids (that are in fact hundreds of times brighter than the Sun) pulsate with a very rapid period of only 1 or 2 days, while the brightest (as much as 30,000 times brighter than the Sun) have a much slower period of about 100 days. The correlation between the pulsation period and luminosity is called the *period-luminosity relation*. If a star can be identified as a Cepheid, and its period measured, then its luminosity and absolute magnitude can be determined. This can then be used, along with its apparent magnitude, to determine its distance.

The percentage of metals in a Cepheid star's outer layers will determine how it pulsates. This occurs because the metals can have a substantial effect

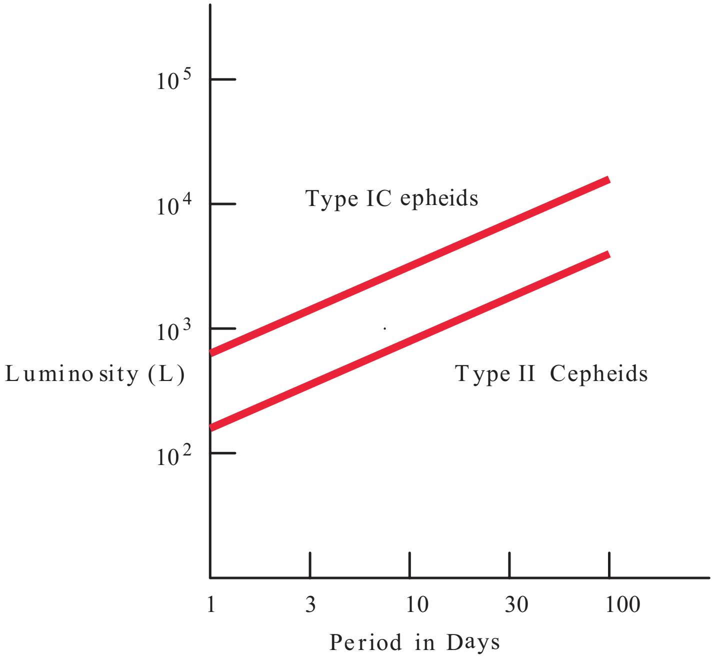


Fig. 10.11. Period-luminosity relationship for the two types of Cepheid variable star

on the opacity of the gas. They can then be classified according to their metal content. If a Cepheid is a metal-rich *Population I* star,[19] it is called a *Type I Cepheid*, and if it is a metal-poor *Population II* star, it is called a *Type II Cepheid*. Figure 10.11 shows a period-luminosity diagram for the two types of Cepheids. So, an astronomer must first determine what type of Cepheid he or she is observing before the period-luminosity relationship can be applied.

---

[19] Population I stars are bright supergiant main sequence stars with high luminosity, such as O- and B-type stars, and members of young open star clusters. Molecular clouds are often found in the same location as Population I stars. They are usually in the disk of a galaxy and concentrated in the spiral arms, following nearly, but not always, circular orbits. Population I stars include stars with a range of ages, maybe 10 billion years old, or 1 year old. Population II stars, on the other hand, are usually old stars. Examples include RR Lyrae stars and the central stars of planetary nebulae. This type of star has no correlation with the location of the spiral arms. They are also found in globular clusters, which are almost entirely in the halo and central bulge of the galaxy. Therefore, they represent the oldest stars, which formed very early in the history of the galaxy.

10.9. Temperature and Mass of Cepheids

The period-luminosity relationship comes about because the more massive stars are also the most luminous stars as they cross the H-R diagram during core helium burning. These massive stars are also larger in size and lower in density during this period of core helium burning, and the period with which a star pulsates is larger for lower densities; so, the massive pulsating stars have the greater luminosities and the longest periods. This is shown in Fig. 10.12.

We have seen that old, high-mass stars have evolutionary tracks that cross back and forth on the H-R diagram and thus will intercept the upper end of the instability strip. Such stars become Cepheids when the helium ionizes at just the right depth to drive the pulsations. Those stars on the left (high temperature) of the instability strip will have helium ionization occurring too close to the surface and will involve only a small fraction of the stars' mass. The stars on the right (low-temperature) side will have convection in

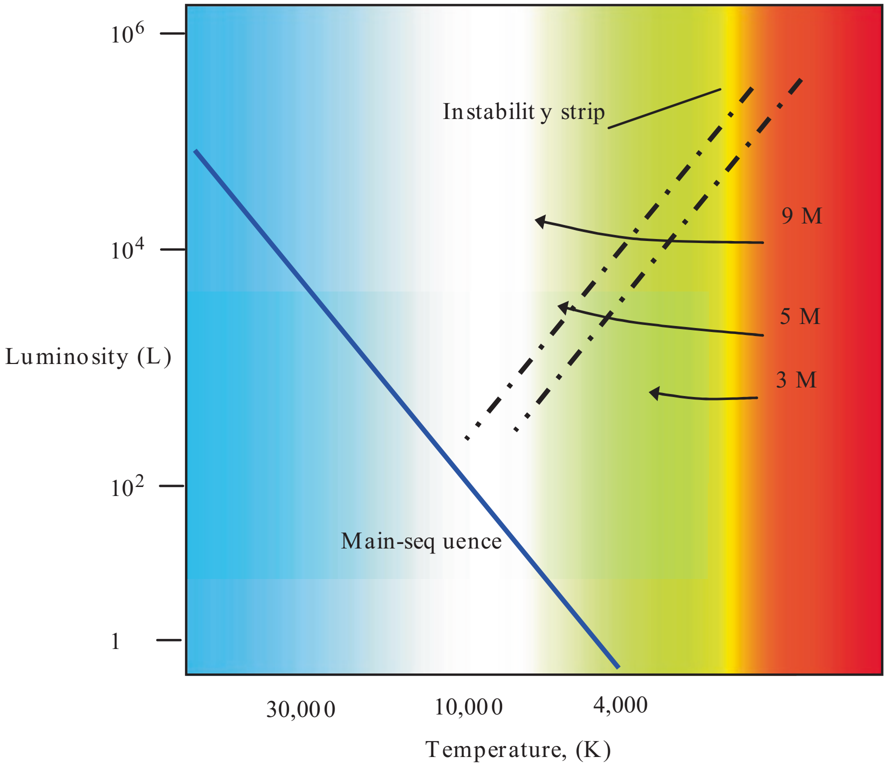


Fig. 10.12. Instability strip and evolutionary tracks for stars of different masses

the star's outer layers, and this will prevent the storage of the heat necessary to drive the pulsations. Thus, Cepheid variable stars can only exist in a very narrow temperature range.

Thought Question 10.5

Cepheid variable have a range of luminosities and corresponding periods. Given a choice, would you observe a low luminosity or high luminosity Cepheid in order to determine its characteristics?

10.10. RR Lyrae and Long-Period Variable Stars

The faintest and hottest stars on the instability strip are the RR Lyrae stars. These stars have a much lower mass than Cepheids. After the helium flash occurs, their evolutionary tracks pass across the lower end of the instability track as they move across the horizontal branch of the H-R diagram. The RR Lyraes, named after the prototype in the constellation Lyra, all have periods shorter than Cepheids, of about 1.5 h to 1 day. They are small and dense stars compared with the Cepheids (but are nearly 10 times larger and about 100 times more luminous than the Sun), and all have the same intrinsic luminosity, about $50L_\odot$. The RR Lyrae region of the instability strip is in fact a segment of the horizontal branch. They are all metal-poor Population II stars, and so many are found in globular clusters.

Thought Question 10.6

How can RR Lyrae stars be used to determine distances within the Milky Way?

The long-period variables are cool red giant stars that can vary by as much as a factor of 100, in a period of months or even years. Many have surface temperatures of about 3500 K and average luminosities in a range of 10 to as much as 10,000 $L_\odot$. They're situated on the middle right-hand side of the H-R diagram. Many are periodic, but there are also a few that are not. A famous example of a periodic long-period variable star is Mira (o Ceti) in Cetus. A famous non-periodic long-period variable star is Betelgeuse (α Orionis) in Orion. After all that has just been written, it may come as a surprise to you to know that we do not fully understand why some cool red giant stars become long-period variable stars.

There are many pulsating stars that can be observed by the amateur astronomer, and in fact several organizations exist throughout the world that cater specifically to this pastime.[20] However, we shall just describe the brightest members of each of the three aforementioned classes: Cepheid, RR Lyrae and long-period variable stars. All that is needed in observing these stars is a degree of patience, as the changes in magnitude can take as little as a few days to several hundred days, and, of course, clear skies.

The nomenclature used in this list is the same as that used before, with a few changes. The apparent magnitude range of the variable is given, along with its period in days.

η Aquilae	HD 187929	$19^h\ 52.5^m$	+01° 00′	July
3.48 to 4.39 m	−3.91 m	7.17 days	F6 to G4	

This is a nice Cepheid to observe, as its variability can be seen with the naked eye. The rise to brightest magnitude takes 2 days, with the remainder of the time in a slow fading. The nearby star Beta β Aquilae (3.71 m) is often used as a comparison. It is the third brightest Cepheid (in apparent magnitude), after Delta Cephei and Polaris. The actual period is 7 days, 4 h, 14 min and 23 s!

δ Cephei	HD 213306	$22^h\ 29.1^m$	+58° 25′	August
3.48 to 4.37 m	−3.47 m	5.37 days	F5 to G3	

This is the prototype star of the classic short-period pulsating variables known as Cepheids. It was discovered in 1784 by the British amateur astronomer John Goodricke. It is an easy favorite with amateurs, as two bright stars also lie in the vicinity—Epsilon (ε) Persei (4.2 m), Zeta (ζ) Persei (3.4 m), Zeta (ζ) Cephei (3.35 m), and Eta (η) Cephei (3.43 m). The behavior of the star is as follows: it will brighten for about 1.5 days and will then fade for 4 days, with a period of 5 days, 8 h and 48.2 min. Delta Cephei is also a famous double star, with the secondary star (6.3 m) a nice white color, which contrasts nicely with the yellowish tint of the primary.

RT Aurigae	HD 45412	$06^h\ 28.6^m$	+30° 29′	December
5.29 to 6.6 m	−2.65 m	3.73 days	F5 to G0	

Also known as 48 Aurigae, this star was discovered to be variable in 1905 by T. Astbury, a member of the British Astronomical Association. The rise to maximum takes 1.5 days, with a diminishing over 2.5 days. Easy to observe in binoculars, it lies midway between Epsilon (ε) Geminorum

---

[20] A list of organizations is in the appendices in this book.

(3.06 m) and Theta (θ) Auragae (2.65 m). The period has been measured to an astounding accuracy to be 3.728261 days!

α Ursae Minoris	ADS 1477	$02^h\ 31.8^m$	+89° 16′	October
1.92 to 2.07 m	−3.64 m	3.97 days	F7:Iib to Iiv	

Everyone knows this one, possibly the most famous star in the sky—Polaris, or the Pole Star. Polaris A, the supergiant primary component, is a classic Population I Cepheid variable, although it was once thought to be a Population II due to its high galactic latitude. The magnitude changes are very small and therefore not really detectable with the naked eye. It is also a nice double, consisting of a yellowish primary and a faint whitish-blue secondary at a magnitude of 8.2. The primary is a Population II Cepheid variable[21] and a spectroscopic binary. Although claims have been made to the effect that the system can be resolved in an aperture as small as 4 cm, at least 6 cm will be required to split it clearly. Polaris has several distinctions that are always good to declare to unsuspecting members of the public at star parties: it will be closest to the actual pole in A. D. 2102, it also is the closest Cepheid to the Solar System, and it is located less than a degree from the celestial pole.

Other Cepheid variable stars that can be observed with amateur equipment are: U Aquilae, Y Ophiuchi, W Sagittae, SU Cassiopeiae, T Monocerotis and T Vulpeculae.

Here, now, are some bright RR Lyrae variables.

RR Lyrae	HD 182989	$19^h\ 25.3^m$	+42° 47′	July
7.06 to 8.12 m	1.13 m	0.567 days		

This is the prototype of the RR Lyrae class of pulsating variable stars. These are similar to Cepheids but have shorter periods and lower luminosities. There are no naked-eye members of this class of variable, and RR Lyrae is the brightest member. There is a very rapid rise to maximum, with the light of the star doubling in less than 30 min, with a slower falling in magnitude. From an observational viewpoint, it is a nice white star, although detailed measurements have shown that it does become bluer as it increases in brightness. There is some considerable debate as to the changes in spectral type that accompany the variability. One source quotes A8 to F7, while another A2 to F1. Take your pick. There is also some indication that there is another variability period along with the shorter one, which has a period of about 41 days.

---

[21] There is some doubt about this, again due to its high galactic latitude.

Other RR Lyrae variable stars are: RV Arietis, RW Arietis, and V467 Sagittari; however, all of these stars are faint and so will present a considerable challenge to observers.

Here, now, are some long-period variables.

MIRA	o CET	02^h 19.3^m	−02° 59′	OCTOBER
2.00 to 10.1 m	−3.54 m	331.96 days	M9 to M6e	

An important star, and maybe the first variable star ever observed. Written records mentioning it certainly exist as far back as 1596. The prototype of the long-period pulsating variable, it varies from third to tenth magnitude over a period of 332 days, and it is an ideal star for the first-time variable star observer. At minimum, the star is a deep red color, but, of course, faint. It now has a lower temperature of 1900 K. The period, however, is subject to irregularities, as is its magnitude, and can be longer, or shorter, than the quoted average of 332 days. It has been observed for maximum light to reach first magnitude—similar to Aldebaran! One of the oddities about Mira is that the change in spectral class does not occur exactly with maximum but rather a few days later! Another oddity is that when Mira is at its faintest, it apparently is also at its largest, when you would think the opposite to be true. A reason for this has been put forward recently: the star produces titanium oxide in its atmosphere as it cools and expands. The compound then acts as a filter, blocking out the light. The name *Mira* is Arabic for "wonderful star." For those of you determined to observe this famous star, here are a couple of dates for your diary for predicted maxima: 2015—May 21–31, and 2016—April 20–30.

Other Mira-type stars are R Leonis and R Leporis, both of which have been described in earlier sections.

Problems

1. Determine the lifetime of a 100 $M_\odot$ main sequence star.
2. Determine the lifetime of a 0.08 $M_\odot$ main sequence star.
3. Determine the lifetime of a 28 $M_\odot$ Red supergiant star.

Chapter 11

Star Death: White Dwarfs and Planetary Nebulae

11.1. The Death of Stars

Stars live for millions, billions and even hundreds of billions of years,[1] and so you may be thinking how on Earth can we know anything about how a star dies? After all, our planet is only 4.5 billion years old, and we have been studying astronomy for about the last 500 years. Well, fortunately for us, it is nevertheless possible to observe the many disparate ways in which a star can end its life.

Once again, it is the mass of a star that decides how it will end its life, and the results are spectacular and sometimes very strange indeed. Low-mass stars can end their lives in a comparatively gentle manner, forming beautiful and apparently delicate structures that we know as planetary nebulae, before proceeding to small and ever-cooling white dwarf stars. At the other end of the scale, high-mass stars tend to end their lives in a far more spectacular fashion by exploding! These are the rare supernovae.

We begin our journey by looking at stars that have a low mass.

---

[1] Current theories predict that low-mass M-type stars will stay on the main sequence for trillions of years!

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_11

11.2. The Asymptotic Giant Branch

Let's recap briefly how low-mass stars (and by this we mean stars with a mass of about 4 $M_\odot$ and less), behave after leaving the main sequence. When core hydrogen burning ceases, the core will shrink, and this heats up the surrounding hydrogen gas, and so hydrogen-shell burning begins. The outer layers of the star will expand but also cool, and so the star becomes a red giant. The post-main sequence star will move up and to the right on the H-R diagram as its luminosity increases and temperature falls. We can say that the star now lies on the red giant branch of the H-R diagram. The next stage involves the onset of helium burning in the core. If a star has a high mass (greater than about 2–3 $M_\odot$), then this starts gradually, but if the star has a lower mass, this stage begins suddenly, in what is called the helium flash. But no matter which way it starts, a result of the helium burning is that the core actually cools down, with a resulting slight decrease in luminosity. The outer layers of the star also contract a little, heating them up in the process, and so the evolutionary track of the red giant now moves left across the H-R diagram. The luminosity during this phase remains more or less constant, so the path is nearly horizontal, and so is called the horizontal branch. Stars on the horizontal branch are stars in which helium burning is occurring in the core, which in turn is surrounded by a shell of hydrogen burning. Many such stars are often found in globular clusters.

We can now look at the next stage of a star's life. Recall that the byproducts of the triple α process are the elements carbon and oxygen. So after a suitably long period of time, maybe 100 million years, we could expect all of the helium in the core to have been converted into carbon and oxygen. This would mean that core helium burning would cease. A similar process to that (which was explained earlier) then begins. The absence of nuclear fusion results in a contraction of the core because there is no energy source to provide the internal pressure necessary to balance the force of gravity. The core contraction is stopped, however, by degenerate electron pressure, which is something we met earlier. A result of the core contraction is a release of heat into the helium gas surrounding the core, and so helium burning begins in a thin shell around the carbon-oxygen core. This is, aptly enough, called shell helium burning.

Now an extraordinary thing happens: the star enters a *second* red giant phase. It is as if history has repeated itself. Stars become red giants at the end of their main sequence lifetimes. The shell hydrogen-burning phase provides energy, causing the outer layers of the star to expand and cool. In a similar fashion, the energy from the helium-burning shell also causes the outer layers to expand, and so the low-mass star rises into the red giant region of the H-R diagram for a second time. But this time it has an even greater luminosity.

This phase of a star's life is often called the *asymptotic giant branch* phase, or *AGB*. Thus, these stars are called AGB stars and are on the asymptotic giant branch of the main sequence.

The structure of an AGB star is shown in Fig. 11.1. Its central region is a degenerate carbon-oxygen mix surrounded by a helium-burning shell, which in turn is surrounded by a helium-rich shell. A hydrogen-burning shell further surrounds this. A hydrogen-rich outer layer further encompasses all of this. What is truly remarkable is the size of these objects. The core region is about the same size as Earth, while the hydrogen envelope is immense. It can be as large as the orbit of Earth! When the star has aged, however, the outer layers, which are expanding, cause the hydrogen-burning shell to also expand and then cool, and so the nuclear reactions occurring therein may cease, albeit temporarily.

The luminosity of these stars can be very high indeed. For example, a 1 $M_\odot$ AGB star may eventually attain a luminosity of 10,000 $L_\odot$. Compare this with the luminosity of only 1000 $L_\odot$ it achieved when it reached the helium-flash phase and the poor 1 $L_\odot$ when it resided on the main sequence. It is sobering to think ahead and imagine what will happen when the Sun becomes an AGB star about 8 billion years from now!

There are many AGB stars that can be observed by amateurs, and in fact we have already mentioned and described several of them. The archetypal

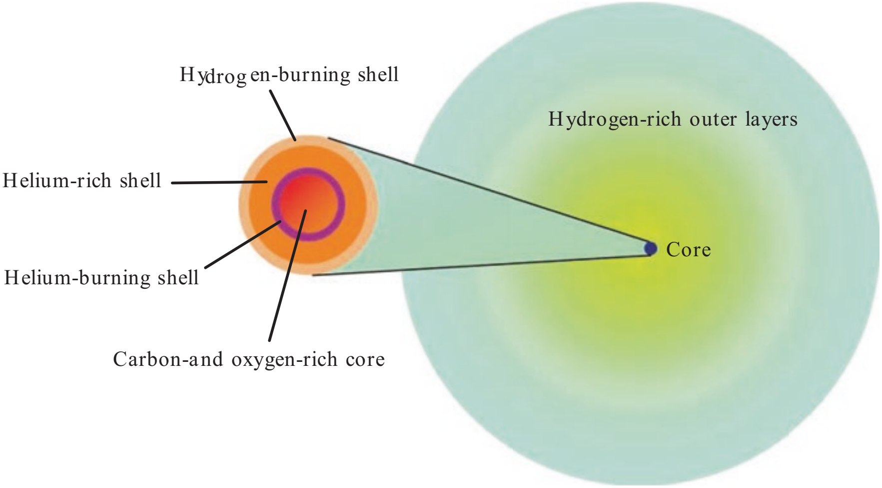


Fig. 11.1. The structure of an AGB star

AGB star is Mira (o Ceti), but there are also R Leonis, R Leporis, R Aquarii, and R Cassiopea. In addition, there are a few others, although they are somewhat fainter. These include χ Cygni, W Hydrae, S Pegasi and TT Monocerotis.

11.3. Dredge-Ups

We have seen that energy and heat are transported from a star's core to the surface by one of two methods: convection or radiation. Convection is the motion of the star's gases moving upwards towards the surface, and then a cooling of the gases so that they fall downwards. This method of energy transfer is very important in giant stars. Radiation, or radiative diffusion as it is also sometimes called, is the transfer of energy using electromagnetic radiation, and is only important when the gases in a star are transparent (the opacity is low). When a star ages and has left the main sequence, the convective zone can increase substantially in size, and sometimes extend right down to the core. This means that the heavy elements, or metals, that are formed there can be carried to the star's surface by convection. This process has the very unglamorous name of *dredge-up* and has three phases:

- The *first* dredge-up begins when the star becomes a red giant for the first time (i.e., core hydrogen-burning phase has stopped). The byproducts of the CNO[2] cycle of hydrogen are transported to the surface because the convective zone now reaches deep into the core regions.
- A *second* dredge-up begins when the helium-burning phase ends.
- Then, during the AGB phase, a *third* dredge-up occurs, but only if the mass of the star is greater than 2 $M_{\odot}$, when a large amount of newly formed carbon is carried to the star's surface.

The spectrum of a star that has such a carbon-enriched surface exhibits very prominent absorption bands of carbon-rich elements, such as C_2, CH and CN. Such stars that have undergone a third dredge-up are often called *carbon stars*.

---

[2] In stars that are hotter and have a higher mass than the Sun, the chain of reactions that leads to hydrogen fusion is called the CNO cycle, where C, N and O stands for carbon, nitrogen and oxygen, respectively. The amount of energy produced in this reaction is exactly the same as that produced by the proton-proton reaction discussed earlier, but it occurs at a much more rapid rate.

11.4. Mass Loss and Stellar Winds

As a star continues to rise up the AGB, it increases in both brightness and size, and, consequently, it develops a very strong stellar wind. This blows the star's outer layers into interstellar space. Thus, the star undergoes a substantial mass loss during this phase, maybe as much as 10^{-4} $M_\odot$ per year. (This is about 1000 times greater than the mass loss of a red giant star, and about 10 billion times the mass loss of the present-day Sun.) The cause of these extreme stellar winds is still a puzzle, although the surface gravity of AGB stars is very low because the stars are so large; thus, any sort of disturbance on the star surface is capable of expelling material outwards. The outer layers of the star flow outward at 10 km per second (about 2% of the speed of the solar wind), cooling as they move from the star. Dust particles can thus form in the cooler surrounding gas out of the ejected carbon-rich molecules. In fact, it is believed that tiny grains of soot are formed! Many carbon stars have been observed, surrounded by cocoons of carbon-rich matter. In some cases, the dust cloud is so thick that it can totally obscure the star, absorbing all the emitted radiation. The dust then heats up and re-emits the energy, but this time in the infrared.

Thought Question 11.1

If stars of about 0.8 $M_\odot$ *where the only type that ever existed in the Milky Way, would you be reading this book?*

11.5. Infrared Stars

It may come as a surprise to know that AGB stars, which can have luminosities 10,000 times that of the Sun, were, until the 1960s, hardly known. The reason for this is simple: the dust that surrounds the star, and re-emits the radiation, is so cool that the reradiated energy is almost entirely in the infrared part of the spectrum. This is, of course, invisible to the naked eye, and it has only been explored in detail in the past 50 years. This is shown schematically in Fig. 11.2. These stars are very faint, or even invisible in the visible part of the spectrum. By comparison, the Sun is very bright in the visible part of the spectrum and very faint in the infrared.

The surface of an infrared star can be thought of as starting at the surface layer of the dust cloud, and for some AGB stars, this can have a radius as much as 500 AU, which is about ten times the size of the Solar System.

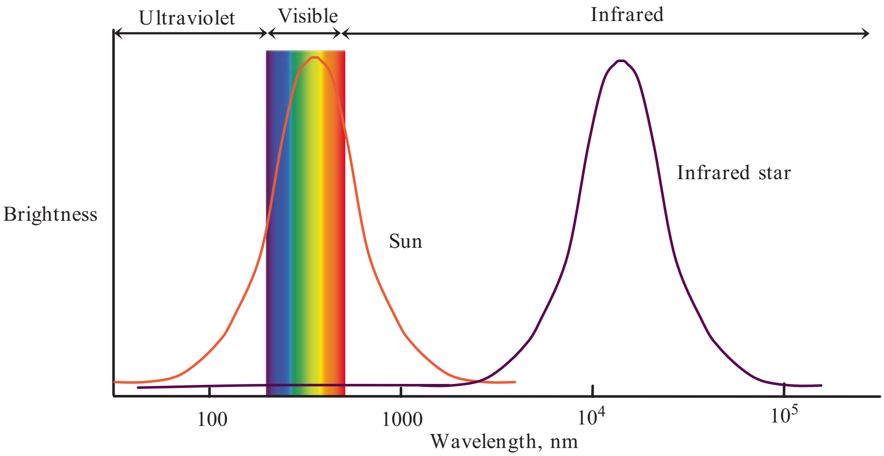


Fig. 11.2. Idealized spectra of an infrared star compared with the Sun

These outer layers of the star are extremely tenuous and hold only a fraction of the star's total mass. The vast majority of the mass is in the carbon-oxygen core and the layers that surround it. So, we can picture an infrared star as having a very small and dense central part and an enormous, low-density outer layer.

Most of the energy emitted by the Sun is in the visible part of the spectrum. In comparison, nearly all of the energy radiated from the dust surrounding an infrared star is invisible to the naked eye.

11.6. The End of an AGB Star's Life

As it ages, an AGB star continues to grow in size and increase its luminosity, along with an increase in the rate at which it loses mass. As mentioned earlier, the mass loss can be 10^{-4} $M_{\odot}$ per year, which means that if the Sun lost mass at this rate, it would only last for 10,000 years. So, obviously, even giant stars cannot carry on this way for very long. If a star has a mass of less than about 8 $M_{\odot}$, its stellar wind will soon strip away the outer layers almost down to the degenerate core. Therefore, a loss of the outer layers would signal the end of the AGB phase. For stars that are greater than 8 $M_{\odot}$, the end of the AGB phase arrives in a much more spectacular event—a supernova—but we leave that discussion for a later section.

We will end this section on the AGB part of a star's life with a sobering and amazing thought. Recall that carbon stars enrich the interstellar medium with not only carbon but some nitrogen and oxygen as well. In fact, carbon can only be formed by the triple α process that occurs in helium burning, and carbon stars are the main means by which the element carbon is dispersed throughout the interstellar medium.

The part that amazes me is when you consider that the carbon in your body, and in fact in the body of every living creature on Earth, was formed many billions of years ago, inside a giant star undergoing the triple α process. It was then dredged up to the star's surface and expelled into space. Later, by some means, it formed the precursor to the Solar System and made the Sun and planets, and all life on Earth. This means…

We are made of the stuff of stars!

One of the benefits of carbon stars, from the point of view of an observer, is that many of them are visible in the night sky with amateur instruments. We have already come across a few of these: R Leporis, RS Cygni, and 19 Piscium. But there are several more that are worth seeking out, and these are listed below. The one aspect of these stars that will be immediately apparent to you is their color; all of them are strongly colored red. They are, in fact, the reddest stars visible to the amateur astronomer, and some have even had their color likened to a drop of blood!

X Cancri	HD 76221	08^h 55.4^m	+17° 14′	February
6.12VM	B-V:2.97	C6		

This is an orange-colored star, and, as with most other carbon stars is a semi-regular variable star, classification SRB, has a period of 180–195 days, and has been observed to range in magnitude from 5.6 to 7.5.

V Hydrae	Lalande 16	10 h 51.6 m	−21° 15′	March
7.0VM	B-V:4.5	C9		

This star, another classic long-period variable, has a period about 533 days and varies in brightness between 6 and 12 magnitudes. It also has a second periodicity of 18 years. One of the rare carbon stars that is visible in amateur instruments, its color has been described as a "magnificent copper red." Note however it is difficult to observe owing to its large magnitude range.

La Superba	ϒ CVn	12 h 45.1 m	+45° 26′	April
5.4VM	B-V:2.9	C7		

The color of this star (red) is best seen through binoculars or a small telescope. With a period of 159 days, and varying in magnitude between 4.9 and 6.0 m, this red giant has a diameter of 400 million kilometers.

V Pavonis	HD 160435	17^h 43.3^m	−57° 43′	June
6.65_v m	B-V:2.45	C5		

A red giant variable star, class SRB, varying in brightness from 6.3 to 8.2 magnitude over a period of 225.4 days. It also has a secondary period of about 3735 days. It has a glorious deep-red color.

RS Cygni	HD 192443	20^h 13.3^m	+38° 44′	July
8.1vm	B-V:3.3	C5		

A red giant star with a persistent periodicity, class SRA, it has a period of 417.39 days, with a magnitude range of 6.5–9.5. A strange star where the light curve can vary appreciably, the maxima sometimes doubling. Another one of the deeply red-colored stars.

Garnet Star	μ Cephei	21^h 43.5^m	+58° 47′	August
4.08vm	B-V:2.26	M2Ia		

Located on the northeastern edge of the nebulosity IC1396, the Garnet Star, or μ Cephei, named by William Herschel, is one of the reddest stars in the entire sky. It has a deep orange or red color seen against a backdrop of faint white stars. It is a pulsating red supergiant star, with a period of about 730 days, varying from 3.4 to 5.1 magnitude. It is in the last stages of its life and could go supernova anytime now, relatively speaking, of course. Another added bonus is that it is one of the largest stars visible to the naked eye, with a radius 1000 times that of the Sun.[3]

R Sculptoris	HD 8879	01^h 26.9^m	−32° 33′	October
5.79vm	B-V:1.4	C6		

A semi-regular-period variable star, with a period of between 140 and 146 days; it varies in brightness from 5.0 to 6.5.

U Camelopardalis	–	03^h 41.8^m	+62° 39′	November
8.3$_v$m	B-V:4.9	N7		

A semi-regular variable star, period 412 days with a magnitude range of 7.7–9.5 m. It has a very deep red color.

Hind's Crimson Star	R Leporis	04^h 59.6^m	−14° 48′	December
7.71vm	B-V:3.4	C7		

The star, a classic long-period variable, period about 432 days, varies in brightness between 6.0 and 9.7 magnitude. At maximum brightness it displays the famous ruddy color that gives it its name. Discovered in 1845 by

---

[3] Recent research suggests it may have a radius ranging from 650 to 1420 times that of the Sun, so choosing a figure of 1000 seems reasonable.

J. R. Hind with a color described as "intense smoky red," many amateurs regard this to be the reddest star in the entire sky.

R Corona Borealis	HD 141527	15h 48.6m	+28° 09′	May
5.89vm	B-V: 0.608	G0Iab:pe		Corona Borealis

Although not strictly a carbon star, R Cor Bor, as it is affectionately known, should be mentioned here. It is the prototype variable star of the class RCB. But what makes this star so special is that it is an irregular variable, usually seen at maximum brightness, but then suddenly fading down to twelfth magnitude, which can last for several weeks, months, or even as long as a year.[4] Then, just as suddenly, it can return to its normal brightness. The reason for this strange behavior is that carbon grains condense out in the star's atmosphere, thus blocking out the light from the star. Radiation then causes the grains to dissipate, and so the star returns to its usual magnitude. The cycle then begins again with the grains building up over time. Other stars that show a similar behavior are RY Sagittarii (6.5 m), SU Tauri (10 m) and S Apodis (10 m).

Thought Question 11.2

Why do you think astronomers initially had significant difficulties when trying to correlate the spectra of carbon stars to their effective temperatures?

11.7. Planetary Nebulae

At the end of the AGB phase, all that will remain of the star is the degenerate core of carbon and oxygen, surrounded by a thin shell in which hydrogen burning occurs. The dust ejected during the AGB phase will be moving outward at tens of kilometers per second. As the debris moves away, the hot, dense and small core of the star will become visible. The aging star will also undergo a series of bursts in luminosity, and during each burst, eject a shell of material into interstellar space. The star now begins to move rapidly towards the left of the H-R diagram, at an approximately constant luminosity but an increasing temperature. It will only take, say, a few thousand years for the surface temperature to reach 30,000 K. Some stars achieve temperatures of 100,000 K. At these high temperatures, the exposed core of

---

[4] On one occasion, it remained at minimum magnitude for 10 years!

the star will emit prodigious amounts of ultraviolet radiation, which can excite and ionize the expanding shell of gas. The shell of ionized and heated gas will begin to glow and produce what is called a planetary nebula.

We know that as the helium in the helium-burning shell is depleted, the pressure that supports the dormant hydrogen-burning shell decreases. Therefore, the hydrogen-burning shell contracts and heats up, thereby initiating hydrogen burning. This newly started hydrogen burning creates helium, which falls down upon the temporarily dormant helium-burning shell. If the shell temperature reaches a specific value, it reignites in what is called the helium-shell flash, similar to (but less intense than) the helium flash that occurs in the evolution of low-mass stars. The newly created energy pushes the hydrogen-burning shell outward, cooling it as it does so, which results in a cessation of the hydrogen burning, and the shell becomes dormant once again. The process then starts all over again.

The luminosity of the AGB star increases quite substantially when the helium shell flash occurs, although it is only for a relatively short time. This short-lived burst is called a *thermal pulse*. After a thermal pulse has occurred, the star resumes its former appearance until enough helium builds to allow another thermal pulse to occur. With each thermal pulse, the mass of the degenerate core, consisting of carbon and oxygen, will increase.

For very massive stars, the thermal pulse occurs in the very deep interior of the star and produces only a slight, temporary change in luminosity. For a star of mass 1 $M_{\odot}$, a thermal pulse would be close enough to the surface to cause the luminosity to increase by a factor of 10 and last about 100 years. The time between thermal pulses varies depending on the star's mass, but calculations predict that they would occur at ever-decreasing intervals, perhaps as short as 100,000–300,000 years, while the luminosity of the star during this time would slowly increase overall.

Significant mass loss can also occur during thermal pulses. A star's outer layers can separate completely from the carbon- and oxygen-rich core, and as the ejected material disperses into space, grains of dust can condense out of the cooling gas. The radiation from the very hot core can propel the dust grains farther and so the star sheds its outer layers completely. In this manner, a 1 $M_{\odot}$ star can lose about 40% of its mass. Even more mass is lost by the massive stars. As the dying star loses its outer layers, the hot core is exposed, and it illuminates the surrounding dust and gas cloud.

The evolution of the remaining core is itself of interest as it progresses rapidly to its final state. There are two factors that can influence the rate at which the core evolves. First, due to the star's extreme luminosity (which can be as high as 100,000 $L_{\odot}$), it consumes its hydrogen at a very fast rate. Second, little hydrogen remains in the thin hydrogen-burning shell that surrounds the degenerate core, so there is hardly any fuel left to be consumed. The central stars of some planetary nebulae have as little as a few millionths of a solar mass of hydrogen left to burn, and so they fade very rapidly. In fact, some can have their luminosities decrease by as much as 90% in as little as 100 years, whereas others may require a bit longer, perhaps a few thousand years. As the source of ionizing photons decreases over time, the planetary nebula grows darker and eventually fades away.

Planetary nebulae[5] are some of the most interesting and beautiful objects in the sky, and they have a lot to offer the amateur. They range across the whole of the observational spectrum. Some are easy to find in binoculars, while others require a large aperture, patience and maybe even specialized filters to be distinguished from the background star fields. These small shells of gas, once the atmospheres of stars, come in a variety of shapes, sizes, and brightnesses. Many have a hot central star within the nebula, which is visible in amateur equipment and is the power source, providing the energy for the gas to glow.

Several nebulae have a multiple-shell appearance, and this is thought to be due to the red giant experiencing several periods of pulsation where the material escapes from the star. The strong stellar winds and magnetic fields of the star are also thought to be responsible for the many observed exotic shapes of the nebulae. Planetary nebulae are only a fleeting feature in our galaxy; after only a few tens of thousands of years, they will have dissipated into interstellar space and so no longer exist. Thus, the planetary nebulae we observe today cannot be older than about 60,000 years. However, this aspect of a star's evolution is apparently very common, and there are more than 1400 planetary nebulae in our part of the galaxy alone!

Visually, the nebulae are one of the few deep-sky objects that actually appear colored. About 90% of their light comes from the doubly ionized oxygen line, OIII, at wavelengths 495.9 nm and 500.7 nm. This is a very characteristic blue-green color, and it so happens, the color at which the dark-adapted eye is at its most sensitive. The specialized light filters are also extremely useful for observing planetaries, as they isolate the OIII light in particular, increasing the contrast between the nebula and the sky background, thus markedly improving the nebula's visibility.

Such is the variety of shapes and sizes that there is something to offer all types of observers. Some planetaries are so tiny that even at high magnification, using large-aperture telescopes, the nebulae will still appear star-like. Others are much larger. For instance, the Helix Nebula, Caldwell 63, is half

---

[5] The name "planetary nebulae" was first applied to these objects by Herschel, who thought that the nebula looked like Jupiter when seen in a telescope.

the size of the full Moon but can only be observed with low magnification, and perhaps only in binoculars, as any higher magnification will lower its contrast to such an extent that it will simply disappear from view. Many, such as the Dumbbell Nebula, M27, in Vulpecula, exhibit a bipolar shape. See Photo 11.1. Still others show ring shapes, such as the ever-popular Ring Nebula, M57, in Lyra.

Photo 11.1. Dumbbell Nebula, Messier 27

An interesting aside is the possibility of observing the central stars of the nebulae. These are very small subdwarf and dwarf stars. They are similar to main sequence stars of types O and B, but, as they are running down their nuclear reactions, or in some cases, no longer producing energy by nuclear reactions, they are consequently fainter and smaller. These two characteristics make observation very difficult. The brightest central star is possibly that of NGC 1514 in Taurus, at 9.4 magnitude, but the majority are at magnitude 10 or fainter.

An image of a planetary nebula can be seen at the end of the chapter.

The Vorontsoz-Vellyaminov classification system can be used to describe the appearance of a planetary. Although it is of limited use, it will be used here.

Planetary Nebulae Morphology Types

1. Starlike
2. Smooth disc-like appearance
    - (a) bright towards center
    - (b) uniform brightness
    - (c) possible faint ring structure
3. Irregular disc-like appearance
    - (a) irregular brightness distribution
    - (b) possible, faint ring structure
4. Definite ring structure
5. Irregular shape
6. Unclassified shape
    - (i) *can be a combination of two classifications,
    - (ii) (e.g., 4 + 3, ring and irregular disc)

The usual information is given for each bright planetary nebula listed below, with the addition of morphology class [⊙] and central star brightness [✸]. In addition, the magnitude quoted is the magnitude of the planetary nebula as if it were a point source. This last parameter can often be confusing, so even if a nebula has a quoted magnitude of, say, 8, it may be much fainter than this and, consequently, hard to find. Finally, bear in mind that the size of the nebula is that observed in a telescope, and not the size measured from images taken with the Hubble Space Telescope or the Spitzer infrared telescope!

Caldwell 39	NGC 2392	$07^{h}\ 29.2^{m}$	+20° 55′	January
10.1 m	⊕ 47 | 43″	⊙ 3B + 3B	✸ 9.8	

This is a small but famous planetary nebula, which can be seen as a pale blue dot in a telescope of 10 cm, although it can be glimpsed in largish binoculars as the apparent southern half of a double star. Higher magnification will resolve the central star and the beginnings of its characteristic "Eskimo" face. With an aperture of 20 cm, the blue disc becomes apparent. Research indicates that we are seeing the planetary nebula pole-on, although this is by no means certain. Its distance is also in doubt, with values ranging from 1600 to 7500 light years. Also known as the Eskimo Nebula.

Caldwell 59	NGC 3242	$10^{h}\ 24.8^{m}$	−18° 38′	February
8.6 m	⊕ 45 | 36″	⊙ 4 + 3 m	✸ 12.1	

One of the brighter planetary nebulae and the brightest in the spring sky for northern observers, this is a fine sight in small telescopes. Visible in binoculars as a tiny blue disc. With an aperture of 10 cm, the blue color becomes more pronounced along with its disc, which is approximately the same size as that of Jupiter in a similar aperture. The central star has a reported temperature of about 100,000 K. Also known as the Ghost of Jupiter.

Messier 97	NGC 3587	$11^{h}\ 14.8^{m}$	+55° 01′	March
9.9 m	⊕ 202 | 196″	⊙ 3 m	✸ 15	

Sadly, this is not visible in binoculars due to its low surface brightness; apertures of at least 20 cm will be needed to glimpse the "eyes" of the nebula. At about 10-cm aperture, the planetary nebula will appear as a very pale, blue-tinted circular disc, although the topic of color in regard to this particular planetary nebula is in question. Also known as the Owl Nebula.

Caldwell 6	NGC 6543	$17^{h}\ 58.6^{m}$	+66° 38′	June
8.1 m	⊕ 23 | 17″	⊙ 3A + 2	✸ 11	

Seen as a bright oval planetary nebula with a fine blue-green color, this is one of the planetary nebulae that became famous after the HST published its image. Visible even in a telescope of 10 cm, but a large telescope (20 cm) will show some faint structure, though to observe the central star requires a 40-cm aperture. The incredibly beautiful and complex structure is thought to be the result of a binary system, with the central star classified as a Wolf-Rayet star. Also known as the Cat's Eye Nebula.

Messier 57	NGC 6720	$18^{h}\ 53.6^{m}$	+33° 02′	July
8.8 m	⊕ 86 | 62″	⊙ 4 + 3	✸ 15.75	

Probably the most famous of all planetary nebulae, the Ring Nebula is, surprisingly, visible in binoculars. However, it will not be resolved into the famous “smoke-ring” shape seen so often in color photographs; rather, it will resemble an out-of-focus star. It is just resolved in telescopes of about 10-cm aperture, and at 20 cm the classic smoke-ring shape becomes apparent. At high magnification (and larger aperture), the Ring Nebula is truly spectacular. The inner region will be seen to be faintly hazy, but a large aperture and perfect conditions will be needed to see the central star.

Messier 27	NGC 6853	$19^h 59.6^m$	+22° 43′	July
7.5 m	⊕ 8 | 5.7′	⊙ 3 + 2	✹ 13.8	

Also known as the Dumbbell Nebula, this famous planetary nebula can be seen in small binoculars as a box-shaped hazy patch, and many amateurs consider it the sky’s premier planetary nebula. In apertures of 20 cm, the classic dumbbell shape is apparent, with the brighter parts appearing as wedge shapes that spread out to the north and south of the planetary nebula’s center, and a central star may be glimpsed. As an interesting aside, the central star, which is a white dwarf, is believed to have a radius which is 0.055 ± 0.02 that of the Sun. This means it is the largest currently known white dwarf. See Photo 11.1.

Caldwell 63	NGC 7293	$22^h 29.6^m$	−20° 50′	August
7.6 m	⊕ 880 | 720″	⊙ 4 + 3	✹ 13.5	

Thought to be one of the closest planetary nebulae to Earth, at about 700 light years, it has an angular size of over 1/4°—half that of the full Moon. However, it has a very low surface brightness and is thus notoriously difficult to locate. With an aperture of 10 cm, low magnification is necessary, and averted vision is useful for glimpsing the central star. The use of an OIII filter will drastically improve the image. This was the first planetary nebula discovered to contain cometary knots, which are highly radially symmetric blobs of material and, are described as “cometary” in appearance. It has been estimated that there are over 20,000 knots in the nebula. Also known as the Helix Nebula.

Caldwell 22	NGC 7662	$23^h 25.9^m$	+42° 33′	September
8.6 m	⊕ 32 | 28″	⊙ 4 + 3	✹ 13.2	

This nice planetary nebula is visible in binoculars due to its striking blue color, but it will only appear stellar-like. Research indicates that the planetary nebula has a structure similar to that seen in the striking HST image of the Helix Nebula, showing Fast Low-Ionization Emission Regions (FLIERS). These are clumps of above-average density gas ejected from the central star before it formed the planetary nebula. Also known as the Blue Snowball.

Messier 76	NGC 650	$01^h\,42.4^m$	+51° 34′	October
10.1 m	⊕ 163 | 107″	⊙ 3 + 6	✹ 15.9	

This is a small planetary nebula that shows a definite non-symmetrical shape. In small telescopes of aperture 10 cm, and using averted vision, two distinct “nodes” or protuberances can be seen. With apertures of around 30 cm, the planetary nebula will appear as two bright but small discs that are in contact. Even larger telescopes will show considerably more detail. Also known as the Little Dumbbell Nebula.

Herschel 53	NGC 1501	$04^h\,07.0^m$	+60° 55′	November
11.5 m	⊕ 56 | 48″	⊙ 3	✹ 14.4	

A very nice blue planetary nebula, easily seen in telescopes of 20 cm and glimpsed in apertures of 10 m. With a larger aperture, some structure can be glimpsed, and many observers liken this planetary nebula to that of the Eskimo Nebula. Has been called the Oyster Nebula.

PK221–12.1	IC 2165	$06^h\,21.7^m$	−12° 59′	December
10.5 m	⊕ 28″	⊙ 3B	✹15	

This is another planetary nebula that needs a high magnification in order to resolve its non-stellar properties. With aperture of 20 cm, the small, faintly blue disc can be seen but using larger apertures will resolve the non-circular shape along with a slight brightening at its center.

Thought Question 11.3

Astronomers now think the Planetary Nebulae have, more-or-less the same shape, but when images are taken, they exhibit many different shapes. Can you think of a reason why?

11.8. White Dwarf Stars

We now look at the endpoint for low-mass stars, and it is a very strange end indeed. We have seen that stars with a mass of less than around 4 $M_\odot$ never manage to produce the internal pressure and temperature necessary to provide the means to burn the carbon and oxygen in the core. What happens instead is an ejection of the star’s outer layers, leaving behind the very hot carbon-oxygen-rich core. In such a scenario, the core has stopped producing energy by nuclear fusion and so just cools down, admittedly over a vast time scale. These cooling relics are the white dwarf stars. In many instances, they are no bigger than Earth.

11.8.1 Electron Degeneracy

Experience tells us that as the mass of an object increases, so does its size, and this applies for many astronomical objects, such as stars on the main sequence. However, the opposite is true for white dwarfs. The more massive a white dwarf star, the smaller it becomes. This contrary behavior has to do with the electron structure of the material of the white dwarf. Increasing the density of an object will lead to an increase in pressure, as observed in main sequence stars, but the pressure in a white dwarf star (which is, remember, the core of a once much larger star) is produced by degenerative electrons.[6] This electron degenerate pressure supports the star. An increase in density, however, also leads to an increase in gravity. For the white dwarf star, this increased gravity will exceed the increase in pressure, and so the star will contract. As it gets smaller, both the gravity and pressure increase further and come into balance with each other, but at a smaller size for the white dwarf. This means the more massive a white dwarf star, the smaller it is. As an example, a 0.5 $M_\odot$ white dwarf star is about 90% larger than Earth, whereas a 1 $M_\odot$ white dwarf star is only about 50% larger than Earth. If the white dwarf is 1.3 $M_\odot$, then it is only 40% as large as Earth.

11.8.2 The Chandrasekhar Limit

White dwarf stars have a very unusual mass-radius relationship, shown in Fig. 11.3. As you can see, the more degenerate matter you put into a white dwarf star, the smaller it gets. However, you cannot do this ad infinitum, as there is a maximum mass that a white dwarf can have. This mass, which is about 1.4 $M_\odot$, is called the *Chandrasekhar limit*, named after the Indian astrophysicist who was the first to pay serious attention to the behavior of white dwarfs. It is the mass for which the mass-radius relationship drops to zero, so that a white dwarf star with a mass equal to the Chandrasekhar limit will shrink to a very small size.

---

[6] See Appendix 1 for a full explanation of degeneracy.

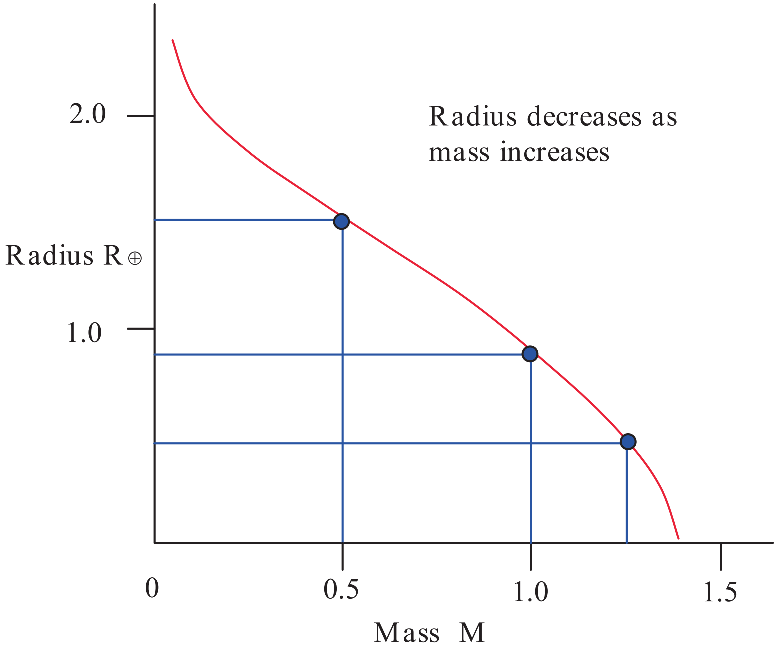


Fig. 11.3. Mass-radius relationship for white dwarf stars. The radius is given in terms of Earth's radius. The more massive a white dwarf, the smaller it will be. Note that on this graph, the size of a white dwarf will fall to zero if it has a mass of 1.4 $M_\odot$

However, no star with a mass greater than 1.4 $M_\odot$ can be supported against the crush of gravity by the pressure of the degenerate electrons. This means that the main sequence stars of type O, B and A, which have masses greater than the Chandrasekhar limit, will need to shed mass if they are to become white dwarf stars. This they do while becoming AGB stars, as we saw earlier. But not all stars do achieve the necessary mass loss, and in such cases where the contraction cannot be stopped by degenerate electrons, the stars collapse even further to become neutron stars and perhaps even black holes.

Before we leave the Chandrasekhar limit, it is important to note that maybe the limit isn't the last word on white dwarf masses.

Recent observations seem to suggest that a few white dwarfs apparently had a mass greater than 1.4 $M_\odot$, which would contradict the Chandrasekhar limit. Notice I used the word *had*, this is because what was actually detected were white dwarf supernova (see Chap. 12). The first candidate was detected in 2003, by the Supernova Legacy Survey.[7] It was a type Ia supernova, designated SNLS-03D3bb, found in a galaxy approximately 4 billion light years away. The data suggested that the origin of the supernovae was a white dwarf that had somehow accumulated twice the mass of the Sun before exploding and given the name the "Champagne Supernova".

Two possible solutions to this higher mass anomaly were proposed.

1. it may have been spinning so fast that a centrifugal trend may allow it to exceed the Chandrasekhar limit, with the added support effectively increasing the critical mass.
2. the supernova may have been the result of the merger of two white dwarfs, so that the limit was only violated for a very short time.

A further idea was proposed to explain the problem, in that the explosion was not spherical, expanding outward in a symmetrical manner, but rather, asymmetrical, exploding preferentially in one direction. However, spectroscopic analysis of the data discounted this idea.

Several more type 1a supernova have been detected, whose progenitors were white dwarfs that had a mass great than 1.4 $M_\odot$. These *Super-Chandrasekhar mass white dwarfs* are thought to have had masses up to 2.4–2.8 $M_\odot$.

These apparent violations of the Chandrasekhar Limit are more than just an anomaly, however. White Dwarf supernovae are used as "standard candles" because the brightness of a type Ia supernovae is thought to be broadly uniform and can be used to measure distances in the universe. Thus, an uncharacteristic type Ia supernova could throw distance measurements into question, along with a possible reassessment of the expansion history of the Universe. Further analysis and observations in the future will hopefully explain these apparent contradictions to The Chandrasekhar Limit.

Thought Question 11.4

An astronomer publishes a paper that says he has detected evidence of a massive star, 12 $M_\odot$*, that has ended its life as a white dwarf. Are you convinced?*

---

[7] The Supernova Legacy Survey Program was a project designed to investigate dark energy, by observing approximately 2000 high-redshift supernovae between 2003 and 2008.

11.8.3 White Dwarf Origins

It is now believed that most, if not all, white dwarfs have evolved directly from the central stars of planetary nebulae. These, in turn, are the former cores of AGB stars. We saw earlier that during the AGB phase, a star will lose much of its mass via a cool stellar wind. If the star has sufficient mass stripped away so it is lower than the Chandrasekhar limit, a carbon-oxygen rich core of matter surrounded by a very thin layer of helium-rich gas is the result. In some cases, there may even be a thin outer layer of hydrogen-rich gas. The star and expelled gas are now a planetary nebula, and at the moment nuclear fusion ends, a white dwarf is born.[8] But even though theory matches well with observations, there is still uncertainty as to the mass that the star may have originally had in order to lose enough to become a white dwarf star. Current ideas suggest a mass limit of 8–10$M_\odot$. Those main sequence stars that have a mass of between 2 and 8 $M_\odot$ produce white dwarfs of masses 0.7 and 1.4 $M_\odot$, whereas main sequence stars less than 2 $M_\odot$ will produce white dwarfs of mass 0.6–07 $M_\odot$. If a white dwarf star has a mass less than 0.6 $M_\odot$, the progenitor main sequence star will have a mass less than 1 $M_\odot$.

11.8.4 Composition of White Dwarfs

A question that is often asked is, “What is a white dwarf star made of?” The answer is surprising. The actual composition of a white dwarf will depend on the initial mass of the star. These can be roughly placed into three types.

- *Very low mass stars*: If the mass of a main-sequence star is lower than approximately half a solar mass, it never becomes hot enough to fuse helium in its core. Over a lifetime that significantly exceeds the age of the Universe (around 13.8 billion years), the star will burn all its hydrogen, becoming, for a while, a blue dwarf, and end its evolution as a helium white dwarf composed chiefly of helium-4 nuclei.
- *Low to medium mass stars*: If the mass of a main-sequence star is between 0.5 and 8.0 $M_\odot$(similar to the Sun), its core becomes hot enough to fuse helium into carbon and oxygen via the triple-alpha process, but never becomes necessarily hot enough to fuse carbon into neon. Near the end of

---

[8] Recent observations have detected a star (V4334 Sagittarii) that was well on its way to being a white dwarf when it underwent a final helium flash, grew to red giant size once again, and is now ejecting more gas. Another star that has shown a similar behavior is V605 Aquilae.

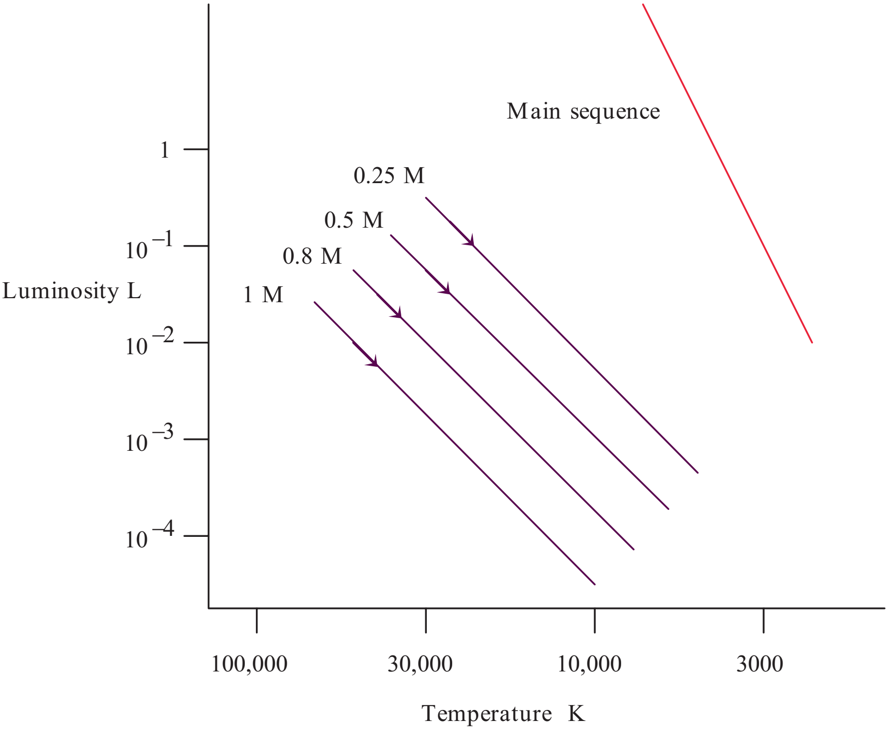


Fig. 11.4. White dwarf evolutionary tracks

this time period, it will have a carbon–oxygen core with no fusion reactions, surrounded by an inner helium-burning shell and an outer hydrogen-burning shell. It then expels most of its outer material, creating a planetary nebula, until only the carbon–oxygen core is left. This process is responsible for the carbon–oxygen white dwarfs which form the vast majority of observed white dwarfs

- *Medium to high mass stars*: A few main-sequence stars, of around 8 and 10 $M_\odot$ although sufficiently massive to fuse carbon to neon and magnesium, may not have enough mass to fuse neon. Such a star may form a white dwarf composed chiefly of oxygen, neon, and magnesium, with the proviso that its core does not collapse, and fusion doesn't occur so violently that blows the star apart in a supernova. A few white dwarfs have been so identified, but most of the evidence for their existence is from novae called *ONeMg* or *neon* novae. Their spectra display abundances of neon, magnesium, and other intermediate-mass elements that can only be explained by the accretion of material onto an oxygen-neon-magnesium white dwarf.

As discussed above, the matter making up the white dwarf star consists mostly of ionized oxygen and carbon atoms, along with a few oxygen-neon-magnesium stars, which are floating in a sea of fast-moving degenerate electrons. As the star continues to cool, the particles in this matter slow down, resulting in electric forces between the ions beginning to dominate the random thermal motions they may have originally had. These ions no longer move freely though the white dwarf but instead are aligned in orderly rows, rather like a giant crystal lattice. It is appropriate to think of the white dwarf as being "solid," with the degenerate electrons still moving freely in the crystal lattice, just as electrons move in, say, a copper wire.

Another interesting thing to note is that a diamond is a crystal lattice of carbon, so a cooling white dwarf star can also be thought of as a (sort of) giant spherical diamond. The density in a white dwarf star is immense, typically 10^9 kg m^{-3}. This is about one million times the density of water. One of the statistics astronomers like to throw out is that one teaspoon of white dwarf matter weighs about 5.5 tons, equal to the weight of an elephant—providing, of course, that you could get a teaspoon of the matter to Earth in the first place.

11.8.5 White Dwarf Evolution

When the white dwarf shrinks to its ultimate size, it will no longer have fuel available for nuclear fusion. It will, however, still have a very hot core and a large reservoir of residual heat. For example, the surface temperature of one famous white dwarf star, Sirius B, is about 30,000 K. Time will pass, and with it, the white dwarf will cool down as it radiates its heat into space. As it does so, it will also grow dimmer, as shown in Fig. 11.4, where white dwarf stars of differing mass are plotted on an H-R diagram. The more massive a white dwarf star, the smaller its surface area vs. less-massive white dwarfs. This means that massive white dwarfs are less luminous for a given temperature, so their evolutionary tracks are below those of the less-massive white dwarf stars.

Theoretical models of the evolution of white dwarfs have been constructed, and they show that a white dwarf with a mass of 0.6 $M_\odot$ will fade to 0.1 $L_\odot$ in about 20 million years. Any further reductions in luminosity take progressively longer amounts of time. This means that it will take 300 million years to fade to about 0.01 $L_\odot$ and a billion years to get to 0.001 $L_\odot$. It will take about 6 billion years for the white dwarf to reach a luminosity of 0.0001 $L_\odot$. At this point, the white dwarf will have the same temperature and color as the Sun. It will be so faint, however, that unless it was within a few parsecs of Earth, it would be undetectable. Those white dwarf stars with masses greater than 0.6 $M_{\odot}$ have more internal heat and so will take even longer to cool down and grow faint.

In the case of the Sun, it will eject most of its mass into space and eventually end up about the same size as Earth, but its luminosity will change dramatically, perhaps only achieving one-tenth the brightness it presently has. As it ages further, it will continue to grow even fainter. When about 5 billion years have passed, the Sun will only be able to achieve one ten-thousandth of its present luminosity. As time passes into the unimaginable future, it will simply fade from view!

A white dwarf will cool and grow fainter, so it moves downwards and to the right on the H-R diagram. The more massive a white dwarf, the smaller and fainter it will be. Therefore, the track for a 1 $M_{\odot}$ white dwarf will lie below the track for less-massive white dwarfs. Note that although a white dwarf may have the same temperature as a main sequence star, it will be fainter because it is small and thus has a smaller surface area.

What is incredible about these lower-mass stars is that their main sequence lifetimes are so incredibly long, the universe is not yet old enough for them to have evolved into white dwarf stars. This means that there are no white dwarf stars with a mass less than about 0.6 $M_{\odot}$. The timescale for the evolution from giant star to white dwarf can take between 10,000 and 100,000 years.

11.8.6 Spectral Types of White Dwarfs

Just as there are spectral classifications for Main Sequence and Giant stars, there is also a dedicated spectral classification for white dwarfs, with a classification D (which stands for Degenerate). These white dwarfs can then be subdivided into subdivisions.

- DA- The atmosphere of these stars is abundant with hydrogen, which is indicated by the presence of Balmer hydrogen in their spectral lines.
- DB- The atmosphere of these stars is abundant with helium, which is indicated by the presence of neutral helium (He I) in their spectral lines.
- DO- The atmosphere of these stars is abundant with helium, which is indicated by the presence of ionized helium (He II) in their spectral lines.
- DQ- The atmosphere of these stars is abundant with carbon, which is indicated by the presence of atomic or molecular carbon in their spectral lines.

- DZ- The atmosphere of these stars is abundant with metals, which is indicated by the presence of metal in their spectral lines.
- DC- There is no indication of any of the categories above, as the spectral lines are not strong.
- DX- Spectral lines are not clear enough to categorize.

Due to their faintness and small size, white dwarf stars present a challenge to observers. There are, of course, many of them in the night sky, and those amateur astronomers with large telescopes of, say, aperture 25 cm and larger will have no problem locating and observing them. On the other hand, there are a handful that, given the right conditions, can be seen with much more modest instruments. These are the ones outlined below. The symbol ⊕ indicates the size of the white dwarf star as compared with Earth (thus, $0.5_{\oplus}$ would mean that it is half the size of Earth).

Let's look at a few White Dwarfs.

Sirius B		$06^h\ 45.1^m$	−16° 43′	January
8.3 m	11.2 M	0.92⊕	27,000 K	

The companion star to the brightest star in the sky (Sirius) is a white dwarf known as the Pup, the first ever to be discovered. It is a difficult, though not impossible, star to observe for two main reasons. First, the dazzling primary star overcomes it, and so the light from Sirius often needs to be blocked out by some means. In fact, if Sirius B were not a companion to Sirius, it would be easily visible in binoculars. Second, its orbit changes over a period of 50 years. This means that at certain times it will be too close to Sirius to be detected with amateur instruments. The next time it will be at maximum separation is the year 2025.

Procyon B	α Canis Minoris	$07^h\ 39.3^m$	+05° 13′	January
10.7 m	13.2 M	1.05⊕	8700 K	

This dwarf star is not easily visible in small amateur telescopes, having a magnitude of 10.8 and a mean separation of 5 arc seconds. Note that it has a low temperature compared with other white dwarfs. It is the second-closest white dwarf to Earth. Definitely a challenge to observers.

Van Maanen's Star	Wolf 28	$00^h\ 49.1^m$	+05° 23′	September
12.4 m	14.1 M	0.9(?)⊕	6000 K	

One of the few white dwarfs stars visible to amateurs, it is the closest solitary white dwarf, at only 13.9 light years distant. It is located about 2° south of δ (delta) Piscium. Discovered by A. Van Maanen in 1917 due to its large proper motion of 2.98 arc seconds per year, it is the third closest white dwarf to us.

o^2 ERADANI	40 ERIDANI B	04^h 15.2^m	−07° 39′	NOVEMBER
9.5 m	11.0 M	1.48⊕	14,000 K	

Thought Question 11.5

You observe many white dwarfs, and none of them are surrounded by a planetary nebula. Can you explain why?

Even though this is a challenge to split with binoculars, it is nevertheless the easiest white dwarf star to observe. The star will be in a prime observing position relative to its brighter primary star for the next 50 years or so. What makes this system so interesting is that the secondary is the brightest white dwarf star visible from Earth. In addition, under high magnification, it will be seen to have a companion star of its own—a red dwarf star! Thus, it is a nice triple-star system.

1. A binary system consisting of a 10 $M_{\odot}$main sequnce star, and a white dwarf at the Chandrasekhar Limit has an orbital period of 5 years, along with an average semimajor axis of 7 AU. Make an estimate of the mass of the main sequence star in solar units.
2. A binary system consists of two identical stars, with nearly identical spectra. One of the stars is a white dwarf, The system has an orbital period of 5 years, along with an average semimajor axis of 4.12 AU. Calculate their combined mass.
3. What type of star do you think the companion star in question 3 is?

Chapter 12

Star Death: Supernovae, & Neutron Stars

12.1. High-Mass Stars and Nuclear Burning

Throughout the entire life of a low-mass star (that is, one that is less than 4 $M_\odot$), only two nuclear reactions occur: hydrogen burning and helium burning, and the only elements besides hydrogen and helium that are formed are carbon and oxygen. Stars that have a zero-age mass greater than 4 $M_\odot$ begin their lives in a similar manner, but theory predicts that due to the increased mass, and therefore higher temperatures involved, other nuclear reactions will occur. The tremendous crush of gravity is so overwhelming that degeneracy pressure is never allowed to come into play. The carbon-oxygen core is more massive than the Chandrasekhar limit of 1.4 $M_\odot$, and so the degenerate pressure cannot stop the core from contracting and heating.

The nuclear reactions that take place in the star's final phase of its life are very complex, with many different reactions occurring simultaneously. But the simplest sequence of fusion involves what is termed *helium capture*; this is the fusing of helium into progressively heavier elements.[1] The core

---

[1] Some helium nuclei do remain in the star's core, but these are not in sufficient number to initiate helium burning to any great degree.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_12

continues to collapse with an accompanying rise in temperature to about 600 million K. At this high temperature, the helium capture can give rise to *carbon burning*, and the carbon can be fused into heavier elements. The elements oxygen, neon, sodium and magnesium are produced. The carbon fusion provides a new source of energy that, at least temporarily, restores the balance between pressure and gravity.

If the star, however, has a mass greater than 8 $M_{\odot}$ even further reactions can occur. In this phase, the carbon burning may only last a few hundred years. As the core contracts further, the core temperature reaches 1 billion K, and *neon burning* begins. In this manner, the neon produced by the earlier carbon-burning reaction is used up, but at the same time, there is an increase in the amount of oxygen and magnesium in the star's core. This reaction lasts as little as 1 year. As you can imagine, with each stage of element burning, higher temperatures are reached, and further reactions occur; *oxygen burning* will occur when the temperature reaches 1.5 billion K, with the production of sulfur. *Silicon burning* can also occur if the core temperature reaches the staggering temperature of 2.7 billion K. This reaction produces several nuclei, from sulfur to iron.

Despite the very dramatic events that are occurring inside the high-mass star, its outward appearance changes only slowly. When each stage of core nuclear fusion stops, the surrounding shell burning intensifies and therefore inflates the star's outer layers. Then, each time the core flares up again and begins further reactions, the outer layers may contract slightly. This results in the evolutionary track of the star zigzagging across the top of the H-R diagram.

Some of the reactions that occur also release neutrons, which are particles similar to protons only they do not have an electric charge. This neutrality means that they can, and do, collide with positively charged nuclei and combine with them. The absorption of neutrons by nuclei is termed *neutron capture*. In this way, many elements and isotopes that are not produced directly in the fusion reactions are produced.

Each stage during this phase of a high-mass star's life helps to initiate the subsequent phase. As each phase ends due to the star's using up the specific fuel in its core, gravity will cause the core to contract to an ever-higher density and temperature, which in turn is responsible for starting the next phase of nuclear burning. In effect, you can think of each stage burning the "ash" of the previous one.

An interesting point here is that we tend to think of astronomical events taking place over many millions of years. However, theoretical calculations have shown that when we are dealing with high-mass stars, events can proceed at a very fast pace, with each successive stage of nuclear burning proceeding at an ever-increasing rate. One calculation has been made in detail for 20–25 $M_\odot$ zero-age stars, and the results are very surprising. The carbon-burning stage can last for about 600 years, while the neon-burning stage can be as short as 1 year. Then things start to speed up! The oxygen burning lasts only 6 months, and the silicon burning only a day!

At each phase of core burning, a new shell of material is formed around the core of the high-mass star, and after several such stages, in, say, a very massive star of 20–25 $M_\odot$, the internal structure of the star can resemble an onion (shown in Fig. 12.1).

Nuclear reactions are taking place in several different shells simultaneously, and the energy released does so at such a rapid rate that the star's outer layers can expand to an immense size. The star can now be called a *supergiant* star. The luminosity and temperature of such stars are much higher than those of a mere giant star.

Many of the brightest stars in the night sky are supergiants, including several we met earlier in the book. These include Rigel and Betelgeuse in Orion, and Arcturus in Scorpius. Rigel has a temperature of 11,000 K, while Betelgeuse is only 3700 K (or even cooler) and is an example of a red supergiant. Thus, although Betelgeuse is cooler, it must be correspondingly larger for it to be as bright as Rigel. Oddly enough, red supergiants are rare, perhaps even rarer than O-type stars. One current estimate predicts that there is only one red supergiant star for every million stars in the Milky Way, and only about 200 have ever been studied.

What makes these stars stand out is their immense size. The radius of Betelgeuse has been measured to be about 700 times that of the Sun, or 3.6 AU. This can be better appreciated by thinking of it this way: If it were

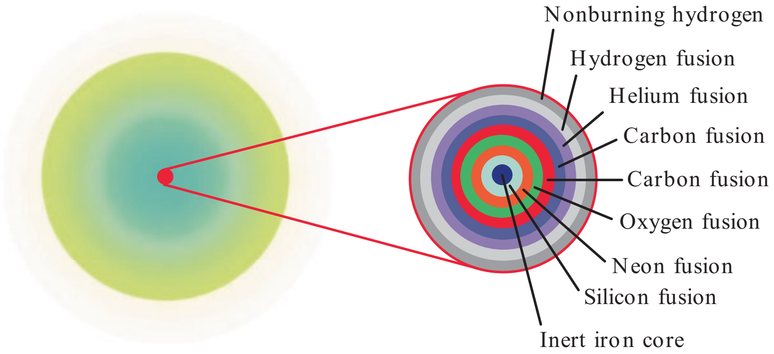


Fig. 12.1. The multiple-layer structure of an old high-mass star

placed in the Solar System, it would extend past the Asteroid Belt to about half-way between the orbits of Mars and Jupiter. Antares would extend nearly to Jupiter! Alpha Herculis is only 2 AU in radius. The record, however, must go to VV Cephei, which is an eclipsing binary star. Its radius is a staggering 1900 times that of the Sun, or 8.8 AU. This means that it would nearly extend to Saturn.

As was mentioned earlier, supergiant stars are quite rare, but fortunately for us observers, there are some we can see with the naked eye. Some are listed below.

CE Tauri	HD36389	05^h 32.2^m	−58° 58′	December
4.32 m	−4.77 M	M2 Iab-lb		

Also known as 119 Tauri, this star has a radius of 300 times that of the Sun and lies about 1800 light years from us. It has the odd distinction of being classified as both a semi-regular and irregular variable star, meaning it is an erratic variable star, and so its period is difficult to predict with any certainty. It lies within a field of stars of similar brightness, which makes it difficult to locate unless a good star atlas is handy. From an observer's point of view, it has a very red color and has been named the Ruby Star.

Other stars that are supergiants and were mentioned in earlier sections are: Mu Cephei, Eta Persei, ψ^{-1}Aurigae, VV Cephei, Antares, Betelgeuse, Mira, La Superba, Deneb, Rigel, Canopus and Aldebaran.

Before we leave supergiant stars, it should be mentioned that there is a class of stars that are similar to supergiants, and these are the Wolf-Rayet stars. These are very hot, very luminous supergiant stars, similar to O-type stars, but they have very strange spectra that show only emission lines and, strangely enough, no hydrogen lines. Wolf-Rayet stars are believed to be precursors to the formation of planetary nebulae. They are few and far between, with perhaps only 1000 in our galaxy. They experience terrific mass losses, and images from large telescopes show these stars surrounded by rich clouds of ejected material. Fortunately for us, there is a very bright example that can be easily observed.

γ^2 Vel	HD 68273	08^h 09.5^m	−47° 20′	Dec- **Jan** -Feb
1.99_v m	0.05 M	WC 8		Vela

The brightest and closest of all Wolf-Rayet stars, γ^2 *Vel* is an easy double, of colors white and greenish-white.

This aspect of a supergiant's life, whereby several layers of nuclear burning occur, resembling the layers of an onion, cannot go on forever, as there is only a finite amount of material to burn. Thus, a point comes when the high-mass star undergoes yet another change, but this time, with catastrophic consequences. It is star death, but in a spectacular manner—a *supernova*.

12.2. Supernovae and the Formation of Elements

When nuclei in the core collide and fuse, energy is emitted, and it is this energy flowing from the core and surrounding shells of nuclear burning that supports the tremendous weight of material making up the star. The energy is a consequence of the strong nuclear force of attraction between neutrons and protons, or *nucleons*, as they are sometimes called. But you may recall that protons also repel each other by what is called the weak electric force. This has profound consequences for the life of the high-mass star.

Up to this point, energy has been released. (For example, the energy has been in the form of output, but, due to the repulsive effect, if any protons are added to nuclei larger than iron, which has itself 26 protons, then energy must be input to the system.) What this means is that any nuclei greater than and including iron will not release any energy. Therefore, the various stages of nuclear burning end with the production of silicon. After that, iron can be formed, but there will be no release of energy associated with its formation. The result is an iron-rich core that has no nuclear reactions taking place within it.

Of course, surrounding this inert core of iron will be the various shells of nuclear burning.[2] However, this is a state of affairs that cannot continue for very long.

Astronomers use a variety of techniques to find out about the life of a star. Observations are made, and then theoretical models are devised so that they fit the observations. In the case of supernovae, it can be said that most, if not all, of what we know about supernovae comes from theoretical and mathematical calculations. After all, it is not easy to see what is happening in the central regions of a star! You will also see that we are now talking about densities, pressures and velocities that will stagger our comprehension. With this in mind note that the following descriptions of the events in a high-mass star are theoretical predictions, although ones that seem to fit the observations.[3]

During these final days of a star, shells of silicon, oxygen, neon, carbon, helium and hydrogen surround the core of inert iron, in which there are no nuclear reactions taking place. The pressure of its degenerate electrons

---

[2] The entire energy-producing region in the star is now in a volume about the same size as Earth, one million times smaller in radius than the size of the star.

[3] Don't think that astronomers know all there is to know about what gives rise to a supernova. Prior to the famous supernova in 1987 ("SN 1987A"), astronomers believed that only red supergiants could form supernova. They were thrown into some confusion when it was discovered that the progenitor star of SN1987 was a blue supergiant!

supports the core, which can be thought of as essentially a white dwarf star surrounded by the outer layers of a red giant star. But recall there is a limit to the mass of a white dwarf star—the Chandrasekhar limit—and so when the core surpasses this limit, its weight becomes too great to be supported by the degenerate electrons, and it collapses.

A consequence of the core contraction is an increase in the density, which in turn gives rise to a process called *neutronization.* This is a process whereby electrons react with the protons in the iron nuclei to form neutrons, as shown below.

$$e^- + p \rightarrow n + \nu$$

Each neutronization reaction will also produce a neutrino. Now more and more electrons will react with the protons, and so there are fewer left to support the core and resist the compression. This results in a speeding up of the contraction and actually could better be termed a collapse of the core. It only takes about a second for the core to collapse from a radius of thousands of kilometers to about 50. Then, in only a few more seconds, it shrinks down to a 5-km radius. The core temperature also increases during this time to about 100 billion K. The gravitational energy released as a result of the core collapse is equal to the Sun's luminosity for several billion years. Most of this energy is in the form of neutrinos, but some is also in the form of gamma rays, which are created due to the extremely hot core temperature. These gamma ray photons in turn have so much energy that when they collide with the iron nuclei, the nuclei are broken down into alpha particles (which are ^{4}He nuclei). This process is called *photo disintegration*.

After a short interval of time, which is thought to be about 0.25 s, the central 0.6–0.8 $M_\odot$ of the collapsing core will reach a density equal to that of atomic nuclei, that is, some 4×10^{27} kg m^{-3}. At this point, the neutrons become degenerate and strongly resist any further attempts at compression. To get an idea of what this density means, Earth would have to be compressed to a sphere 300 m in diameter. For all intents and purposes, the core of the star can now be thought of as a neutron star, and the innermost part of the core suddenly becomes rigid, causing the contraction to abruptly halt. This innermost part actually rebounds outward and pushes back against the rest of the infalling core, driving it outward in a pressure wave. This is called the *core bounce*, and it is illustrated in Fig. 12.2.

The core also cools at this stage, and this causes the pressure to decrease significantly in those regions that surround the core. If you recall, there is a balancing act between the pressure pushing upward and gravity falling downward, and so a consequence of this reduced pressure is that the

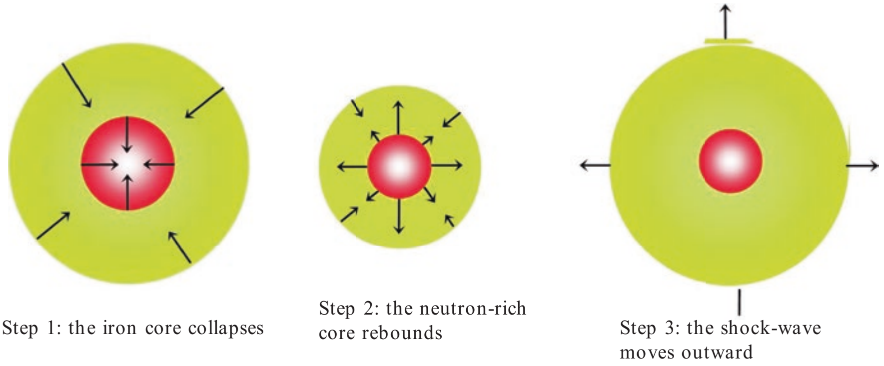


Fig. 12.2. Evolution of a supernova explosion

material surrounding the core now falls inward at a velocity close to 15% of the speed of light. This inward-moving material encounters the outward-moving pressure wave, which, incidentally, can be moving at one-sixth the speed of light. In just a fraction of a second, the falling material now moves back outward towards the star's surface.

Surprisingly enough, this wave of pressure would soon die out long before it reached the surface of the star if it weren't for the fact that helping it along is the immense number of neutrinos that are trying to escape the star's core. The upward-moving wave of pressure speeds up as it encounters the less dense regions of the star and achieves a speed in excess of the speed of sound in the star's outer regions. The pressure wave now behaves like a shockwave.

The neutrinos actually escape from the star in a few seconds, but it takes a few hours for the shockwave to reach the surface. Most of the material of the star is pushed outward by this shock wave and is expelled from the star at many thousands of kilometers per second. The energy released during this event is a staggering 10^{46} J, which is 100 times more than the entire output of the Sun during the last 4.6 billion years. It will surprise you to know that the visible light we observe is only about 1% of the total energy released during the event.

Recent studies have proposed that up to 96% of the material making up the star may be ejected into the interstellar medium that, of course, will be used in future generations of star formation. But before this matter is ejected, it is compressed to such a degree that new nuclear reactions can occur within it, and it is these reactions that form all the elements that are heavier than iron. Elements such as tin, zinc, gold, mercury, lead and uranium, to name but a few, are produced, and this has profound implications because it means that the stuff that makes up the Solar System, Earth, and in fact us, was formed long ago in a supernova.

The expansion of the star's surface due to the shockwave is the cause of the tremendous increase in luminosity that we observe, but after several months, the surface will cool, and so the brightness will fade. During this later-stage time, the main source of the supernova's light is in fact the radioactive decay of nickel and cobalt nuclei, which were produced in the supernova event. These decaying nuclei are able to keep the supernova shining for many years.

As an observer, it would be delightful if, on a few nights a year, you could go out and just pick a supernova that you wish to look at. Life isn't like that, however. From a statistical point of view, there should be about 100 supernova a year in our galaxy, and so you would think that you would have a good chance of observing one. Think again. The most recent bright supernova, occurring in 1987, was in fact in another galaxy completely—the Large Magellanic Cloud—and the last bright one in our galaxy was several hundred years ago. So why is this? The answer is simple. As we have seen in earlier sections, our galaxy is filled with dust and gas, and it is this material that can block out the light from any supernova that may be happening. That isn't to say that we will never see a supernova—far from it; but we cannot predict with any certainty when one will occur, although there are a few stars that we should keep an eye on, Betelgeuse and Eta Carinae, to name but two.

We can, however, see the remains of a supernova, the *supernova remnant*.

Thought Question 12.1

You decide to get involved in a supernovae search initiative, but concentrate your observations on massive star supernovae in star clusters. Would you observe open clusters or globular clusters?

12.3. Supernova Remnants

The supernova remnant (usually abbreviated to *SNR*) represents the debris of the explosion, the layers of the star that have been hurled into space, and the remains of the core that will now be a *neutron star*. The visibility of the remnant actually depends on several factors, including its age, whether there is an energy source to continue making it shine, and the original type of supernova explosion.

The SNR goes through several stages that are currently a major field of research, with most basic theories proposing a three-stage process for SNR evolution:

- *First Phase*—The front of the expansion is formed from the shockwave interacting with the surrounding interstellar medium (ISM). This phase is characterized by having both a constant temperature within the SNR and constant expansion velocity of the shell. It lasts for a couple hundred years. This phase is known as the *free expansion phase*.
- *Second phase*—The SNR material slowly begins to decelerate and cool. In this phase, the main shell of the SNR is unstable, and the ejected material becomes mixed up with the gas that was just shocked by the initial shockwave. This mixing also has the effect of enhancing the magnetic field that exists inside the SNR shell. This phase lasts 10,000–20,000 years and is known as the *Sedov* or *adiabatic phase.*
- *Third phase*—This starts after the shell has cooled down to about 106 K, at which time electrons recombine with heavier atom such as oxygen, with the result that the shell radiates energy more efficiently. This cools the shell faster, making it shrink and become denser. As more of the shell cools, more atoms recombine, to create a snowball effect, and thus the SNR quickly develops a thin shell and radiates most of its energy away as optical light. After a few hundreds of thousands of years the expansion stops and the SNR begins to collapse under its own gravity. Millions of years later, the SNR will be absorbed into the interstellar medium. This phase is known as the *snow plow* or *radiative phase*.

Some images of supernova remnants are at the end of the chapter.

In addition to the processes outlined above, the remnant's velocity of expansion will decrease as it ages, usually from 10,000 km s^{-1} to maybe 200 km s^{-1}. The object will, of course, fade during this time. A few SNRs have a neutron star at their center that provides a replenishing source of energy to the far-flung material. The classic archetypal SNR that has undergone this process is the Crab Nebula, M1 in Taurus. What we are seeing is the radiation produced by electrons traveling at velocities near the speed of light as they circle around magnetic fields. This light is called *synchrotron radiation* and is the pearly, faint glow we observe. Some SNRs glow as the speeding material impacts dust grains and atoms in interstellar space, while others emit radiation as a consequence of the tremendous kinetic energies of their exploding star material.

Caldwell 34	NGC 6960	20h 45.7 m	+30° 43′	August
7 m (blue)	• 3 to 5	⊕ 70ǀ6′		

This is the western portion of the Great Cygnus Loop, which is the remnant of a supernova that occurred about 30,000 years ago. It is easy to locate because it is close to the star 52 Cygni, though the glare from this star makes it difficult to see. The nebulosity we observe is the result of the shockwave from the supernova explosion impacting on the much denser interstellar medium. The actual remains of the star have not been detected. Also known as the Veil Nebula (western section).

Caldwell 33	NGC 6992	20h 56.4 m	+31° 43′	August
7 m (blue)	• 2 to 5	⊕ 60|8′		

A spectacular object when viewed under good conditions. It is the only part of the Loop that can be seen in binoculars and has been described as looking like a fish hook. Using a telescope, it becomes apparent why the nebula has been the named the Filamentary Nebula, as lacy and delicate strands will be seen. However, there is a downside—it is notoriously difficult to find. Patience, clear skies and a good star atlas will help. A showpiece of the summer sky (when you have finally found it). Also known as the Veil Nebula (eastern section).

NGC 1909	IC 2118	05h 02m	−07° 54′	December
	• 3 to 5	⊕ 180|60		

This is a very faint patch of nebulosity, which is apparently the last of a very old supernova remnant. It resembles a long ribbon of material, which can be glimpsed with binoculars. It is glowing by reflecting the light of nearby Rigel. Very rarely mentioned in observing guides. Also known as the Witch Head Nebula.

Messier 1	NGC 1952	05h 34.5m	+22° 01′	December
8.5 m	• 1 to 5	⊕ 6|4′		

The most famous supernova remnant in the sky, it can be glimpsed in binoculars as an oval light of plain appearance. With telescopes of aperture 20 cm, it becomes a ghostly patch of gray light. In 1968, in its center was discovered the Crab Pulsar (the source of the energy responsible for the pearly glow observed), a rapidly rotating neutron star that has also been optically detected. The Crab Nebula is a type of supernova remnant called a *plerion*, which, however, is far from common among supernova remnants.

Sharpless 2–276	05h 31m	−04° 54′	December
• 6.5?	⊕ 600′		

Often mentioned in books but very rarely observed, this is a huge arcing loop of gas located to the east of the constellation Orion. It encloses both the sword and belt of Orion, and if it were a complete circle it would be about 10° in diameter. Currently believed to have originated from a supernova that occurred 2–3 million years ago. In addition, it may also have given rise to several runaway stars, that include AE Aurigae, μ (Mu) Columbae and 53 Arietis. It continues to glow due to a group of hot young stars in the Orion OB1 Association. Observationally, the eastern part of the loop is well defined, but the western part is exceedingly difficult to locate and has never, as far as we know, been seen visually, only being observed by the use of photography or using a CCD. Impossible to see it through a telescope, recent rumors have emerged that it has been glimpsed by a select few, by using either an OIII filter or an ultra-high-contrast filter. Needless to say, perfect conditions and very dark skies will greatly heighten the chances of it being seen. This may be possibly the greatest observing challenge to the naked-eye observer. Also known as Barnard's Loop.

12.4. Types of Supernovae

Before we bid a fond farewell to supernovae, and hello to the final phase of a star's life, it should be mentioned (albeit briefly) that there are two types of supernovae, or rather, there were two types!

Over the past few years, it is no exaggeration to say that there has been a plethora of discoveries concerning types of supernova. It seems that whenever a supernova is discovered, it has anomalies in its spectrum that give it a unique classification. It is as if every supernova is a new class in its own right. To describe the many disparate types is beyond the scope of this book, but there is an excellent book by David Stevenson (details in appendices) that covers most of what we know, or think we know, about supernovae.

Here is a brief list of the many differing types of supernova, but no doubt it will be added to, and expanded, over the next few years.

Type Ia
Type. Ia ("point one a")*
Type Ib
Type Ibn*
Type Ic
Type II-P
Type II-L
Type IIn*
Type IIa

As you can begin to comprehend, it is a tad complicated.[4] To try to simplify this, we will just outline the basics of supernova type, without delving too deeply into the subtle spectroscopic anomalies that have given rise to the new subclasses, although we will mention what makes each one unique.

The classification system used to distinguish the two types is a rather obtuse (for the non-professional astronomer) system based upon the simple premise of whether the supernova has emission lines of hydrogen in its spectrum. *Type I* supernovae do not have these lines, whereas *Type II* does. The supernovae we have previously discussed are Type II supernovae. This class of supernova involves the final death throes of a very evolved and massive star. These stars, as we have seen, have quite a lot of hydrogen left in their outer layers, hence the classification as a Type II.

Type I supernovae, on the other hand, do not have hydrogen emission lines and can be further divided into Types Ia, Ib and Ic, along with the recent additions Type. Ia and Type Ibn. Type Ia has absorption lines of ionized silicon, whereas Types Ib and Ic do not. An example of this type of spectra is shown in Fig. 12.3.

The difference between Ib and Ic is also a spectroscopic one, in that the former has a helium absorption line, whereas the latter does not. Type. Ia ("point one a") are about one-tenth as bright as a typical Type Ia, and Type Ibn has narrower lines of helium in its spectrum.

There is a further twist to this story. Types Ib, Ic, and II are massive stars, but Type I stars have had their outer layers stripped away either by a strong stellar wind or by the action of a nearby star (so that the progenitor supernova star is in fact part of a binary star system), and mass is transferred from one star to another. Furthermore, all three are usually found near sites of star formation. This is to be expected, as we know that massive stars have short lives, and so we do not expect them to move far from their birthplace.

Type I supernovae are a different beast altogether from Type II. They are usually (but not always) found in galaxies where star formation may be minimal or has even stopped altogether. You can now see that this implies that they originate not from the final phases of massive stars, as described above, but from some other new phenomenon. Actually, in an earlier section, we discussed stars that are the originators of Type Ia supernovae.

---

[4] Those with a * against them are recent additions to the classification system, showing subtle spectroscopic differences to the "mainstream" types. Type IIa begin with spectra similar to Type Ia, and then progress to Type II spectra. Type IIn have narrow emission lines not seen in the archetypical Type II. Types II-L have no plateau in their light curve, and Type II-P are the "typical" Type II supernovae. To further complicate matters, there are the very recent additions to the list, namely the *super-luminous supernovae*, or SLSN, along with its subdivisions of SLSN-1, SLSN-II and SLSN-R.

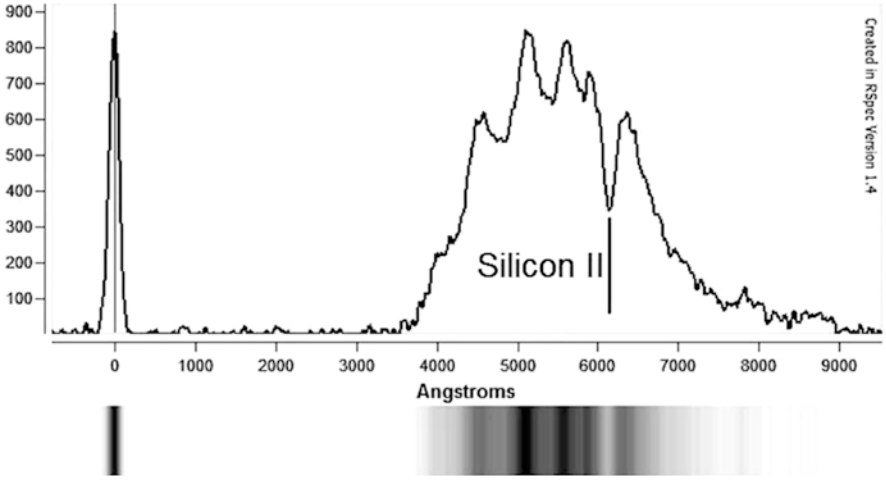


Fig. 12.3. Spectra of supernova SN2011fe. This spectrum was obtained with just a 9-in. SCT telescope and 30 × 25 s integration time. It is an excellent example of what can be obtained with amateur telescopes, a star analyzer and associated software. (Spectrum courtesy of David Strange, UK)

These are white dwarf stars that literally explode by thermonuclear reactions. Now, you may think that this is contradictory to what you read earlier, where it was said that white dwarf stars do not have any nuclear reactions occurring within them. This is absolutely correct, but in this case, the carbon-rich white dwarf star is part of a binary system where the other star is a red giant.

Recall that as a red giant star evolves, its outer layers expand, and it can overflow what is called its Roche lobe. This is the region around a star in which the gravity of the star dominates. Any matter within the Roche lobe is gravitationally bound to the star, but if the Roche lobe is filled, then matter can overcome the gravitational attraction of the star and in fact can "flow" or be transferred to a companion star.[5] This material then falls onto the white dwarf star. A consequence of this extra material is that the Chandrasekhar limit can be reached, and the increase in pressure will cause carbon burning to commence deep in the star's interior. Of course, this is accompanied by a resulting increase in temperature.

Normally, this temperature increase would mean a further increase in pressure, resulting in an expansion of the white dwarf, resulting in the star's

---

[5] The subject of binary stars, Roche lobes, and other assorted ephemera could fill a book in itself! Any interested readers will find references to books on such topics in the appendices of this book.

cooling down and the carbon burning ceasing. However, as we have seen, white dwarfs are not normal. They are made of degenerate matter, which means that the increase of temperature just results in the carbon-burning reactions proceeding at an ever-increasing rate. This is reminiscent of the helium flash process seen in low-mass stars mentioned earlier. The temperature soon gets to be so high that the electrons in the white dwarf become non-degenerate, and, to put it simply, the white dwarf blows itself to bits!

So, we can see that Type I supernovae involve nuclear energy and emit more energy in the form of electromagnetic radiation,[6] whereas Type II involve gravitational energy and emit an enormous number of neutrinos.

All of this, of course, has no bearing on the amateur astronomer who wishes to observe a supernova, as this will not help or hinder you. Nevertheless, it is important to know, in case you read in journals or on the web that the latest supernova is of Type Ib (or Ic or Ia or. Ia or Ibn or even one of the Type II's).

Thought Question 12.2

The brightest star in the night sky is Sirius, which is itself a binary system with a white dwarf companion. One could imagine that eventually a Type I supernova could occur, but can you think of a simple reason as to why this is unlikely to occur in this particular system?

12.5. Hypernovae

However, before we move onto the next section that deals with the objects left after a supernova has occurred, it would be instructive to describe an addition to the genre—the *hypernova*.

The hypernova is a type of supernova explosion with energy far in excess to that of the standard supernovae we have discussed above, and is the origin for the term "super luminous supernovae" (SLSN). In the same manner of supernovae in general, these hypernovae are believed to be the result of several differing types of stellar explosion: some well modeled and observed in recent years, some still tentatively suggested for observed hypernovae and some entirely theoretical. These are:

---

[6] As there is no core collapse in a Type I supernova, there will be no neutrinos emitted.

- *Collapsar model.* This is a type of hypernova that produces a black hole. When the core collapses in a star with a core at least around 15 times as massive as the sun, the resulting explosion is insufficient to expel the outer layers of the star, and it will collapse into a black hole without producing a visible supernova outburst. An example of a collapsar hypernova is Sn1998bw, which was associated with the gamma ray burst (GRB) 980,425. It is officially classified as a type Ic supernova due to distinctive spectral properties in the radio spectrum that is an indicator of relativistic matter.
- *Circumstellar material (CSM) model.* Nearly all of the (so far) observed hypernovae have spectra similar to either a Type Ic or Type IIn supernova, discussed previously. The Type Ic hypernovae are believed to be produced by jets from material falling back onto the black hole, whereas Type IIn hypernovae have significantly different light curves and are not associated with gamma ray bursts. Type IIn supernovae all appear embedded in a dense nebula that was probably expelled from the progenitor star itself, and this circumstellar material (hence the name) is believed to be the origin of the extra luminosity. When this material, which was expelled in an initial "normal" supernova, collides with the denser material of the nebula, close to the star, the shockwave converts kinetic energy efficiently into visible radiation, and so we observe an extremely luminous supernova of extended duration, even though the initial energy from the event was the same as that of a "normal" supernova.

There are also a few models that are only tentatively observed or entirely theoretical. These are the *pair-instability supernova*, *magnetar energy release*, and explosions resulting from binary systems, with white dwarf or neutron stars that are in unusual arrangements and/or undergoing mergers.

Using the spectra of a supernova, it is possible to measure the rate of expansion of the supernovae shell. We shall use Fig. 12.3 to determine the expansion rate. See Math Box 12.1.

Math Box 12.1 Supernova Shell Expansion Rate

Use the Doppler effect to determine the shift of the spectral line Silicon II:

$$v = \frac{\Delta\lambda}{\lambda_{rest}} \times c$$

$\Delta\lambda$ is the apparent wavelength shift.

$$\Delta\lambda = \lambda_{rest} - \lambda_{observed}.$$

v is the radial velocity.
c is the speed of light, 300,000 km s^{-1}.
Rest wavelength, λ_{rest}, of Silicon II is 6355 Å.
Observed wavelength, $\lambda_{observed}$, of Silicon II is 6150 Å.

$$v = \frac{\Delta\lambda}{\lambda_{rest}} \times c$$

$$Velocity = \frac{6150 - 6355}{6355} \times 300{,}000$$

$$Velocity = -9677.42 \text{ km s}^{-1}$$

This means the SNR is moving toward us at approximately 22,000 mph. An astounding figure.

Notice the negative sign in the answer. This indicates a blue shift in the line and means the material emitting the line is moving toward us.

(Spectra and data courtesy of David Strange.)

We now move on to the final phase of a star's life, the end result of millions, and sometimes billions, of years of stellar evolution, the end of a long journey.

Thought Question 12.3

Astronomers detect a supernova. What technique would they employ to decide whether it is a Type I or Type II supernova?

12.6. Neutron Stars and Pulsars

The endpoint of a star's life is now in sight, and although these objects are probably forever beyond the vision of amateur astronomers, it is important that we discuss them for the sake of completeness. These objects are either very small (maybe 10 km in radius) or invisible, so for all intents and purposes they are not observable with amateur equipment.[7] They represent the conclusion of a star's evolution and, until fairly recently, had never been seen, only predicted. The fascinating properties of these objects could fill a book in itself, so we shall just briefly describe those properties that are relevant to the evolutionary story.

Recall that in a Type II supernova the central 0.6 $M_\odot$ of the collapsing core has a density equal to that of the nuclei of atoms, and the neutrons become degenerate. This central core region has become a *neutron star*. In fact, after a supernova explosion has flung all the outer layers of the star into space, what remains (usually) is just this central core region. These neutron stars were actually predicted as far back as 1939 by Robert Oppenheimer and George Volkoff, who calculated the properties of a star made entirely of neutrons.

The actual structure of the star is not completely known, but there are many theoretical models that accurately describe the observations, although it might be possible to deduce some details by studying the neutron-star oscillations using a technique known as Asteroseismology.

The current view of neutron star structure goes like this.

- The surface of a neutron star is composed of ordinary nuclei crushed into a solid lattice with a sea of electrons flowing through the gaps between them. If the surface temperature exceeds 10^6 kelvins (i.e., a young pulsar), the surface will be a fluid instead of the solid phase that might exist in cooler neutron stars (temperature less than 10^6 kelvins).
- The "atmosphere" of a neutron star is theorised to be at most several micrometres thick, and its dynamics are fully under the control of the neutron star's magnetic field. Below the atmosphere one encounters a solid "crust".
- The crust is exceptionally hard and very smooth (maximum surface irregularities ~ millimetres or less), due to the extreme gravitational field.
- Continuing towards the centre, one would find nuclei with ever-increasing numbers of neutrons, kept stable by the tremendous pressures. The nuclei become increasingly small (gravity and pressure overwhelming the strong nuclear force) until the core is reached, defined as the point where mostly neutrons exist.
- The composition of the superdense matter in the core remains unknown. There are however several models describing the core.
    - superfluid neutron-degenerate matter (mostly neutrons, with some protons and electrons).

---

[7] No doubt I shall soon be corrected on this point, when an amateur images the Crab Nebula pulsar. It is only a matter of time!

- degenerate strange matter (containing strange quarks in addition to up and down quarks).
- matter containing high-energy pions and kaons in addition to neutrons.
- ultra-dense quark-degenerate matter.

Many of their properties are similar to those of white dwarf stars. For instance, an increase in mass of a neutron star will result in a decrease in radius, with a range of radii from 10 to 15 km. The mass of a neutron star can be from 1.4 to 2.6 $M_\odot$. But, of course, these figures depend on the calculation being used at the time. Nevertheless, they give a good picture of the star, see Fig. 12.4.

Two properties of neutron stars that we can describe in confidence are its rotation and magnetic field. A neutron star rotates at a very rapid rate, as many as hundreds and even thousands of times per second. It does this because of a law of physics called the *conservation of angular momentum*. Although it is a complicated law, it is easy to visualize. Just picture an ice skater spinning around; as she pulls in her arms, she spins faster. It is the same with the neutron star. The Sun rotates about once every 30 days, but if it were shrunk to the size of a neutron star, it would have a rotation rate of 1000 times per second. We also know that every star has a magnetic field but imagine compressing that field to the size of a neutron star, making it enormous. Again, using the Sun as an example, its magnetic field would increase by 10 billion times if it shrank to the size of a neutron star. The strength of the field of a typical star is about 1 Tesla, whereas in a neutron star it can be as high as 100 million Tesla.

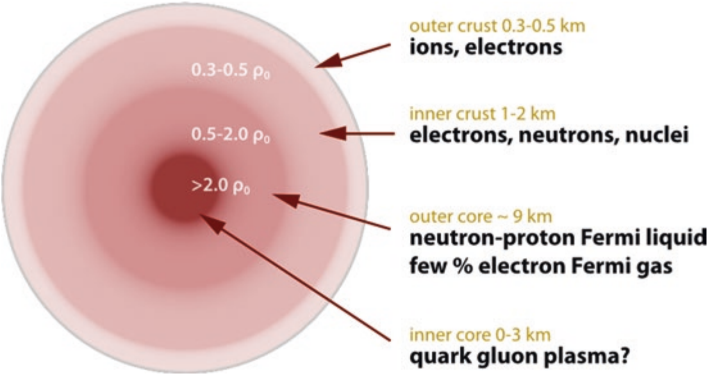


Fig. 12.4. Neutron star structure. Courtesy of Robert Schultze, Creative Commons. Densities are given in units of the nuclear matter saturation density ρ_0

Some neutron stars are believed to be part of a binary system, and material can be transferred from the companion star onto the magnetic pole regions of the neutron star with the matter traveling at perhaps nearly half the speed of light. The material literally crashes onto the star and results in “hot spots.” The temperature of these hot spots is high, in the range 10^8 K, and can result in the emission of X-rays. In fact, to casually say they emit X-rays is a bit misleading, as the amount of X-ray emission is tremendous. The total amount of X-ray luminosity can be as high as 10^{31} watts, nearly 100,000 times greater than the total amount of energy emitted by the Sun at ALL wavelengths! These *X-ray bursters* typically flare up and last from a few hours to a few days. Each burst lasts only a few seconds, and then declines in energy and brightness. This type of binary system is called an *X-ray binary*, and examples are Hercules X-1 and Centaurus X-3.

This leads rather nicely to the subject of *pulsars*. In 1967, a young graduate student in Cambridge (Jocelyn Bell) discovered a source of very evenly spaced pulses of radio emission. The period of the pulses was 1.337 s and was very constant to an accuracy of about 1 part in 10 million. The object, designated PSR 1919 + 21, was the first pulsar discovered! The problem was trying to explain what this object was. Some theories predicted a neutron star that pulsated in a manner similar to that of a Cepheid variable star, where its size actually changed. One proposal even suggested that these pulsars were in fact messages from an alien civilization. Not surprisingly, this last idea was discounted. Another model was that of a rotating white dwarf star. All of these plausible explanations were eventually discounted and the correct one accepted.

The generally accepted model of a pulsar is one in which the magnetic axis of the neutron star is tipped with respect to its rotation axis. Very energetic particles travel along the magnetic field lines and are literally beamed out from the magnetic poles. As the neutron star rotates around its rotation axis, the beamed radiation sweeps across Earth and is the pulse that we detect. In some instances, two pulses can be observed per rotation if the beams from both magnetic poles sweep past Earth.

As time passes, the period of a pulsar increases. For instance, a pulsar with a period of 1 s will slow down to a rate of 2 s in 30 million years. One observational point to make is that although we know of many pulsars, there are none that have a period of, say, 5 s or longer; this would imply that the pulse mechanism must cease after a period of time. So, neutron stars only exist as pulsars for the first tens of millions of years after the supernova explosion.

It was mentioned earlier that neutron stars are the remains from a supernova explosion, and that pulsars are rotating neutron stars. So, we would expect to find pulsars at the center of supernova remnants, or SNRs. We do, but so far there are only a handful of SNRs with associated pulsars. There are two reasons for this. To detect a pulsar, the beams have to sweep past Earth, and if they don't, we will not detect them. Second, the supernova remnant will only last a relatively short time, perhaps 100,000 years, before it merges into the interstellar medium and disappears from view. On the other hand, a pulsar can last for millions of years. So, many of the pulsars we observe now are old, with their SNRs having already been dispersed.

An example of a pulsar at the center of an SNR, and probably the most famous, is the one in the Crab Nebula designated PSR 0531–21. In fact, the energy from the pulsar is responsible for the pearly glow and appearance of the nebula and is caused by synchrotron radiation produced by high-velocity electrons spiraling around the magnetic field. An SNR that has a filled-in appearance as opposed to a shell-like appearance is termed a plerion, as mentioned earlier.

Thought Question 12.4

Are all neutron stars pulsars?

It was mentioned in the section on white dwarfs that there is a limit to a white dwarf's mass (the Chandrasekhar limit), beyond which the star cannot support the weight of the material making it up. Not surprisingly, there is also a limit to the mass a neutron star can support.[8] Current estimates put this in the range[9] of ~1.4–2.6 $M_\odot$. In some supernovae, the most massive outer layers may not have been dispersed into space during the explosion, and matter may fall back onto the already dense core. This extra material may push the neutron star core above its limit, and neutron degeneracy pressure will not be able to fend off gravity.

The core will continue to collapse catastrophically, and not even the increasing temperature and pressure can halt the inevitable result. In fact, according to Einstein's famous equation, $E = mc^2$, energy is equivalent to mass, and so the energy associated with the incredible pressure and temperatures concentrated in the now-tiny core acts like additional mass, thus hastening the collapse.

---

[8] The approximate upper limit of neutron star mass is called the Tolman–Oppenheimer–Volkoff limit.

[9] However, it is now believed there is an interval of a few tenths of a solar mass, where the masses of low-mass neutron stars and high-mass white dwarfs could overlap.

To the best of our knowledge, nothing can stop the crush of gravity. The core collapses without end, forming a *black hole*.[10] We have reached the end of a star's life.

Thought Question 12.5

An astronomer announces she has detected a binary system consisting of a main sequence star of mass 2 $M_\odot$ *and a compact object of 8* $M_\odot$ *that she believes is a neutron star. Is she correct in her assumptions.*

Problems

1. The Andromeda galaxy is moving through space and has a shift in its spectral lines. Take the reference wavelength to be 550 nm and the measured wavelength to be 549.45 nm. Determine the galaxy's velocity.
2. Is this a blue shift or red shift?
3. A galaxy has a measured Hα line at 699 nm and the rest wavelength of Hydrogen α is 656 nm. Determine its redshift.
4. Determine the velocity in question 3 as a percentage of the speed of light (Photos 12.1, 12.2, and 12.3).

---

[10] Black holes are discussed in detail in their own Chap. 14.

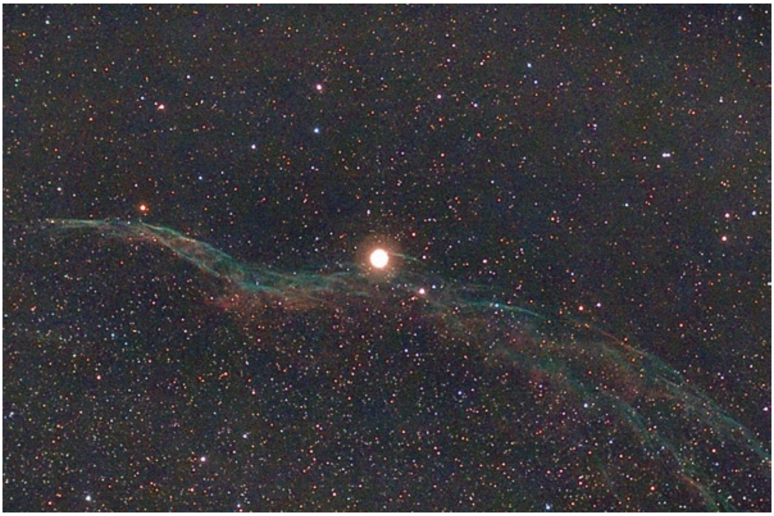


Photo 12.1. NGC 6960 Veil Nebula west

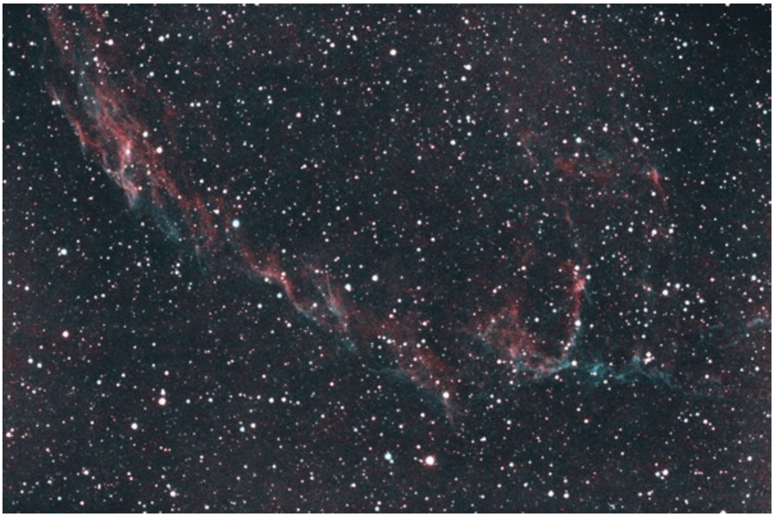


Photo 12.2. NGC 6992 Veil Nebula east

Photo 12.3. Messier 1, Crab Nebula

Chapter 13

An Interlude: Special & General Relativity

13.1. Introduction

Before we delve into the mysterious and remarkable topic of star death, it may be wise to discuss a topic which actually has a bearing on several of the following chapters, namely Special and General Relativity.

To give a detailed and full of explanation of both theories of relativity is far beyond the scope of this book. Nevertheless, I shall introduce some very easy mathematics in the section of *Special Relativity* that hopefully can be understood by all. In fact, the mathematics in special relativity is quite straightforward, but the concepts are I think difficult to explain, and understand, whereas the same cannot be said at all about *General Relativity.* The mathematics here is very advanced and to be truly understood would need a science degree, but perversely, the concepts are easier to understand. I shall endeavour to discuss both theories in a manner that will be understood by everyone, leaving out nearly all the mathematics, and giving examples of how and where these theories apply in nature, but especially in astrophysics.

We often say Relativity, as if it is just one theory, but in fact Einstein developed his theory in two parts: *The Special Theory of Relativity*, published in 1905, shows that both space and time are intertwined into what we now call *Spacetime*, but ignored the effects of gravity, whereas *The General Theory of Relativity*, published in 1915, offered an astonishing new description of gravity. Or to put it another way, the special theory is so

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_13

named because it describes the special case in which gravity is ignored, while the general theory is so named because gravity is included.

Relativity pervades all of existence, however, its effects only become apparent to us when we are either moving close to the speed of light or discussing objects that have a mass comparable to that of stars, or even greater, galaxies. So let us begin our discussion of one of the greatest achievements in science, by looking at Special Relativity.

13.2. Special Relativity

To begin our simple look at special relativity, we should first mention two ideas, or postulates, that Einstein developed.

1. First Postulate of Special Relativity
   The laws of physics are the same and can be stated in their simplest form in all inertial frames of reference.[1] This can even be simplified further, "The laws of nature are the same for everyone".

2. Second Postulate of Special Relativity
   The speed of light, c, is a constant, independent of the relative motion of the source. In a simpler form, "The speed of light is the same for everyone".

These may look like very simple ideas, but they have consequences that affect all of physics. The first postulate relates to reference frames, as all velocities are measured relative to some frame of reference such as a planet's orbit is measured relative to the star it is orbiting around. Note that the simplest frames of reference are those that are not accelerated and are not rotating. Furthermore, not only are the laws of physics the simplest in inertial frames, but they should be the same laws of physics in all inertial frames.[2]

The second postulate deals with the speed of light. But first we need to look at some history. Towards the end of the nineteenth century, the major principles of classical physics were well recognized, and the two most significant were the laws of electricity and magnetism and Newton's laws. The laws of electricity and magnetism predicted that light travels at

---

[1] An inertial frame of reference is a reference frame in which a body at rest remains at rest and a body in motion moves at a *constant speed* in *a straight line* unless acted on by an outside force.

[2] If one was to fully appreciate the mathematics that arise from this postulate, one would eventually arrive at the famous equation $E = mc^2$.

$c = 3 \times 10^8$ ms^{-1} but they do not specify the frame of reference in which light has this speed. Newton's laws on the other hand, describes velocities which add up. Here's an example; if you are on train moving exactly at 100 kms^{-1}, in a straight line, and you throw a cricket ball down the train at 25 km^{-1}, to an observer standing outside the train, the ball would be moving at 100 + 25 = 125 km^{-1}, but to you, the ball is only moving at 25 km^{-1}.

Now consider the same scenario but this time you are shining a laser beam down the train, which, incidentally, is moving at the speed of light. For you, the light is, as to be expected, is moving at c, whereas for the stationary observer, the laser beam would appear to be moving at $c + c = 2c$, clearly something is wrong. Thus, Newton's addition of velocities is not acceptable for light, and Einstein established that this is correct; any object that possesses mass cannot travel at a speed greater than c.[3] The conclusion is simple—light in a vacuum *always* travel at speed c relative to any observer.[4]

13.3. What Is Relative?

When Einstein was developing his theories, he realised that it would be nigh on impossible to conduct experiments to test his ideas, so he used "*Gedankenexperiment*" or "thought experiments". It is just a way to think through the consequences of simple ideas. Here's an example of a thought experiment that will illustrate why makes the special theory of relativity so astonishing.

Imagine you walk to work using the same path every day, and you have measured the distance very accurately, over several months, to be 3 km. Your partner also takes this path to his work, but instead he cycles the path, and has determined once again, that the distance is 3 km. You both agree that distance travelled either by walking or cycling is the same. You also agree that it may take you 60 min to walk that distance, as you both set off at the same time, and you text your partner when you arrive at work. So, this shows that in ordinary life, the distances, and times you measure appear absolute and distinct. You both agree on the distance between any two points, between say, home and work, and the time between two events, such as when you leave home and arrive at work.

---

[3] We mention mass here, otherwise the nature of photons would become a problem; therefore, they have a zero mass. What a photon actually is, is another problem, as recent research suggests it is a point source. Think about that if you can.

[4] The speed of light is a constant $c = 3 \times 10^8$ ms^{-1} *in a vacuum*.

Einstein however, revealed that these viewpoints are not precisely correct because if one were to make very, very accurate measurements of both the distance and time, the measurements you obtain would *differ* from those obtained by your partner on his bicycle. To be completely accurate, the differences will be so small as to be unnoticeable, but if your partner was cycling close to the speed of light, these differences of distance and time, would be considerable.

Many people, mistakenly, believe that what Einstein's theory tell us is "everything is relative." However, what it really means arises from the idea that motion is always relative. Here is another thought experiment that is often quoted, which illustrates just this point, and goes like this.

You are on a new aeroplane that can fly at a speed of 1670 kmh^{-1} from Nairobi, Kenya, that is located on the Earth's equator. Someone asks how fast is the plane moving? At first glance this appears to be a very trivial question as we just said we have just stated it is travelling at 1670 kmh^{-1} However, consider alien spectators looking at the solar system as a whole. They would see the plane travelling at a speed of more than 100,000 kmh^{-1}, because that is Earth's orbital speed. Furthermore, onlookers living in another galaxy see the aeroplane moving at about 800,000 kmh^{-1} due to the Milky Way rotation. There is only one thing all observers, whether alien or not, can agree on is that the aeroplane is traveling at 1670 kmh^{-1} relative to the surface of the Earth.

Now imagine that the aeroplane is flying from Nairobi, Kenya to Quito, Ecuador, both these cities being approximately located on the equator. Furthermore, the aeroplane is flying east to west, but the Earth rotates from west to east, at the same speed as the plane—1670 kmh^{-1}. Your partner, now on the Moon, and looking at the Earth would see the plane appear to stay put while Earth rotated beneath it. Thus, we have two alternative viewpoints about the aeroplane's flight. You would say that the aeroplane is traveling westward across the surface of the Earth, whilst your partner on the Moon says the aeroplane is stationary while Earth rotates eastward beneath it. The point being made here is that both viewpoints are equally valid.

Our final thought experiment is another that is often used to discuss relative motion. Imagine the situation where you find yourself floating freely in a spaceship, and because you feel no outward awareness of motion, you perceive yourself to be at rest. Now, looking out of a porthole you see your partner in his spaceship, moving away at a *constant*[5] speed of 100 kmh^{-1}. This seems all well and correct, but let's look at it from the viewpoint of

---

[5] It is vitally important here that the word constant is stressed. If any acceleration is involved, then special relativity cannot be used to describe the situation.

your partner. He thinks he is stationary as he believes, just like you, that he is floating freely in a spaceship with no outward awareness of motion. Therefore, he believes *he* is at rest and that *you* are moving away from him at 100 km/h. Both points of view are equally valid, and in fact you could both argue infinitely about who is moving and who is stationary, but it's pointless because all motion is relative.

Any measurement of motion, as well as any measurement of both time and space that we make, only makes sense when described against whom or what they are being measured relative to. Thus, when one asks questions like "Who's really moving?" and "How fast is one going?" can have no absolute answers.

We shall limit ourselves to the above discussions, and not pursue any further scenarios when objects are moving at speeds that approach the speed of light. The reason for this is that the chapter would approach the length of the rest of the book. Rather I will mention the results of such an approach.

We have discussed that information cannot travel faster than the speed of light (in a vacuum), and no object can even attain the speed of light. But if we were to go pursue this line of thought, we would see how special relativity would lead us to the following conclusions:

(i) Observe anything moving by you close to the speed of light, and you will see that time runs more slowly for the moving object. This leads you to the conclusion that a moving human being ages slower than you, or a clock tick slower than your clock. Note that the objects, whether it be a human or just about anything else (computer, tree, cat, etc., etc..) has to be moving close to the speed of light for you to notice any appreciable effect.

(ii) Observing two events to occur simultaneously, for example a flash of light that is in two different places at the same time, can result in someone that is moving by you close to the speed of light, not agreeing with your view that the two events were simultaneous.

(iii) If you were to measure the length of, say, a ruler,[6] moving by you close to the speed of light, you would find that its length (*in the direction of its motion*) will be shorter than if the ruler were not moving.

(iv) Imagine you are able to measure the mass of something (kitten, smartphone, spaceship, etc., etc.) that is moving by you close to the speed of light, you would discover that its mass will be greater than the mass it would have, if it were stationary.[7]

---

[6] It doesn't necessarily have to be a ruler, as the length of anything will appear shorter. It just makes life easier to imagine a ruler.

[7] In fact, it is this line of reasoning that eventually leads to possibly the most famous equation ever; $E = mc^2$.

13.4. Time Dilation, Length Contraction and Relativistic Mass

13.4.1 Time Dilation

As mentioned above, if you are stationary, and another person is going past you close to the speed of light, and you were to compare your watch with theirs, you would see that their clock is ticking slower than yours. This is Time Dilation, and can be stated thus; *Time dilation is the phenomenon of time passing slower for an observer who is moving relative to another stationary observer.*

We could follow this statement with some subtle, but easily followed derivation of formulae that can be used to illustrate this phenomenon. However, as astronomers, we are more inclined to see the end result, so I will not go down this path, but instead present the end result.

First, we are going to talk about the Lorentz Factor, γ, which is given like this

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

where

v is the velocity of a moving object, say, a spaceship.

c is the speed of light.

As you see this isn't a very difficult equation, and in fact is derived from simple geometry. Some further analysis shows that if you consider the interval of time measured in a moving object, and compare the time in a stationary object we get the following

$$\Delta t_{so} = \gamma \Delta t$$

where

Δt is the time measured by an observer in, say, a spaceship.

Δt_{so} the time measured by an observer on the Earth, i.e., you.

This can be further simplified

$$\Delta t_{so} = \gamma \Delta t$$

$$\Delta t_{so=}\frac{\Delta t}{\sqrt{1-\frac{v^2}{c^2}}}$$

And we can reorder this to give

Maths Box 13.1 Time Dilation

Question. Consider a spaceship moving at 90% the speed of light, 0.9c, or 270,000 km^{-1}. Calculate the time dilation observed by a stationary observer.

Note that v^2/c^2 is a ratio so we can just use the percentages of each value, although you could use the actual velocities. I find it is easier to do the former. We are only interested in the dilated time factor and so can set the stationary time, Δtso, to be 1. Also, I drop the percentage symbols, and c is equal to 100% of the speed of light.

Using

$$\Delta t_{so}\sqrt{1-\frac{v^2}{c^2}}=\Delta t$$

We get

$$1\,\mathrm{x}\sqrt{1-\frac{90^2}{100^2}}=\Delta t$$

This gives us

$$\sqrt{1-\frac{8100}{10{,}000}}=\Delta t$$

$$\sqrt{1-0.81}=\Delta t$$

$$\sqrt{0.19}=\Delta t$$

$$0.436=\Delta t$$

Thus, at 90% of the speed of light, the time aboard the space craft, or local time as it is referred to, has slowed down to 43.6% of that relative to a stationary observer. Or to put it another way, if we launch a rocket into space on a 50-year mission at a velocity of 90% of the speed of light, the rocket, and any astronauts on board will have only aged by

Answer $= 0.436\,\mathrm{x}\,50 = 21.8\,\mathit{years}$

Note that everyone on Earth will have aged 50 years.

$$\Delta t_{so}\sqrt{1-\frac{v^2}{c^2}} = \Delta t$$

These are in fact, the formulae for time dilation. An example would be helpful here, See Maths Box 13.1.

13.4.2 Length Contraction

The analysis for length contraction is more or less the same as for time dilation, and we end up with similar equations. To be precise, *Length Contraction* is when a moving object's length is measured to be shorter than its proper length, which is the length as measured in the object's own rest frame, as if it were not moving. Also known as the *Lorentz* contraction or *Lorentz–FitzGerald* contraction, after Hendrik Lorentz and George Francis FitzGerald, it is usually only noticeable when something is moving at a substantial fraction of the speed of light. An important aspect to note is that the phenomena of Length contraction only occurs in the direction in which the body is travelling. Now, for most objects in life, the effect is negligible when moving at the speeds we encounter every day, and thus can be ignored. It only becomes meaningful if an object approaches the speed of light relative to the observer. Here are the formulae.

$$\Delta l_{so} = \gamma \Delta l$$

where
Δl is the length measured by an observer in, say, a spaceship.
Δl_{so} the length measured by an observer on the Earth, i.e., you.
This can be further simplified

$$\Delta l_{so} = \gamma \Delta l$$

Maths Box 13.2 Length Contraction

Question. Consider a spaceship that is moving with a velocity 99.94% the speed of light, or 0.9994c, and is travelling to α Centauri (Alpha Centauri), a distance of 4.300 light years, as measured from the Earth. Calculate the distance from Earth to Alpha Centauri as measured by the astronaut.

Using the same approach as in the example shown in Maths Box 13.1.

Using

$$\Delta l_{so}\sqrt{1-\frac{v^2}{c^2}} = \Delta l$$

Calculating Δl

$$\Delta l_{so}\sqrt{1-\frac{99.94^2}{100^2}}$$

This gives us

$$\Delta l_{so}\sqrt{1-\frac{9988.00360}{10{,}000}}$$

$$\Delta l_{so}\sqrt{1-0.99880036}$$

$$\Delta l_{so}\sqrt{0.00119964}$$

$$\Delta l_{so}0.0346358$$

$$\Delta l = \Delta l_{so}0.034635$$

$0.0346358 \times 4.300 = 0.1490$ light years.

Thus, at 99.94% of the speed of light, the distance to α Centauri as measured aboard the space craft, has shortened to 3.46358% of that relative to a stationary observer. Or to put it another way, if we launch a rocket into space travelling to α Centauri at 99.94% of the speed of light, the distance as observed by an astronaut will be 0.149 light years.

$$\Delta l_{so=}\frac{\Delta l}{\sqrt{1-\frac{v^2}{c^2}}}$$

And we can reorder this to give

$$\Delta l_{so}\sqrt{1-\frac{v^2}{c^2}} = \Delta l$$

Let's look at an example, See Maths Box 13.2.

Note that in Maths Box 13.2 I have used numbers to several significant figures. This is important as rounding off a number to, say, 2 decimal points can have a significant effect on your final answer.

In addition, when working these problems be sure to think beforehand as to what an answer should be like. Furthermore, in the previous two examples, both the time dilation and length contraction should result in the times and lengths under consideration to be smaller than the originals. So, if you get your Δl and Δl_{so} and Δt and Δt_{so} mixed up, just think about the ideas of special relativity.

13.4.3 Relativistic Mass

Our final discussion in this section deals with a topic that often leaves readers scratching their heads and saying something along the lines of "does this *really* happen?". Well, yes, it does, but in a way different to what you expect. The topic is of course Relativistic *Mass Increase*. In a nutshell, as one moves with a speed approaching the velocity of light, one's mass increases. That's it really, and it has to do with Einstein's famous equation, $E = mc^2$. Without delving too deeply into the theory (and mathematics), energy is equivalent to mass, or, scientifically speaking, mass–energy equivalence states that all objects having mass, have a corresponding intrinsic energy, even when they are stationary. This is the rest mass. However, when one is moving at speeds close to the speed of light, the mass, as observed by a stationary observer, would seem to increase. In fact, as the mass approaches the speed of light, it would increase a lot, until, at the speed of light, its mass would be infinite.

What people tend to ask is, does the mass actually increase? Relativists get around this by using the mass-energy equivalence we mentioned in the previous paragraph, as well as the concept of momentum. They tend to say—it's the energy and momentum that increases, and leave it that, which really is quite an unsatisfactory answer to most people.[8] However, to fully explain the concepts would take several of the book's chapters, so we will leave the explanation as it is, and scratch our heads.

The equations are very similar to the previous discussion of time dilation and length contraction. They are

$$\Delta m_{so} = \frac{\Delta m}{\gamma}$$

---

[8] It is probably a better idea to not use relativistic mass as a concept at all, and just state that any object will become harder to accelerate as it gains speed because of its increasing momentum, not its mass. However, to give a description of momentum in special relativity would be more than one is willing to tackle at this level, so let's leave it at that.

Maths Box 13.3 Relativistic Mass Increase

Question. Consider a spaceship that is moving with a velocity 99.94% the speed of light, or 0.9994c. On board is an astronaut with a mass of 50 kg.[9] Calculate the mass of the astronaut as measured by the observer.

Using

$$\Delta m_{so} = \frac{\Delta m}{\gamma}$$

$$\Delta m_{so=} \Delta m \sqrt{1 - \frac{v^2}{c^2}}$$

We solve for Δmso

$$\Delta m \sqrt{1 - \frac{99.94^2}{100^2}}$$

This gives us

$$\Delta m \sqrt{1 - \frac{9988.00360}{10,000}}$$

$$\Delta m \sqrt{1 - 0.99880036}$$

$$\Delta m \sqrt{0.00119964}$$

$$\Delta m_{so=} \Delta m\, 0.0346358$$

$$\frac{\Delta m_{so}}{0.0346358} = \Delta m$$

50/0.0346358 = 1443.6 kg.

Thus, at 99.94% of the speed of light, the mass of the astronaut has increased by an extraordinary amount relative to a stationary observer.

where
Δm is the mass measured by an observer in, say, a spaceship.
Δm_{so} the mass measured by an observer on the Earth, i.e., you.
This can be further simplified

---

[9] Recall that weight and mass are not the same, your mass remains (more-or-less) constant, however, your weight is the effect of the force of gravity acting on your mass.

Table 13.1 Relativistic effects

Velocity (v) as a fraction of the speed of light (c) $\frac{v}{c}$	Length (l) of a moving object as a fraction of its rest length (l_o) (length contraction)	Mass (m) of a moving object compared to its rest mass (m_o) (mass increase)	Length of a time interval measured on the moving clock compared to the time interval measured on Earth (time dilation)
0	1.000	1.000	1.000
0.1	0.995	1.005	0.995
0.3	0.954	1.048	0.954
0.5	0.867	1.155	0.867
0.6	0.800	1.200	0.800
0.7	0.714	1.400	0.714
0.8	0.600	1.667	0.600
0.9	0.436	2.294	0.436
0.95	0.312	3.203	0.312
0.99	0.141	7.089	0.141
0.999	0.045	22.366	0.045

The terms rest-length (l_o), and rest-mass (m_o) refer to the length and mass which the moving object would have if it were stationary relative to the Earth. Table courtesy of Iain Nicolson

$$\Delta m_{so} = \frac{\Delta m}{\gamma}$$

$$\Delta m_{so=} \frac{\Delta m}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Let’s look at our final example for Special Relativity, See Maths Box 13.3.

To see how the length, mass, and time changes as one approaches the speed of light, refer to Table 13.1.

As I mentioned earlier, it would appear that the effects of special relativity only occur at if you’re travelling close to the speed of light. However,

---

[10] There are in fact three sources of redshift or blueshift: relativistic Doppler shifts; gravitational redshifts, due to light exiting a very strong gravitational field; and the cosmological expansion (where space itself stretches). This article concerns itself only with Doppler shifts.

[11] The relativistic Doppler effect is different from the non-relativistic Doppler effect, such as we hear as a police siren changes pitch approaching, and then receding from you. The difference here is that the equations include the time dilation effect of special relativity and do not involve the medium of propagation as a reference point, such as the atmosphere.

there are a few examples that we can discus, and mention, that may surprise you. Let's begin where special relativity occurs in astronomy.

A topic we have already discussed is *Redshift* or *Blueshift,*[10] where one measures the shift in a spectral line due to the motion of, say, a star. This phenomenon is due to the Doppler effect. To be accurate, it is more properly known as the *Relativistic Doppler Effect,*[11] caused by the change in frequency and thus wavelength, of light, due to the relative motion of the source, which could be a star or galaxy, and the observer, when we take into account the effects we have just described due to the special theory of relativity.

Electromagnetism is another example of special relativity. In fact, Einstein mentions this in the first paragraphs of his famous paper—On the Electrodynamics of Moving Bodies—where he states "the electromagnetic output produced by the relative motion of a magnet and an induction coil is the same whether the magnet is moved or whether the coil is moved. This is a finding of James Clerk Maxwell's electromagnetic theory, and of course Maxwell's theories are concerned with light. Therefore, one could say that without special relativity, light as we know it would not exist. A professor of physics at Pomona College in Claremont, California, Thomas Moore, used the principle of relativity to demonstrate why Faraday's Law, a famous law that most physics students will be familiar with, to show whether or not a changing magnetic field creates an electric current, is true. He says, "Not only would magnetism not exist but light would also not exist, because relativity requires that changes in an electromagnetic field move at a finite speed instead of instantaneously," and furthermore, "If relativity did not enforce this requirement … changes in electric fields would be communicated instantaneously … instead of through electromagnetic waves, and both magnetism and light would be unnecessary."

A couple of examples that have a more earthbound application will be discussed now, and one of these will really surprise you I think,

One thing that is in use every day that is affected by relativity is the GPS system. The level of precision from GPS satellites must be known to an accuracy of 20–30 nanoseconds but remember that the satellites are constantly moving relative to observers on the Earth, thus effects predicted by both the Special and General theories of Relativity[12] must be taken into account to achieve the desired 20–30 nanosecond accuracy.

An observer on the ground sees satellites in motion relative to them, and special relativity predicts that one should see their clocks ticking slower, by about 7 microseconds per day due to the slower time rate caused by time dilation. In addition, the satellites are in orbits high above the Earth, where

---

[12] See next section on General Relativity.

the curvature of spacetime due to the Earth's mass is less than it is at the Earth's surface. Taking into account one of the predictions of General Relativity is that clocks closer to a massive object will seem to tick more slowly than those located further away, as we will see later in the section on Black Holes. Thus, when observed from the surface of the Earth, the clocks on the satellites appear to be ticking faster than identical clocks on the ground, by 45 microseconds per day.

Combining these two relativistic effects means that the clocks on-board each satellite should tick faster than identical clocks on the ground by about 38 microseconds per day. The high precision required of the GPS system requires nanosecond accuracy, and 38 microseconds is 38,000 nanoseconds, so if the effects are not accounted for, any navigational fix based on the GPS system, would be inaccurate after only 2 min, and errors in global positions would continue to accrue at a rate of about 10 kilometres each day.

Our final example of the application of special relativity, and indeed the final part of our discussion on Special Relativity is a discussion about Gold!

Gold has an individual, soft, yellow colour, and its well-deserved stunning lustre will now seem even more exotic to you when you find out that it's actually due to relativistic effects. Imagine calculating the frequency of light that gold emits but not taking special relativity into consideration, your calculation would show that gold should have a silver shine. Yet, the colour of gold in reality leans towards the red end of the spectrum. How is this so?

Well, this incongruity can be explained if we look at the electrons in the atoms of gold orbiting in their shells.[13] Consider the orbital closest to the nucleus where the electrons have to move at the shockingly fast speed. They need to do this to avoid being drawn into the nucleus by the formidable positive charge from all the protons in the nucleus, and this can cause a lot of relativistic effects, because the electrons are moving so fast, roughly half the speed of light. Due to this high speed, the separate electron shells appear to be closer than they actually are, and for any electron to jump to a higher energy level it would need to absorb a very specific wavelength of light. In the case of gold, the wavelengths that absorbed are normally in the ultraviolet range—way beyond our normal range of vision. Nonetheless, if we take into account the relativistic effects that appear to squeeze the shells nearer to each other, we see that gold actually absorbs light with a smaller frequency, and that of course can be blue light. Thus, the blue light is absorbed, with only the reddish colours, red, yellow, and orange, reflected into our eyes. Consequently, gold has a glamorous, yellowy lustre.

---

[13] In an atom of gold, there are a total of 79 electrons orbiting a nucleus that contains 79 protons.

That is now the end of our discussion on the first section of Einstein's Relativity. We will now talk about General Relativity.

13.5. General Relativity

The second aspect of Einstein's theories of relativity is *General Relativity*. It describes gravity in a completely new way, and is, in my opinion, one of the greatest intellectual achievements of all time.

Prior to this, gravity was believed to be a property of mass, in that using Isaac Newton's laws of Gravitation, one could predict the orbits of planets, or the path of a cricket ball if one new the mass of, say a planet, the sun or a cricket ball, and the distance between any two objects under consideration. This was discussed in Chap. 2, where we introduced Newton's Universal Law of Gravitation and here it is again.

$$F_1 = F_2 = G\frac{\left(M_1 M_2\right)}{d^2}$$

where

M_1 – Mass of first object (kg)
M_2 – Mass of second object (kg)
d – distance apart (m)
G – Gravitational constant = 6.67×10^{-11} N·(m/kg)^2

As you can see it is a very simple equation and describes adequately the gravitational attraction of any two bodies. It actually works quite well and gives very good answers, but there is a slight problem. Let us now have another thought experiment. Imagine the Earth in orbit around the Sun, and everything is well. Suddenly, the Sun disappears, and at the exact same instant the Earth flies of on a tangent into space. Well, this is what would happen if Newton's ideas held sway. But the key point here is the use of the words "same exact instant". Using Newtons ideas, the gravitational attraction, or rather lack of it, if the Earth were to disappear, would be felt

---

[14] Note that the precession of planetary orbits is not peculiar to Mercury, all the planetary orbits precess to some degree or other. The effect getting smaller the further away from the Sun.

[15] Even when taking into account the effects from the other planets, as well as a very slight deformation of the sun due to its rotation, as well as the fact that the Earth is not an inertial frame of reference,

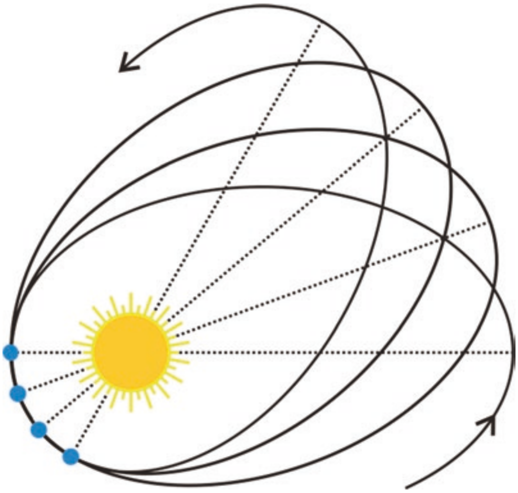


Fig. 13.1. The precession of the perihelion of mercury

instantaneously by the Earth. This means that the effect of gravity propagates faster than the speed of light, and we all know that this cannot be true.

A more tangible situation that could not be explained by Newton's Laws is something called the *Precession of Perihelion of Mercury*. As Mercury orbits the Sun, it travels on an elliptical orbit as we discussed in Chap. 2. However, it had been known, even during Newton's lifetime, that the point of closest approach of Mercury to the Sun, or Perihelion, does not always occur at the same place but slowly moves around the sun, and this is called precession, see Fig. 13.1.

The precession of the orbit is not peculiar to Mercury, all the planetary orbits precess.[14]

The observed precession of Mercury's orbit is 5601 s of arc per century, but when Newton's equations are applied,[15] they predict a precession of

Thought Question 13.1

Why do we keep using Newton's Equations even though we now know that Einstein's ideas are the correct description of gravity (as far as we can tell).

---

[16] He is best known for predicting the existence and position of Neptune using only mathematics.

5557 s of arc per century. There is a discrepancy of 43 s of arc per century, or about 7% You may think that this isn't too much, but it is a significant amount. Over the next 200 years, various explanations were put forward, such as a large but unobserved amount of dust that lies between the Sun and Mercury, and this extra gravity would cause the observed discrepancy.

Then in 1859, the French astronomer and mathematician, Urbain Le Verrier[16] showed that Mercury's orbit could never be explained using Newton's equations. He did, however, suggest that another planet might exist in an orbit even closer to the Sun than that of Mercury, and this would account for observed perturbation. Due to his success of his work on predicting the position of Neptune led some astronomers to agree with him, and the hypothetical planet was named Vulcan. However, as we know, no such planet was ever found.

The correct explanation was announced by Einstein in November 1915, when he showed that General Relativity could account for the observed precession. This was the first validation of his Theory.

Thought Question 13.2

If the Sun were to collapse and form a black hole, (i) what would happen to the Earth's orbit? and (ii) as an observer, how would this affect your planetary observations?

Before we move onto a new section it is worth mentioning just how important General Relativity is, and what it predicted. For instance,

- Gravity arises from distortions of spacetime. It is not an enigmatic, shadowy force that acts at a distance, as Newton stated. The presence of any

---

[17] This enigmatic statement will be discussed in Chap. 13.

[18] It would be remiss of me if I failed to mention the detection of gravitational waves. The first indirect existence of gravitational waves was the observed orbital decay of a binary pulsar, that matched the decay as predicted by general relativity as energy is lost to gravitational radiation. For this discovery, Russell A. Hulse and Joseph Hooton Taylor Jr. received the 1993 Nobel Prize in Physics. Direct observations of gravitational waves came about in 2015, when a signal generated by the merger of two black holes was received by the LIGO gravitational wave detectors in Livingston and in Hanford. For this direct detection of gravitational waves, Rainer Weiss, Kip Thorne and Barry Barish were awarded the 2017 Nobel Prize in Physics.

[19] E. F. Taylor and J. A. Wheeler, Spacetime Physics, second ed. (Freeman, 1992).

mass will cause the distortion, that determine how other objects move through spacetime.
- Time passes slower in gravitational fields. The stronger the gravity, the more slowly time passes.
- Black holes will exist in spacetime and falling into a black hole will result in you leaving the observable universe.
- The universe has no boundaries and no centre, but it might still have a finite volume, or it may not. It could be infinite in both time *and* space.[17]
- A sizeable mass that undergoes a rapid change in motion or structure will emit gravitational waves that travel at the speed of light.[18]

13.6. Spacetime

> *Space is different for different observers.*
> *Time is different for different observers.*
> *Spacetime is the same for everyone.*[19]

We live in a 3-dimensional universe that we are all familiar with, and that everything can be explained in terms of length, width, and depth, but if I were to say to you - try to imagine a 4-dimensional space, it would be impossible to do, and so you could not visualize any direction, or dimension, that is different from length, width, and depth. Our brains are not capable of this, but this doesn't mean they do not exist. Although we have no hope of picturing a 4-dimensional space, it can however be described mathematically. In fact, we could describe 5-dimensional space, or 6, or 7 and so on. Any space that has more than three dimensions is called a *hyperspace,* which means "beyond space."

From his work on Special Relativity, Einstein realized that space and time are intimately linked. Specifically, special relativity says that the 3 dimensions of space and the one dimension of time together form an intimate, 4-dimensional amalgam that we call spacetime. Thus, *Spacetime* is a 4-dimensionl space that is described by length, width, depth, and *time.*[20] Einstein realized that when he expanded his theory of relativity so that it included gravity, he discovered that matter shapes the structure or fabric of spacetime. A common analogy often used is that of a heavy weight placed upon a 2-dimensional rubber sheet, with the sheet being distorted, or the Earth in space, see Fig. 13.2. The larger the mass, the greater the distortion.

It has to be said that this is a challenging concept, with most people finding it difficult to comprehend, as we are unable to perceive the underlying

---

[20] Do not make the common error of thinking that time is the fourth dimension. It isn't, it is simply one of the four. You could say spacetime is length, width, time, and depth.

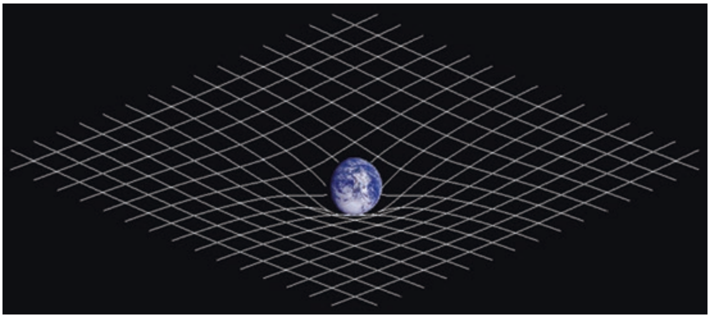


Fig. 13.2. The space-time distortion made by Earth. (Image Courtesy of Creative Commons)

curvature of the space we live in. Even though one can discuss spacetime as being analogous to a sheet of rubber, the analogy breaks down because a rubber sheet is 2-dimensional, whilst spacetime is four dimensional. Remember, it isn't just a warping in space that the analogy represents, but also a warping in time. In fact, the complex equations used in general relativity are so complex that if someone finds a solution to any of the equations, the solution is often named after them, and they become world famous.

It may be wise to note that the rubber sheet analogy has several limitations. For instance, the rubber sheet represents the universe, but it is incorrect to think of placing a mass like the earth *on* the universe. Think instead think of the mass, or earth as being *within* the rubber sheet. Furthermore, the analogy does not show the time part of spacetime, because after each orbit of the Sun, the Earth returns to the same place in space, relative to the Sun, but also to a time that is 1 year later.

Actually, the deduction that spacetime is a single framework wasn't one that Einstein reached by himself, but originated with the German mathematician Hermann Minkowski, who in a 1908 colloquium, said the famous words "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

Before we move onto the topic of curved spacetime, we can make an attempt to picture spacetime ourselves. It is impossible to picture in our minds the 4-dimensions of spacetime but imagine if we could. The 3 spatial dimensions of spacetime that we are familiar with would be visible, but

---

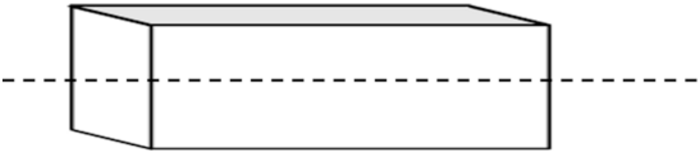
[21] I apologise for the poor quality of the two drawings but trying to find a curved rectangular block in 3 dimensions was nigh on impossible.

every single object would be stretched out through the dimension of time. Basically, those objects that we see as 3-dimensional in our ordinary lives today, would appear as 4-dimensional objects in spacetime. If it were possible to "see" in 4 dimensions, it would be possible to "see" every event in a person's life.

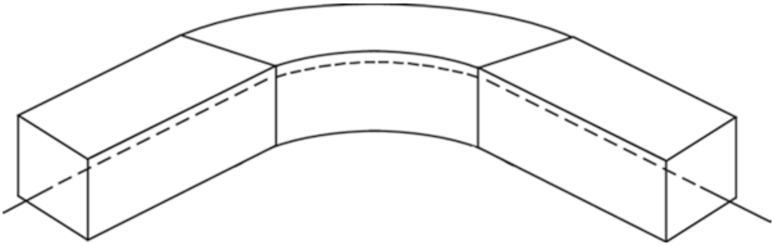
13.7. Curved Spacetime

We are now going to discuss the concept of Curved Spacetime. This is one of those ideas that, like Spacetime itself, is just as difficult to picture in our minds, but there is a simple analogy that may help. Before that, however, lets recap spacetime. As Einstein stated, spacetime is a 4-dimensionl space that is described by length, width, depth, and time. Now picture a volume of space with the proviso that there is nothing in that volume of space. Think of it as a place between, say, galaxy clusters, or, as it called, a *Void*. In this volume of space, there are no objects that have a mass, nothing at all, and so we can call this volume of spacetime as being flat. If a laser beam were to pass through this volume of space, it would travel in a straight line.

This is the point where the aforementioned analogy comes into play.[21] Now imagine instead of a volume of space, we have an enormous rectangular block of jelly, or jello. This we can call our volume of spacetime, and running though the centre of this rectangular block, from one end to another, is a length of string, in an utterly straight line, see drawing below.



Now, imagine that we bend or warp the block in an arc, see drawing below.



The piece of string that was in a straight line, is now following the curve of the block. Furthermore, the curvature is in all spatial dimensions, i.e., the x, y, and z axis, although in the drawing it is only curved in the x-axis.

Take this analogy further, and instead of jelly (jello) we have a volume of space. Spacetime would be "flat" in the first drawing, but if we were to introduce an object with mass, the volume of space would be curved about the object, as in the second drawing. Furthermore, if we now have instead of a length of string, we have a laser beam, in flat space it travels in a straight line, however, in the second drawing *it is still travelling in a straight line, but in curved space!* The laser is following the straightest possible path through curved spacetime. This path is called a *Geodesic* and is a line of shortest spatial distance between two points in space.

A good astronomical example of geodesics are the orbits of planets, comets, meteors etc., that could have circular, elliptical, or unbound parabolic or hyperbolic orbits. In fact, they have the same orbits as predicted by Newton's Universal Law of Gravitation. But the correct explanation for the orbits is now very different from Newton's view of gravity. They are not orbiting due to a shadowy force pulling on them by the distant Sun, instead, they orbit because they are following the straightest possible paths, or geodesics, permissible by the shape of the spacetime around them. The (I think) startling conclusion is that the *mass* of the Sun is not grasping them but is simply dictating the shape of spacetime surrounding it.[22]

It is often famously stated like this.

Matter tells space how to curve, and curved space tells matter how to move.

A very important point to make here is that this is just a very simplistic description of what curved spacetime is about. The real shape of curved spacetime is very difficult to picture, and in reality, is not actually as I describe, but it helps to give an idea of what's going on. I hope.

So now we have the intriguing conclusion that what we observe as gravity actually results from the curvature of spacetime.

Finally, let us close this chapter with a short discussion on what is called the equivalence principle.

---

[22] Isaac Newton himself was troubled by the idea that the Sun could somehow dictate how the Earth moved through space. In fact, sometime in 1692/93 he wrote in a letter "*That one body may act upon another at a distance through a vacuum, ... and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man, who has ... a competent faculty in thinking, can ever fall into it*".

[23] Picture yourself in an airplane taking off. You get "pushed" into your chair, even though there is no force pushing you, it is the place accelerating.

13.8. Gravity and the Equivalence Principle

There is a story that Einstein tried to imagine what one would experience if one fell of a ladder, ignoring the end result of hitting the ground. A modern interpretation prefers to put a person in a stationary elevator. One would naturally have one's feet firmly on the floor, but it is more complicated than this. For instance, if you were in a cabin of a spaceship with no windows and was accelerating at exactly 9.81 m per second squared, your feet would, once again, be firmly placed on the floor.[23] However, there is no gravitational force "pulling" you down, as there would be in an elevator. So, to all intents and purposes, you couldn't tell if you were standing in an elevator on the Earth, or in an accelerating spaceship.

If you were to drop your phone, in the elevator, it would fall to the ground. Similarly, if you were to let go of it in the spaceship cabin, it too would "fall" to the ground. From your vantage point standing on the cabin floor, you cannot tell the difference between the two situations: are the phones falling towards a body with considerable mass, like the Earth, or is the cabin floor accelerating towards you?

Thought Question 13.3

You are standing on a weighing scale in an elevator and it reads 100 kg, and the elevator suddenly plummets down. What does the weighing scale now read as you fall to your doom?

This leads us onto the weightlessness. Now, imagine that you are in the elevator, and the cable breaks, and plummets down. You and everything else in the elevator, would be accelerated at exactly the same rate so that no influence of gravity could be detected, and relative to the elevator, you and

---

[24] On the International Space Station (ISS) astronauts appear weightless because along with the station, they are in free fall. Note that it is not the kind of free fall that takes them directly towards the earth, but in free fall that takes them *around* the Earth—in Earth orbit.

[25] To be completely accurate, the equivalence of gravity and acceleration is true only for gravitational fields that are strictly homogeneous. That is, accelerated in exactly the same way, in exactly the same direction and at exactly the same rate and thus a researcher inside a cabin could not distinguish acceleration from gravity. But in actuality, real gravitational fields are always to a certain extent inhomogeneous.

Table 13.2 Interstrellar voyages at constant 1 g acceleration

Earth time	Ship time	Range attained.	
(years)	(years)	(light years)	(parsecs)
1	0.9	0.4	0.2
10	3.0	9	2.75
100	5.2	99	30.6
1000	7.5	999	306
10,000	9.7	9999	3067
10,000	12	10^5	3×10^4
1,000,000	14	10^6	3×10^5
10^{10}	24	10^{10}	3×10^9

The values of elapsed ship time and range attained apply to continuous acceleration, and do not allow for deceleration at the target. Table courtesy of Iain Nicolson

all those objects faithfully keep their relative positions moving at a constant speed). You would feel weightless.

Thought Question 13.4

Imagine that you hold a laser light and sweep it across the sky between two stars that are separated by an angular distance of 45° and are each 20 light-years away from Earth. About 20 years from now, you'd see your laser light first make a dot (an extremely dim one!) on the surface of one star and a few seconds later make a dot on the surface of the second star. That is, your laser dot would seem to have travelled the many light- years from the first star to the second star in just a few seconds—which means at a speed far in excess of the speed of light. Is this correct?

Now imagine that you're floating freely inside the spaceship. and the other objects are floating, as well, and you feel totally weightless. Would this that mean you are far away from all gravitational influences, such as stars, planets, and other massive bodies, somewhere in deep space? It is very difficult to say, as it can be said that in fact you're really in a falling elevator, as the observational result—weightless, is the same.[24]

---

[26] What matters is the size of the region and the duration of the observations. With an infinitely small region of spacetime, one can always find a reference frame—an infinitely small elevator cabin, observed over an infinitely brief period of time—for which the laws of physics are the same as in special relativity.

Therefore, if you are in a free-falling elevator, you cannot decide whether you are in a gravitational field, and that even in a gravity-free region of space, objects would fall towards the floor providing the room or spaceship is being accelerated. Equally, even in a gravitational field, such as on the Earth, you would drift weightlessly through space, as long as the elevator is in free fall.

Such was the strength of Einstein's belief in this idea, that he was totally convinced that this inability to distinguish a region with a gravitational field from one without was not just restricted to observations of falling objects. He stated that no experiment, or manipulation of the laws of physics, can tell us whether we are in free space or in a gravitational field.[25] This statement came to be known as the *Equivalence Principle*, and one of its more important consequences is that in a reference frame that is in free fall, the laws of physics are the same as if there were no gravity at all: specifically, the laws of physics are those of Special Relativity (Table 13.2).[26]

Problems

1. Consider a spaceship moving at 70% the speed of light, 0.7c, or 210,000 kms^{-1}. Calculate the time dilation observed by a stationary observer.
2. If the spaceship in question1is on a 25-year mission at a velocity of 70% of the speed of light, how long will any astronauts have aged?
3. Consider a spaceship that is moving with a velocity 99.0% the speed of light, or 0.99c, and is travelling to Ross 154 at a distance of nearly 10 light years. Light years, as measured from the Earth. Calculate the distance from Earth to Ross 154 as measured by the astronaut.
4. On board the spaceship in question 3 is an astronaut with a mass of 100 kg. Calculate the mass of the astronaut as measured by the observer.

Chapter 14

Black Holes

14.1. Black Hole Basics

We now discuss an object that everyone has heard about, but not many really understand—a black hole.[1] It is one of those things that has gripped the public's imagination, what with its fabled inescapable pull of gravity and its possible use as a means of stellar transportation. However, it will surely come as a surprise to many of you to know that although a rigorous description of a black hole would entail a thorough background in tensor calculus and general relativity, the broad description of such an exotic object is quite simple, and it is very easy to calculate a few of its basic characteristics. Let us begin.

Following the previous section's description of the formation of a neutron star[2] and subsequent supernova, if the core of the star contains about 3 or more solar masses, nothing will stop its collapse even beyond the neutron

---

[1] The term black hole, was coined by John Wheeler in the latter 1960s. However, that such objects might exist is not a new idea, as Cambridge professor and amateur astronomer John Michell wrote in 1783. He suggested that stars with escape velocities exceeding that of light might exist. Furthermore, in1796, the mathematician Pierre-Simon, Marquis de Laplace, made similar calculations using Newton's theory of gravity and called the objects "dark bodies."

[2] There are some theoretical astrophysicists who have suggested that for the most massive stars, a black hole may form at the centre of such stars *before* a neutron star can form.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_14

degeneracy stage. In fact, the core will collapse to an object with zero radius! Consider this statement for a moment. Something with zero radius has no physical size; it is not very tiny, or even really, really tiny. It has no size at all.

The core collapses, and thus its density and surface gravity increase. If it collapses to zero size, then the gravity becomes infinite, and that is a LOT of gravity. This entity, which has no physical size yet has infinite gravity, is called a *singularity*. Note that this is not a black hole.

To fully understand how and why black holes exist is beyond the scope of this book, but we should mention that the great scientist Albert Einstein and his theory of general relativity really started the whole ball rolling. He was the first person to combine space and time into a single entity—space-time—and it is his equations that showed that gravity could be portrayed as a curvature of space-time. Along comes the astronomer Karl Schwarzschild, who uses Einstein's equations and solves them to give the first-ever general relativistic description of a black hole. We now call them *Schwarzschild black holes*, as his solutions are for non-rotating, electrically neutral black holes, to distinguish them from rotating, charged black holes (see Sect. 14.4).

What Schwarzschild showed is that, if the conditions are right (say, if matter is packed into a small enough volume of space) then space-time can curve back on itself. This means that an object (or light) can follow a path (also known as a geodesic) into a black hole, but inside the black hole, the curvature of space-time is so extreme, and there exists no path leading out. Once in, you stay in!

14.2. The Singularity

Before we go any further, we need to make a slight detour to discuss *escape velocity*. This is the velocity an object needs to escape the pull of gravity of a celestial object. For instance, a spaceship has to go about 11 km per second to escape from Earth. What dictates the value for the escape velocity depends on two things: the mass of the celestial body and the distance from the object escaping to the center of the celestial body. Thus, if something were very massive, or very small, then the escape velocity would increase. A point may be reached when even light, which is the fastest-moving thing in the universe, may not be able to achieve escape velocity, and since the light would never escape, the object would never be seen.[3]

---

[3] In 1783, the British astronomer Rev. John Mitchell realized that using Newton's laws of gravity, a situation could occur whereby an object 500 times the radius of the Sun but with the same density would have an escape velocity greater than the speed of light. Although he didn't know it, he was talking about a black hole.

Let us return now to the singularity. At the heart of every black hole lies a singularity, or that is what is currently believed. This is an object[4] that has zero radius, and thus zero volume, and infinite density, and space-time curves infinitely. It is a one-dimensional point which contains a huge mass in an infinitely small space where the known laws of physics cease to apply. As the American physicist Kip Thorne described it, it's "the point where all laws of physics break down".

The fact that a singularity exists at all is often taken as evidence that the theory of general relativity has broken down. This really isn't unexpected when we think about it, as it occurs in conditions where quantum effects become dominant. It may be that some future theory of quantum gravity (such as current research into superstrings) will be able to describe black holes without singularities, but such a hypothesis is as yet undiscovered.[5]

With our current understanding of black holes and the singularities contained therein, it is reasonable to assume that we have a fair description of its overall features, with the singularities that arise hidden within the black hole. But an idea has been developed that maybe there exist singularities that are not hidden inside a black hole and are called naked singularities. Within the laws of general relativity, a naked singularity is, at the moment, a hypothetical singularity but without an event horizon (see later). In a black hole, the singularity is completely enclosed by a boundary known as the event horizon, inside which the gravitational force of the singularity is so strong that light cannot escape. Thus, objects inside the event horizon—including the singularity itself—cannot be directly observed. By contrast however a naked singularity would be observable from the outside.

Naked singularities, if they did exist, would be significant because it means that it would be possible to actually observe the collapse of an object to infinite density. However, this idea does cause problems for general relativity because general relativity cannot describe what happens to space-time near a naked singularity. In the black holes we have discussed, this is not a problem, as we, as outside observers can never observe the space-time within the event horizon.

So far naked singularities have not been observed in nature although there are a few circumstances that could give rise to one, but these are purely mathematical predictions. For instance, if we consider rotating black holes, then a singularity, spinning rapidly, can become a ring-shaped object, resulting in two event horizons which draw closer together as the spin of the

---

[4] If something that has no size can correctly be called an object.

[5] It has been suggested that by its very nature, we will never be able to fully describe or even understand the singularity at the centre of a black hole.

singularity increases. When both event horizons meld, they shrink toward the rotating singularity and ultimately expose it to the universe. Also, a singularity rotating very fast might arise by the collapse of dust or by a supernova of a fast-spinning star. It is currently believed that no naked singularity exists in the universe.[6]

14.3. The Event Horizon

There will be an area of space (usually spherical[7] in simple descriptions)) surrounding the singularity where the escape velocity will be so high not even light cannot escape. This sphere of space within which the escape velocity is equal to, or higher than, the escape velocity of light is what we call a black hole.[8] Thus, inside a black hole, you would have to go faster than light to escape its gravitational pull; outside a black hole, you would not. The boundary between these two regions is called the *event horizon*; any event that occurs within the horizon is forever invisible to an outside observer, and according to many gravitational theories, an event horizon forms before the singularity of a black hole.

Another way to think about it is like this. If you are inside the event horizon, all the paths that light could take are warped by the immense gravity, which eventually causes the light to fall farther into the black hole. Once a particle (or photon) is inside the horizon, moving further inward is as unavoidable as moving forward in time.

Thought Question 14.1

Does the Earth have an Event Horizon?

14.4. Rotating and Kerr Black Holes

So far, we have discussed an ordinary black hole, and by ordinary, I mean a non-rotating, stationary black hole. But as you would suspect, life isn't like that, and neither are most black holes; they rotate.

---

[6] The reasons why naked singularities cannot exist has been given the name "The Cosmic Censorship Hypothesis" and deals with the geometry of spacetime. Basically, what is implies is that if they do exist, Einstein's General Relativity needs a revision.

[7] Rotating black holes are not spherical but have an oblate shape. Their description however is far more complex and will be discussed in the rotating black hole section.

[8] In some cases, a supernova remnant that does not have a central pulsar or neutron star may have a black hole at its center.

All objects in the universe—planets, stars, including the Sun, galaxies, black holes—spin. A rotating or spinning black hole is one that possesses angular momentum,[9] and rotates about one of its axes of symmetry. Current ideas suggest that rotating black holes form from either the gravitational collapse of a massive spinning star or from the collapse or collision of a group of compact objects, stars, or gas recalling that everything spins. As all known stars rotate, it is expected that all black holes in nature are rotating black holes. Another possibility as to the origin of rotating black holes comes from the detection of gamma ray bursts, or GRB's, which are believed to be the brightest and most energetic events in the universe, and created by a *Collapsar*, which is the term given to a *Hypernova*. These are extreme supernova events, originating from stars with a mass of greater than 30 $M_{\odot}$ whose kinetic energy is an order of magnitude higher than most supernovae, and a luminosity at least 10 times greater. Needless to say, these are very rare events.

Now, as you can imagine, solutions to Einstein's equations are very complex, but there are four of them that are important here, because two of them describe rotating black holes, and two do not.[10] The two that rotate are called Kerr and Kerr–Newman black holes, named after the people who came up to solutions to the equations. Furthermore, and this may surprise you, but there are only three properties that can describe a black hole. These are the mass, the momentum, and its electric charge. The values of the these characterize the black hole and are determined by observing its electromagnetic and gravitational fields. Any other variations in the black hole either escapes to infinity or are consumed by the black hole. This is because anything happening inside the black hole horizon cannot affect events outside of it.

What all this means is that a Kerr black hole is rotating but has no electric charge, whereas a Kerr-Newman black hole rotates and has an electric charge. Research both observational and theoretical suggests that black holes are expected to rotate, because of the method of formation, such as the collapse of rotating stellar objects, but have an essentially zero electric charge, since any net charge will quickly attract the opposite charge and be neutralized. Thus, the term "astrophysical" black hole is usually reserved for the Kerr black hole and since most, if not all observed astronomical objects do not possess an appreciable net electric charge, only the Kerr solution has astrophysical relevance.

---

[9] Don't worry if you are unfamiliar with angular momentum, it is just a physics attribute that all spinning objects possess with the proviso that the total angular momentum of a closed system remains constant.

[10] These are Schwarzschild and Reissner–Nordström black holes, with the former having no electric charge, and latter having a charge.

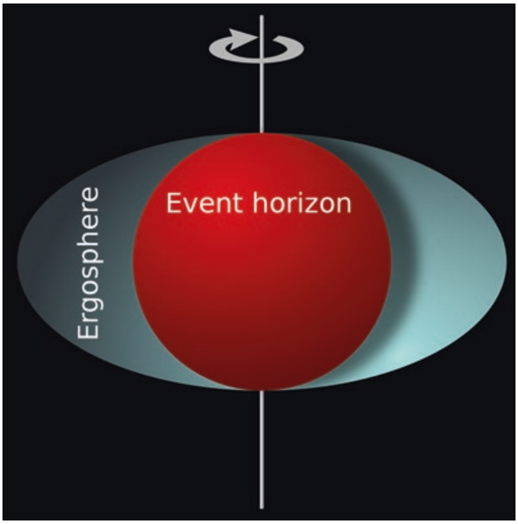


Fig. 14.1. Ergosphere of a rotating black hole, Courtesy of Messer Woland

Before we leave spinning black holes, there is one more phenomenon that is very interesting. Consider a rotating black hole, that has, obviously, an immense gravitational field. Due to both the spinning and gravitational field, the black hole actually "drags" the space around it. This has been called frame dragging, the area that is "dragged" is called the *ergosphere*, see Fig. 14.1. As the black hole rotates, it essentially twists spacetime in the direction of the rotation with a speed that decreases with distance from the event horizon.

Looking at the figure, you will see that there are two surfaces around the rotating black hole. The inner sphere is the event horizon, as well as the inner boundary of a region called the ergosphere. The oval-shaped surface, the touches the event horizon at both poles, is the outer boundary of the ergosphere. What this means is that you were in the ergosphere, it would be impossible for you to remain stationary above the event horizon. An observer located a long way from the black hole would see you rotating around it, as you were "dragged" by the rotating black hole.

In 1969, the famous mathematical physicist Sir Roger Penrose[11]suggested that the ergosphere was the region of spacetime that contains the rotational energy of the black hole, and, hypothetically, one could extract this energy.

---

[11] Sir Roger Penrose was awarded one half of the 2020 Nobel Prize in Physics "for the discovery that black hole formation is a robust prediction of the general theory of relativity".

Notice that the ergosphere is outside the event horizon, and thus it would be possible for objects that had entered that region with sufficient velocity, to actually escape from the gravitational pull of the black hole. By doing this, the object could gain energy by entering the ergosphere, then escaping from it, thus extracting some of the black hole's energy. This is called the Penrose Process, and recent supercomputer calculations now suggest that this process may well be the origin of the observed high-energy particles emitted from quasars and other active galactic nuclei.

We can now give a fairly accurate description of the black hole; it is the area surrounding a singularity where the gravity is so strong that light cannot escape from within the black hole. Furthermore, the event horizon is not a physical entity, like a membrane, it just distinguishes between two areas of space: outside the black hole, and inside it. The astronomer Karl Schwarzschild used Einstein's equations and solved them to give the first-ever general relativistic description of a black hole. What Schwarzschild showed is that, if the conditions are right (say, if matter is packed into a small enough volume of space) then space-time can curve back on itself. This means that an object (or light) can follow a path (also known as a geodesic) into a black hole, but inside the black hole, the curvature of space-time is so extreme, and there exists no path leading out. Once in, you stay in! We now call these black holes *Schwarzschild black holes*, as his solutions are for non-rotating, electrically neutral black holes, to distinguish them from rotating, charged black holes.

14.5. The Lifetime of a Black Hole and Hawking Radiation

It was always believed that black holes would exist forever, until 1974 when the world-famous physicist Stephen Hawking[12] announced to the world that they could "evaporate" albeit over an extremely long time. His theory is, as expected, very complex and mathematical, taking into account both General Relatively and Quantum Mechanics. But I will attempt a simple description. It goes something like this.

Stephen Hawking predicted that due to quantum effects, energy fluctuations arising from the vacuum near the event horizon of the black hole could give rise to particle-antiparticle pairs of virtual particles.[13] One of the par-

---

[12] Stephen Hawking died in March 2014, aged 76, and was undoubtedly one of the world's greatest scientists.

[13] Virtual particles exist in pairs, a particle and antiparticle which can be of any kind. These pairs exist for an extremely short time, and then annihilate. But in some cases, the pair may be torn apart using external energy so that they avoid annihilation and become actual particles, such as in the scenario we are describing.

ticles falls into the black hole whilst the other escapes before they have an opportunity to annihilate each other. The net result is that, to someone observing the black hole from a distance, the black hole has emitted a particle. Now, since the particle that is emitted has positive energy, the particle that gets absorbed by the black hole has negative energy relative to the outside universe. This results in the black hole losing energy, and thus mass, if we use Einstein's famous eq. $E = mc^2$. So, over time, black holes that do not gain mass through via means would shrink and ultimately vanish, in a process called black hole evaporation. For all except the smallest black holes, this would happen extremely slowly.

If the theory is correct, and black holes do evaporate from Hawking Radiation, a 1 $M_\odot$ black hole would evaporate over 10^{64} years which is considerably longer than the current age of the universe,13.8×10^9 years. Some gigantic black holes, such as the ones located in the centers of the most massive galaxies, for instance, TON 618, a hyperluminous, broad-absorption-line, radio-loud quasar located near the border of the constellations Canes Venatici and Coma Berenices, and at a distance of 10.4 billion light-years, has a mass of 6.6×10^{10} $M_\odot$ that implies it evaporate over a timescale of up to 10^{106} years.

However, controversies still surround the hypothesis even though Hawking radiation is generally accepted by the scientific community. Basically, it concerns the loss of information that occurs, if the particle falls into the black hole, as it challenges the principle that information cannot be created or destroyed. So, like most axioms of quantum physics, both observable and testable experiments concerning Hawking Radiation are nearly impossible to conduct, and as such a definitive proof is still inconclusive.

14.6. Size of a Black Hole

The size of a black hole is determined by its mass. The more mass, the bigger the black hole becomes. As we have discussed earlier, the event horizon is the boundary between the universe and the forever isolated region of extreme curved space-time[14], we know as a black hole. The radius of the black hole, that is, the distance from the singularity to the event horizon, is called the *Schwarzschild radius*, R_{Sch}. It is

$$R_{Sch=}\frac{2GM}{c^2}$$

---

[14] There are some relativists who propose that in an unimaginably distant future, black holes will indeed "evaporate." We will long be gone for this issue to worry us.

where G is the gravitational constant, M is the object mass, and c is the speed of light.

We can however simplify this to

$$R_{Sch} = 3\mathrm{M}_{\odot}$$

Which actually gives quite a good result if we make sure we always describe the mass, M, in solar masses, $\mathrm{M}_{\odot}$, and the radius is in kilometres, km.

One point that must be made is that many people believe, erroneously, that black holes will just vacuum you up regardless of your distance from them. This is wrong. For instance, if the Sun were to suddenly turn into a black hole, of diameter 6 km, the orbits of the planets would not change. It would get dark and very cold, true, but the gravitational pull would remain the same. The gravitational effect of a black hole only becomes extreme if one gets very close to it. Or to put it another way—there is a specific distance from a black hole where Newtonian physics fails, and Einsteinian physics takes over. So, providing we are a reasonable distance away, there is nothing, relatively speaking, to worry about.[15]

This distance can be determined of we know the mass of the black hole, and thus its Schwarzschild radius, R_{Sch}.

It is given by.

Distance from event horizon where Einsteinian gravity dominates = 3 R_{Sch}.

It is literally just three times the Schwarzschild radius, R_{Sc}, or nine times the mass of the black hole.[16] However, once you are within this distance, you will, inevitably, plunge toward the event horizon, taking a spiral path, and then cross over into the event horizon into the black hole.

Math Box 14.1 The Size of a Black Hole

To determine the approximate radius of a black hole, known as the Schwarzschild radius ($\mathrm{R}_{\mathrm{Sch}}$), we need to know the progenitor mass, in relation to the Sun's mass, $\mathrm{M}_{\odot}$.

The radius is given by the very simple formula:

$$\mathrm{R}_{Sch} \sim 3\mathrm{M}_{\odot}$$

---

[15] That is, if we ignore the immense amount of radiation being formed around a black hole and the debris from stars that have been literally torn apart.

[16] I imagine that you are surprised at the simplicity of these two equations, considering we are dealing with General Relativity, which has extremely difficult mathematics. I know I was when I first saw and used them.

where R_{Sch} is in kilometres.

Example:
χ^2 Orionis is a B-type supergiant star in the constellation Orion, with an estimated mass of $42M_{\odot}$. Determine the radius of a black hole that may form when the star eventually dies as a supernova and leaves behind a remnant core of mass ~ $5M_{\odot}$.

$$R_{Sch} \sim 5M_{\odot}$$

$$R_{Sch} \sim 3\times 5$$

$$\sim 15\ \text{km}$$

Thus, a 42 $M_{\odot}$ star could form a black hole of diameter 30 km.

Where does Einsteinian gravity take over from Newtonian gravity?

At a large distance, a black hole exerts a gravitational force according to Newton's Law. However, a point is reached whereby Newton's Laws are no longer valid, and the gravitational effects are now explained using Einstein's general relativity. The distance from a black hole where this change occurs is ~3 R_{Sch}.

Example:
χ^2 Orionis could form a black hole with a Schwarzschild radius of approximately 15 km. At what distance from the black hole does gravity increase from what Newton's Law predicts?

Using the above formula, this distance is approximately 3 R_{Sch}, thus:

$$3R_{Sch} \sim 3 \times 15$$
$$\sim 45\ \text{km.}$$

At 45 km from the black hole, the gravitational force will increase to considerably more than that predicted by Newton's laws.

Table 14.1 Black hole classifications

Class	Approximate mass	Approximate radius
Observable Universe (I have included this if the observable universe were to be a black hole, as some people speculate)	$\sim 10^{23}$ $M_{\odot}$	$\sim 13.8 \times 10^{9}$ *ly*
Supermassive black hole	10^5–10^{10} $M_{\odot}$	0.001–400 AU
Intermediate-mass black hole	10^3 $M_{\odot}$	10^3 km ≈ R_{Earth}
Stellar black hole	10 $M_{\odot}$	30 km
Micro black hole	Up to M_{Moon}	Up to 0.1 mm

Thought Question 14.2

If the mass of a black hole were to triple, how much larger would the Event Horizon be?

For an example, see Math Box 14.1.
See Table 14.1. for a very simple black hole classification.

14.7. Spaghettification

This section deals with even more esoteric properties of black holes, if such a thing can be said. The two topics are, naturally, due to the immense and ever-increasing gravitational field that one would encounter approaching the event horizon. We will deal with the odd sounding, but entirely appropriate topic of *Spaghettification*.

Let us imagine a (brave) astronaut approaching a black hole of say, 12 $M_\odot$, at a distance of about 400 km. Nothing untoward is happening and the astronaut will be experiencing gravity as defined by Newtonian physcis. However, as one gets closer, Einstienian physics start to become come into play, and tidal forces start to “tug” on the astronaut, in such a way that the the gravitational pull on one’s feet would be stronger than the pull on one’s head, provinding they are approaching feet first. The astronaut would, literally, be stretched like spaghetti by the gravitational gradient, that is, the difference in strength from head to toe the close one gets to the event horizon.

Also, in addition, to being stretched, the astronuat would be compressed, with the right side pulled to the left, and the left side pulled to the right, horizontally squeezing the person. The astronaut would start to experience this, at about 300 km from the black hole, note that this is well outside the well outside the Schwarzschild radius of 36 km. Thus, the astronaut would be stretched into a thin stream of matter and cross over the event horizon. Note that at the event horizon itself, the tidal disruption is incredible; for a ~ few $M_\odot$ black hole, an astronaut would experience a difference of gravity of about 10^9 Earth gravities, more than enough to tear apart most things.

By a quirk of nature, or rather mathematics, with supermassive black holes, although the phsyics is the same, the point at which spaghettification occurs is actually within the event horizon, so, theoretically, one could travel over the event horizon and surivie, all things considered. For example. With a supermassive black hole of 10,000$M_\odot$ the spaghettification would start at a distance of 3200 km, which is well inside the Schwarzschild radius of 30,000 km. For a supermassive hole, the tidal force at the event horizon is smaller than the example given previously for a ~ few $M_\odot$ black hole by a typical factor 10^{10-16} and could be survivable. But remember however, all paths lead to the singularity, where the tidal forces are *infinite*.

Not unsurprisingly, spaghettification has actually been observed. In 2018, astronomers observing a pair of colliding galaxies called Arp 299, at a distance of 150 million l.y., and using a both radio and infrared telescopes, witnessed a 20×10^6 $M_\odot$ black hole tearing apart a sizable nearby star. Furthermore, in 2021astronomers observed a Tidal Disruption Event (TDE) believed to be a star in the process of being spaghettified, by looking at the spectral absorption lines around the poles of a distant black hole, mass 5.4×10^6 $M_\odot$, located in a distant galaxy.

14.8. Time Travel

One of the stranger aspects of general relativity concerns the passage of time, as measured by two different observers, near a black hole. A classic thought experiment that outlines this idea, well known to astronomers and educators, is where you have one person, stationary, a very large distance from a black hole, and another, brave person who agrees to go on a journey to the black hole. What would each of them see?

What follows is an accurate description of the events of such a scenario; it may seem strange and even unbelievable, but it is real, or would be if we ever get close to a black hole. We also have to make note that light emerging out of a strong gravitational field will show a redshift, called a gravitational redshift. This is due to gravity and not the Doppler effect.

So, let’s begin the journey.

You and the black hole explorer find a black hole with a mass of 10 $M_\odot$ that naturally has a Schwarzschild radius of 30 km. You stay in your mothership, in a circular orbit, at a reasonable distance from the event horizon, of a few thousand kilometres. At this distance the spaceship will be under Newtonian physics in a stable orbit, and so there are no worries about getting “sucked in.”

You both have two identical, extremely accurate and synchronized clocks with numerals glow with blue light. Now, you remain on the mothership spaceship with your clock, whilst the explorer, possibly in a small space pod, gradually flies off toward the black hole. From time-to-time, you glance at her clock as she approaches the event horizon. What you see is striking. Her clock seems to be ticking slower than your clock, on the mother ship, and its numerals are red instead of blue as its light becomes increasingly redshifted. Actually, everything gets red shifted not just the light from the clock. The closer she gets to the event horizon, the slower the clock ticks, and the redder the light becomes. In fact, when she gets to a distance of around 10 km above the event horizon, her clock is ticking only half as fast as your clock on the mothership. Remember that this isn't just an effect on the mechanism of her clock—it is time itself that is running slower.

You keep watching, as she gets ever closer to the event horizon,[17] but now something extraordinary happens. From your point of view aboard the mothership, her clock stops, and she will *never* cross the event horizon. Time will come to a stop for her, and her clock and she will remain, from your point of view, forever fixed above the event horizon. In fact, it would literally take *forever* for her to cross the event horizon.

Now, what would she see from her point of view as she approaches the event horizon?

She is now watching her the clock, but, because both her and her clock are traveling together, time is running normally, and its numerals stay blue. From her point of view, time seems to neither speed up nor slow down. She now glances at your clock on the mothership and observes your time passing progressively faster and faster and your light becoming increasingly blue shifted. But now, she and the clock pass through the event horizon. There is no membrane, barrier, or hard surface. Remember that the event horizon is not a physical boundary, but a mathematical boundary. From her point of view, the clock keeps ticking normally, and she is now inside the event horizon, and has left our observable universe.

There is one situation where she could survive passing over the event horizon, and that occurs with a much larger black hole, such as a supermassive black hole residing in the centres of a galaxy. For instance, $10^9 M_\odot$ black hole has a Schwarzschild radius of 3 billion km (the distance from our Sun to Uranus). With such a supermassive black hole the gravitational force at the event horizon is such that its tidal forces are weaker and hence survivable. She could safely plummet through the event horizon, but once over it, she will inevitably succumb to spaghettification.

The final part of the story is even more fantastical. We are now getting close to pure theory, and it is probable that this may not ever be possible, but there could be a way to travel into the future. Imagine that our explorer has the most powerful spaceship engines ever made, fitted into her little space pod. This time she stays quite a distance from the event horizon, and

---

[17] To be completely accurate, for stellar mass black holes, her light would have been red shifted completely into the radio regime, and so would she be invisible to the naked eye. In fact, the light is so far redshifted that no known telescope could detect it.

even though she is experiencing a considerable gravitational pull from the black hole, the engines prevent her from being dragged completely toward it. From her point of view, time is passing as normal, but from our point of view aboard the mothership, her clock is ticking slower. She decides that she wants to spend a considerable amount of time studying the black hole, say 1 year. Then powering up the engines she returns to the mothership. Whilst for her only 1 year has passed, for us on the mothership, more time will have passed, because from our perspective, her clock was running slower. It could be that from our point of view, she may have been away for 10 or 20 years, it would all depend on how close she got the event horizon, and mass of the black hole. However, the point being made is she will have travelled into the future. She will have aged only 1 year, in the, say, 20 years she has been away.

Throughout all of this discussion we have only lightly touched upon a few important points. Firstly, the process of spaghettification would, with stellar mass black holes, have torn her into atoms, and of course it is thought that all black holes spin, thus the space above the black hole would be "dragged" around, so the mothership would have had to be in orbit about the black hole for us to observe her, and the gravitational redshift would course render her invisible to us.

Nevertheless, bizarre as this seems, it is all correct according to Einstein's theory.

14.9. Finding Black Holes

There remains, however, one slight problem. If black holes are literally invisible to us, how can we ever detect them?

We do this by looking for the effect that they have on objects nearby! First, we search for a star whose motion, determined by measuring the Doppler shift in its spectrum, indicates that it is a member of a binary star. (The Doppler shift is the change in the appearance of light from an object that is moving away from or towards an observer.) If it proves possible to see both stars, give up on that object. The search is for a binary system where one companion is invisible, no matter how powerful the telescope used. However, just because it is unseen does not mean it is a black hole candidate; it may just be too faint to be seen, or the glare from the companion may swamp out its light. It could even be a neutron star.

Thus, further evidence is needed to determine if the invisible companion has a mass greater than that allowable for a neutron star. Kepler's laws are used at this point to determine whether the star, or rather the invisible object, has a mass greater than 3 solar masses. If this is so, then the unseen companion may be a black hole. Further information is still needed, however, and this may appear in the form of X-rays, which can arise either from material flowing from a star into the black hole or from an accretion disc that has formed around the black hole. Either way, the presence of X-ray emission is a good indicator that a black hole may be the unseen companion object.

Of course, the measurements as just stated are a bit more complicated than this. For instance, it is known that neutron stars can emit X-rays and have an accretion disc. So, careful analysis of the data is necessary. However, a few candidates are known, and one is even visible to the amateur astronomer…or perhaps we should say that the companion star to the black hole is visible.

Let us now look at a couple of Black Hole candidates.

Cygnus X-1	HDE 226868	19h 58.4m	35° 12′	July
8.95 m				

This is one of the strongest X-ray sources in the sky, and possibly the most convincing candidate for a black hole. Its position is coincident with the star HDE 226868, which is a B0Ib supergiant of magnitude 9. It lies about 0.5° ENE of Eta Cygni. It is a very hot star, of around 30,000 K, and analysis shows it is a binary with a period of about 5.6 days. Observations by satellite have detected variations in the X-ray emission on a time scale of less than 50 ms. The estimated mass of the unseen black hole companion is in the range of 6–15 solar masses. This would mean that it has a maximum radius of about 45 km. Current X-ray research suggests a mass of 14.8 stellar masses. The system is thought to be a member of the stellar association Cygnus OB3; thus Cygnus X-1 is estimated to be around 5 million years old and probably formed from a progenitor star of more than 40 solar masses. Many researchers suggest that due to the numerous similarities between the emissions of X-ray binaries such as Cygnus X-1 and active galactic nuclei, there is a common mechanism of energy generation that involves a black hole, surrounded by an orbiting accretion disk along with the usual associated jets. Cygnus X-1 is thus identified among a class of objects called *microquasars*, an analog of the quasars,[18] or quasi-stellar radio sources.

Galactic Center	Sgr A*	17^{h} 45.6 ^{m}	−29° 00′	June

This is included here, in spite of the fact that black holes are, to all intents and purposes, invisible, because there is nonetheless one place one can look

---

[18] We discuss these in the chapter on active galaxies.

and exclaim—“There lies the most massive black hole in the entire galaxy!” However, the center of the galaxy has always posed several problems to astronomers, both amateur and professional. Unfortunately, owing to the vast amount of interstellar gas and dust that lies between us and the center of the galaxy, it has been impossible to see it in visible light. However, all is not lost, as gamma rays, hard X-rays, infrared radiation and radio waves are able to pass through the interstellar medium, and so a picture can be built up of this inaccessible region. What has been learned is impressive. There is a radio source located very near or even at the exact center, called Sagittarius A*, [19] (pronounced “Sagittarius A star”), and this was the first cosmic radio source discovered. Measurements of the radio source indicate that it is no bigger than the diameter of Mars’s orbit. One of the (many) surprising results is that Sagittarius A* is stationary, which would indicate that it is very massive.[20] The current estimate for the mass is about 4.6 million solar masses. All this information suggests that at the center of our galaxy lies a supermassive black hole with a diameter of around 41 million km.[21] When you look through a telescope at the region, the field of view will be full of numerous star fields, but these lie much closer to us than to the center. However, when observing this region, you may like to contemplate the almost certain fact that even though you cannot see the center, there in your eyepiece, ever invisible, lies a supermassive black hole,[22] around which you, the Solar System and the Galaxy rotate.

Other black hole binary systems are listed in the following table.

Star	Type	Orbital period (days)	Black hole mass estimates ($M_{\odot \mathrm{Sun}}$)
LMC X-3	B main sequence	1.7	4–11
V616 Mon	K main sequence	7.8	4–15
V404 Cyg	K main sequence	6.5	>6
Nova Sco 1994	F main sequence	2.4	4–15
Nova Velorum	M dwarf	0.29	4–8

---

[19] Sagittarius A* is now believed to be made of two components: SgrA East and SgrA West. The former is a supernova remnant, and the latter is an ultra-compact, non-thermal source, i.e., a black hole.

[20] Recent analysis suggests that the density around the center of the galaxy is about a million times greater than any known star cluster. It is probably made up of living stars, dead stars, gas, and dust, and of course a black hole.

[21] To get a sense of scale, consider that Mercury is 46 million km from the Sun at perihelion.

[22] Now for a surprise. Astronomers have suggested that instead of a supermassive black hole, it is instead a mass of dark matter. Further measurements await for confirmation of this eye-opening statement. “*E A Becerra-Vergara et al, Hinting a dark matter nature of Sgr A* via the S-stars, Monthly Notices of the Royal Astronomical Society: Letters (2021)*”.

Thought Question 14.3

Consider the end results of stellar evolution. What do you think we are likely to find more of, black holes, neutron stars, or white dwarfs?

Problems

1. Determine the Schwarzschild radius of a black hole of 10 $M_\odot$
2. Determine the diameter of a black hole of 150 $M_\odot$
3. At what distance from a 371 $M_\odot$ black hole does gravity increase from what Newton's Law predicts?
4. At what distance from a 3 $M_\odot$ black hole does gravity increase from what Newton's Law predicts?

Chapter 15

Exoplanets

15.1. Introduction

Every day an announcement is made of yet another planet discovered by either Earth-based astronomers or from an orbiting telescope. As of the time of writing of this book[1] there are 5241 confirmed planets, consisting of 9169 candidates along with 3916 multiple-planet systems. No doubt this will change on a daily basis.

There is also the data from the Kepler mission that has several thousand likely candidates that will need further analysis before being confirmed. Soon, with the advent of larger telescopes and improved image-processing techniques, it will only be a matter of time before the first clear and detailed image of a non-Solar-System planet is obtained. That will be a very special day. However, until then any indication as to what these new planets will look like will have to come from the imaginations of artists and scientists (not forgetting amateur astronomers).

What is truly astounding is that the analyses from the research teams currently collecting data, suggest that there is at least one planet for every star in the Milky Way. Now, if we also make the currently held assumption that for every 100 stars that are similar to the Sun, about 20 will have planets

---

[1] January 2023. But as everyone knows, the numbers increase almost daily. A good website that shows a running total is https://exoplanets.nasa.gov

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_15

that will be Earth-sized and in the habitable zone.[2] Furthermore, if one assumes the rather conservative number of 200 billion stars in the Milky Way, this leads us to the conclusion that there are possibly 11 billion Earth-like planets with Earth-like conditions, and if one takes into account red dwarf stars, then this number increases to 40 billion. Finally, if one includes all planets around all stars, including free-floating planets,[3] then the total number could be in the trillions. Incredible!

Surprisingly many of the stars that have been reported as having planets are quite bright, and so easily within reach of small telescopes, and what's more, some host stars can be seen with the naked eye. Even though any sign of planets will be absent if these stars are observed, it is still a sobering and also wonderful thing to be able to view them and to think that circling these bright stellar objects are new worlds. And what else besides, one wonders?

15.2. Types of Exoplanets

Thought Question 15.1

Why do you think we have only discovered exoplanets in the last 30 years, and not before?

This section of the chapter is fraught with problems, as new types are being discovered, or rather, named, with nearly each exoplanet discovery. There also seems to be an overlap of several types, along with a few that are based on the physics of the interior, while others are based on the formation method, with neither school of thought agreeing. Nevertheless, the following table is included for instruction. We will then outline the most populous, and interesting, types of extrasolar planets (indicated by a "*" in the table).

---

[2] The habitable zone is a region around a star where the temperature allows liquid water to exist on a planet; that is to say, not too close to the star, otherwise water would evaporate, and not too far away from the star, or the water would freeze. Of course the energy, and thus the heat, produced by stars will depend on both the size and age of the star, so the habitable zone can be at different distances from the star.

[3] A free-floating planet, also known as an interstellar planet, nomad planet, rogue planet or orphan planet, is a planetary-mass object that orbits the Milky Way directly. They have either been ejected from the planetary system in which they formed or were never gravitationally bound to any star or brown dwarf in the first place.

Terrestrial	Gas giant	Other
Carbon planet	Chthonian planet*	Brown dwarf
Coreless planet	Eccentric Jupiter	Circumbinary planet*
Desert planet*	Helium planet	Double planet
Ice planet*	Hot Jupiter*	Mesoplanet
Iron planet	Hot Neptune*	Planetar
Lave planet	Ice giant	Protoplanet
Ocean planet*	Mini-Neptune	Sub-brown dwarf*
Sub-Earth*	Gas dwarf*	
Super-Earth*	Puffy planet*	
Massive solid planet	Super-Jupiter*	

Following are descriptions of the terrestrial planet types.

- *Desert planet*—A hypothetical type of terrestrial planet with very little water.
- *Ice planet*—A type of planet with an icy surface. There are several extrasolar ice planet candidates, such as OGLE-2005-BLG-390Lb.
- *Ocean planet*—A planet that has a substantial fraction of its mass made up of water. The surface would be completely covered with an ocean of water hundreds of kilometers deep. The most likely known candidate for an ocean planet is extrasolar planet GJ 1214 b. However, the Kepler mission, has recently discovered the ocean planet candidate Kepler-22b.
- *Sub-Earth*—A classification of planets substantially less massive than Earth and Venus. One of the first found was around a millisecond pulsar PSR B1257 + 12. The Kepler mission has since discovered the first three sub-Earths around an ordinary star, Kepler-42. As of the writing of this book, Kepler has found 45 confirmed planets that are smaller than Earth, with 17 of them being smaller than 0.8 the radius of Earth, along with 310 planet candidates with an estimated radius less than that of Earth.
- *Super-Earths*—An extrasolar planet with a mass higher than Earth's but considerably below the mass of the Solar System's smaller gas giants Uranus and Neptune (15 and 17 Earth masses, respectively). The term super-Earth refers only to the mass of the planet and does not imply anything about temperatures, compositions, orbital properties, habitability or environments similar to that of Earth. Three "super-Earth" planets circle Gliese 667C, one of three stars located a relatively close 22 light years from Earth in the constellation of Scorpio, and orbit Gliese 667C in the so-called *Goldilocks zone*—the distance from the star at which the temperature is just right for water to exist in liquid form rather than being stripped away by stellar radiation or locked permanently in ice.

Here are the gas giant planet types.

- *Chthonian planet*—A hypothetical class of celestial objects resulting from the stripping away of a gas giant's hydrogen and helium atmosphere and outer layers, which is called hydrodynamic escape. Such atmospheric stripping is a likely result of proximity to a star. The remaining rocky or metallic core would resemble a terrestrial planet in many respects. The extrasolar planet COROT-7b is likely to be the first Chthonian planet discovered, with a mass nearly five times that of Earth, a diameter 1.7 times that of Earth and an orbital period of 20.5 h.
- *Hot Jupiter*—A class of extrasolar planets whose characteristics are similar to Jupiter, but which have high surface temperatures because they orbit very close to their host star, about between 0.015 and 0.5 AU (2.2×10^6 and 74.8×10^6 km), while Jupiter orbits the Sun at 5.2 AU (7.8×10^8 km), causing low surface temperatures. Probably the best-known hot Jupiter is 51 Pegasi b, given the name Bellerophon, discovered in 1995. Recent research suggests that hot Jupiters are unlikely to have moons, similar to the Jovian moons, but rather are small, asteroid-sized bodies.
- *Hot Neptune*—A type of planet that orbits close to its star (normally less than 1 AU away), having a mass similar to that of Uranus or Neptune. Interestingly, observations have suggested that a larger potential population of hot Neptunes than previously thought may exist. The first hot Neptune discovered was Mu (μ) Arae c (HD 160691 c).
- *Gas dwarf*—A type of planet with a rocky core that has accumulated a thick envelope of hydrogen and helium, with a resulting 1.7–3.9 Earth-radii. The currently known smallest extrasolar planet that is probably a "gas planet" is KOI-314c, that has a mass similar to that of Earth's, but is 60% larger; thus one can infer from its density that it has a thick gas envelope.
- *Puffy planet*—These are gas giants having a large radius and very low density and are sometimes referred to as *hot Saturns*, as their density is similar to that of Saturn. Puffy planets could orbit close to their stars, as the intense heat from the star, along with the internal heating within the planet, may help inflate the planet's atmosphere. Currently, six large-radius low-density planets have been detected using the transit method (see Sect. 13.1 for details of this method). Several examples of puffy planets are HAT-P-1b, COROT-1b, WASP-12b and Kepler-7b.
- *Super Jupiter*—This is a type of planet that's more massive than the planet Jupiter. Exoplanets that exist at the planet/brown dwarf borderline have been referred to as Super Jupiters. The most famous example is the super Jupiter around the star Kappa Andromedae. The planet Kappa

Andromedae b was imaged around the star Kappa Andromedae. It orbits about at a distance 1.8 times that of which Neptune orbits the Sun.

Finally, here are brief descriptions of some other planet types.

- *Circumbinary planet*—This is a planet that orbits not one but two stars. Due to the short orbits of some binary stars, theoretical study suggests that the only way for planets to form in such a system is by forming outside the orbit of the two stars. The first confirmed planet was orbiting the system PSR B1620-26, comprising a millisecond pulsar and a white dwarf, located in the globular cluster M4. Another system is that in the eclipsing binary system HW Viriginis, comprising a sub-dwarf B star and a red dwarf, with the inner and outer planets having masses at least 8.47 and 19.23 times that of Jupiter, respectively, and having orbital periods of 9 and 16 years. Another example is Kepler-16b, 200 light years from Earth, in the constellation Cygnus, thought to be a frozen world of rock and gas, with the mass of Saturn. It orbits two stars that are themselves also circling each other. The planet takes 229 days to orbit the stars, whereas the planet orbits the system's center of mass every 225 days. Interestingly, the results from the Kepler mission suggests that circumbinary systems are relatively common, with all the Kepler circumbinary planets discovered being either close to or actually in the habitable zone. However, none of the planets so far discovered is a terrestrial planet. It has been suggested that in such cases, if these possessed large moons, then they could be habitable.
- *Sub-brown dwarf*—A sub-brown dwarf, or as it is also sometimes referred to, a planetary-mass brown dwarf, is an extrasolar planet formed in a manner similar to that of stars and brown dwarfs (i.e., through the collapse of a gas cloud) but with a mass below the limiting mass for thermonuclear fusion of deuterium. This mass is believed to be around 13 Jupiter masses. There is some debate however as to their true classification, as some astronomers call them free-floating planets, while others call them planetary-mass brown dwarfs. As a compromise it has been suggested that a sub-brown dwarf is a free-floating body found in young star clusters below the lower mass cut-off of brown dwarfs.

Possible candidates are

- DT Virginis c, which if confirmed would be a circumbinary planet, detected by direct imaging and is currently a planetary-mass object with the widest known orbit around a binary star.[4]

---

[4] DT Virginis, is a binary star in the constellation of Virgo. Both stars are low-mass red dwarfs with at least one of them being a flare star.

- The nature of an object that is still controversial is S Ori 52 located in the Delta Orionis cluster, with mass estimated to be between 2 and 8 Jupiter masses. Research suggests it could be a sub-brown dwarf, an isolated planet, or that it could simply be an older brown dwarf lying in the foreground of the Orionis cluster. Take your pick.

It is well known that the size and mass of an exoplanet plays a fundamental role in determining planet types. Astronomers have noticed however what appears to be a unexplained absence in exoplanet sizes, nicknamed the “radius valley,” or Fulton gap, named after Benjamin Fulton, lead author on a paper describing it. Using data from NASA’s Kepler spacecraft indicated that planets of a certain size-range were rare—1.5 and 2 times the size (diameter) of Earth, or super-Earths. It may be that this signifies a critical size in planet formation because planets that reach this size could attract thick atmospheres of hydrogen and helium gas, and grow into gaseous planets, whilst smaller exoplanets are not large enough to hold onto an atmosphere and remain primarily rocky. Alternatively, planets orbiting closer to their host stars may be the cores of Neptune-like exoplanets that have had their atmospheres. Stripped away.

As of the time of writing, the confirmed planet list goes like this.

- Neptune-like 1720
- Gas Giant 1474
- Super Earth 1518
- Terrestrial 186
- Other 20

No doubt the list of extrasolar planetary types will be expanded, along with their descriptions, as more planets are discovered along with their many disparate properties.

Thought Question 15.2

It is easy to see the planets in our solar system, yet very difficult to see planets around other stars. What simple reason is responsible for this?

15.3. Exocomets, Exomoons and Exoasteroids

In addition to the discovery of exoplanets, recent observations have suggested the presence of moons orbiting a few of the exoplanets, these are, naturally, called exomoons, as well as other objects in the exoplanetary system, namely exocomets and exoasteroids.

An exomoon or extrasolar moon[5] is defined as a natural satellite that orbits an exoplanet, and although, to date (September 2021) there have been no confirmed detection of any exomoon, there are many candidates. The problem of course is that it is very difficult to detect such small objects, at vast distances, so new and improved existing techniques have to be devised. The methods include Direct imaging, Microlensing, Pulsar timing, Transit timing effects, and the Transit method, and all require very exacting and careful measurement and analysis. At the moment, there are around 20 possible candidates, see Table 15.1.

An exocomet, is, as one would expect, is a comet orbiting an exoplanetary system, and has the official classification scientific term of *Falling Evaporating Body* (FEB).[6] The first discoveries occurred as far back as 1987, around Beta Pictoris, and there are now around 30 observed or suspected exocomets. The majority of these exocometary systems are located around young A-type stars, but a few were discovered in 2018 around F-type stars. Further discoveries include evidence of exocomets in the relatively old shell star Phi Leonis and the old F2V-type star Eta Corvi. Also, a few late B-type stars such as 51 Ophiuchi, HD 58647are believed to host exocomets.

The old F2V-type star Eta Corvi, may also be host to a cloud of comets from data from the Spitzer Space Telescope which detected the spectral signatures of water ice, organic molecules, and rock around Eta Corvi which are the ingredients of comets found in our own solar system. The star is about one billion years old and could be experiencing a bombardment of comets similar to what occurred in our own solar system around 600 to 800 million years of age, termed the Late Heavy Bombardment.

As well as exocomets, there also exists objects that can broadly be classified as exoasteroids. These are really asteroids that have an interstellar origin. The first such object to be discovered in the Solar System was 1I/‘Oumuamua in 2017, and the second was 2I/Borisov in 2019. Both had a substantial hyperbolic excess velocity, implying that they did not originate in the Solar System.

---

[5] Just as an aside, there are now, as yet hypothetical objects, called orphaned exomoons or ploonets. Formerly exomoons of another planet, ejected from their orbits around their parent planets by tidal forces during planetary migration, becoming planets in their own right. As yet none been detected, but astronomers at Columbia University have suggested that a disrupting detached exomoon may be the cause of the unusual fluctuations in brightness exhibited by Tabby’s Star, a binary stellar system with a red dwarf companion star 880 ± 10 AU from the primary. Tabby’s star has an apparent magnitude of 11.7, thus invisible to the naked eye, but visible with a 5-inch (130 mm) telescope under a suitable dark sky.

[6] Why it can’t be called an exocomet escapes me for the time being.

Table 15.1 List of possible exomoon candidates

Host star of the host planet	Planet designation	Notes
1SWASP J140747.93-394542.6	J1407b	Two possible exomoons residing in small ring gaps around J1407b.
1SWASP J140747.93-394542.6	J1407b	Two possible exomoons residing in small ring gaps around J1407b.
1SWASP J140747.93-394542.6	J1407b	Possible exomoon residing in a large ring gap around J1407b.
DH Tauri	DH Tauri b	Candidate Jupiter-mass satellite from direct imaging. If confirmed, it could also be considered a planet orbiting a brown dwarf.
HD 189733	HD 189733 b	Found by studying periodic increases and decreases in light given off from HD 189733 b. Outside of planet's Hill sphere.
HD 189733	HD 189733 b	Exo-Io candidate; the sodium and potassium data[41][42] at HD189733b is consistent with evaporating exomoons and/or their corresponding gas torus.
Kepler-409	Kepler-409b	Possible exomoon from transit timing variations.
Kepler-517	Kepler-517b	Possible exomoon from transit timing variations.
Kepler-809	Kepler-809b	Possible exomoon from transit timing variations.
Kepler-857	Kepler-857b	Possible exomoon from transit timing variations.
Kepler-1000	Kepler-1000b	Possible exomoon from transit timing variations.
Kepler-1326	Kepler-1326b	Possible exomoon from transit timing variations.
Kepler-1442	Kepler-1442b	Possible exomoon from transit timing variations.
Kepler-1625	Kepler-1625b	Possible Neptune-sized exomoon or double planet, indicated by transit observations.
KOI-268	KOI-268.01	Possible exomoon from transit timing variations.
N/A	MOA-2011-BLG-262L	Found by microlensing; however, it is unknown if the system is a sub-Earth-mass exomoon orbiting a free-floating planet, or a Neptune-mass planet orbiting a low-mass red dwarf star.
N/A	MOA-2015-BLG-337L	Found by microlensing; however it is unknown if the system is a super-Neptune-mass planet orbiting a free-floating planet, or a binary brown dwarf system.

Host star of the host planet	Planet designation	Notes
WASP-12	WASP-12b	Found by studying periodic increases and decreases in light given off from WASP-12b. Outside of planet's Hill sphere.
WASP-49	WASP-49b	Exo-Io candidate; the sodium exosphere around WASP-49b could be due to a volcanically-active Io-like exomoon.
WASP-76	WASP-76b	Exo-Io candidate; sodium detected via absorption spectroscopy around WASP-76b ^{i}s consistent with an extrasolar toroidal atmosphere generated by an evaporating exomoon.
WASP-121	WASP-121b	Exo-Io candidate; the sodium detected via absorption spectroscopy around WASP-121b is consistent with an extrasolar gas torus possibly fuelled by an exo-Io.

Courtesy of Wikipedia

‘Oumuamua, formally designated 1I/2017 U1, was the first known interstellar object ever detected that passed through the solar system, is thought to be an asteroid, and in March 2021, astronomers presented an idea that ‘Oumuamua may be a piece of an exoplanet similar to Pluto, but from beyond our solar system.

2I/Borisov, originally designated C/2019 Q4 (Borisov), [8] is the first observed rogue comet with an origin from beyond the solar system. Unlike ‘Oumuamua, which looked similar to asteroids found in the solar system, 2I/Borisov’s nucleus was surrounded by a coma, a cloud of dust and gas, familiar to solar system comets as they reach perihelion with the Sun. It is estimated that each year, several interstellar objects similar to ‘Oumuamua pass inside the orbit of the Earth, with as many as 10,000 are passing inside the orbit of Neptune on any given day.[7]

Finally, an asteroid belt, just the like the solar system’s was found orbiting a star that is very similar to that of the Sun. By using the Spitzer Telescope infrared capability, a dust disk was detected around the sixth-magnitude star HD 69830 in Puppis. This dust disk indicated a ring of rubble about 25 times more massive than of our solar system’s asteroid belt.

Thought Question 15.3

Are exoplanets still being formed, along with exomoons and exocomets?

15.4. Techniques of Detection

No matter what method is used to detect an exoplanet, the point to understand here is that the light from the planet, which is just reflected light from its host star, is exceptionally faint, when compared to the host stars’ light. This means that the light from the star overwhelms such a small amount of “planetary light.” At the time of writing of this book, there are very few extrasolar planets that have been observed directly, and even fewer resolved from their host star. What this means is that astronomers need other methods to detect exoplanets, and these are referred to as indirect methods. Which method to use is of course dictated by the system being observed, available instrumentation, etc.

---

[7] It is now believed that many comets, and asteroids, now in the solar system, may have their origins in interstellar space, but have subsequently been captured by the sun’s gravity.

To discuss each of the methods in detail is beyond the scope of this book; however, a brief outline of each method will be presented. As you can imagine, all of the methods use a considerable amount of mathematics, as well as an equal amount of statistics. The latter topic is very important, as the possible detection of an exoplanet has to be rigorously analyzed to make sure that what is observed is an exoplanet, and not, for instance, an artefact of the instrumentation or something similar arising from the data reduction techniques used. It may even be a host of other astronomical objects that lie between the star and observer. It's often wise to follow up the possible detection by using an entirely different method to confirm its discovery. You can imagine the problems that arise if a possible exoplanet is discovered using one method but is not confirmed by using any and all of the other methods. In such a case one needs to go back and analyze the data very carefully. Now let's look, briefly, at the various detection methods.

Radial Velocity

One of the main techniques used for detecting these planets makes use of the Doppler effect, which is the change in the wavelength of light from an object that is moving away from or towards an observer. The gravitational pull of a large planet orbiting a star causes the star to wobble slightly. When a star wobbles towards Earth, the star's light appears from Earth to be shifted towards the blue part of the visible light. When the star wobbles away from Earth, the opposite effect occurs.

The Doppler shift is proportional to the speed with which the star approaches or recedes from an observer on Earth. Unfortunately, the Doppler shift caused by the wobbling of stars with companion planets is very small; the wavelength of the star's light changes by only about 1 part in 10 million under the influence of a large, Jupiter-sized planet. For example, the Sun's "wobble speed" is only about 12.5 m/s.

The wobble motion of a star with a planetary companion can provide a great deal of information about the star's companion planet, including an estimate of its mass and the size and frequency of its orbit. The *orbital period* of a planet (the time it takes the planet to complete one full revolution around its star) is equal to the time it takes the star to finish one wobble cycle. The size of the star's wobble is also proportional to the size of the planet's orbit, and by using Kepler's third law of planetary motion, which states that the cube of the average distance between two orbiting bodies equals the orbital period squared, the distance between the star and its companion can be determined. Knowing that the orbital period is the same as the period of the star's wobble, one can calculate the average distance between a star and its companion.

Timing Variations

This technique uses the fact that several astronomical objects have themselves some sort of periodic phenomena associated with them. For example, pulsars, pulsating variable stars, eclipsing binary star systems, etc., and the possible addition of an exoplanet orbiting one of the aforementioned objects will have an observable effect on their periodic phenomena. The method can be quite accurate, especially in the case of the pulsars; however, let's not forget that pulsars aren't that common, so the accuracy is balanced by their rarity. Nevertheless, this technique was used to discover exoplanets around the pulsar PSR 1257 + 12 in 1992, making it the first ever discovery of planets outside our Solar System.

Transit Photometry

Whereas the radial velocity method is an excellent way of measuring the exoplanet's mass, the transit method can be used to determine the radius. This is relatively easy to explain; the exoplanet crosses, or *transits*, across the face of the host star. As a consequence, a minute amount of light is blocked, and thus the total amount of light from the star is reduced briefly. Of course, the amount of light depends on both the size of the star and exoplanet. The method is not without problems, however, as there are two major drawbacks. Firstly, the planet and star have to align with respect to the observer. If not, there will not be a transit. See Fig. 15.1 for an idea of what this means.

The second problem arises from the fact that a single transit is not good enough for confirmation. Other methods will be needed. In the Kepler mission, the percentage of transits that were in fact later found to not be transits at all but caused by other effects was as high as 40%. And the problems don't stop there either. If, say, an exoplanet was detected around a red giant star, then the inherent variability of the star can confuse the issue, as it would be difficult to separate the transit signal from the star's variability signal.

But it's not all doom and gloom, as it still is a very good method for determining the size of a planet, as well as its atmosphere. An example of such a discovery is the often quoted exoplanet, HD 209458, that fades by

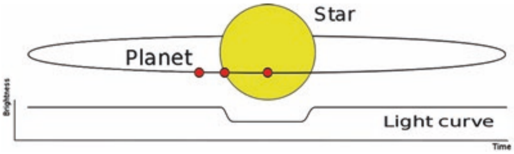


Fig. 15.1. Transit method for detecting exoplanets. (Illustration courtesy of Nikola Smolenski)

about 2% every 3.5 days, and thus can be considered a variable star. Although its detection was first made by the radial velocity method, it was confirmed by the transit method, and was in fact the first-ever exoplanet found by transiting. As a bonus for amateur astronomers, HD 209458, an eighth-magnitude star in Pegasus, is visible with binoculars. The variable star designation for HD 209458 is V376 Pegasi and is the prototype for the variable class EP in the *General Catalogue of Variable Stars*, defined as stars showing eclipses by their planets.

Orbital Brightness Modulation

This method is very easy to explain, as we can compare it, sort of, to the Earth-Moon system. It is just the change in brightness of the system as a planet orbits around the star. Just think about the Moon orbiting Earth; as it does so it will undergo phases, from new to full and back to new again, and thus the amount of light reflected will change, altering the total brightness of the system. It's not possible to see an individual planet, of course, but the combined light can be measured over time and will be seen to be periodic. Once again, though, it does require the system to be aligned in a way that the variations can be measured. If the orbits are face-on, then there will not be any periodic change in brightness. The first planets discovered by this method were Kepler-70b and Kepler-70c.

Gravitational Microlensing

This method makes use of a rather subtle effect of Einstein's General Theory of Relativity. It can be explained like this. The gravitational field of the host star can act like a lens, magnifying the light of a distant background star. If the foreground-lensing star has a planet, then that planet's own gravitational field can make a detectable contribution to the lensing effect (see Fig.15.2). However, this technique does rely on what is, admittedly, a highly unlikely alignment; thus, it is necessary to continuously monitor a very large number of distant stars to detect any planet.

You can see that to make full use of this method requires a lot of background stars, and we are lucky in that there is one place where background stars are plentiful—the galactic center. Because the effect will only occur if

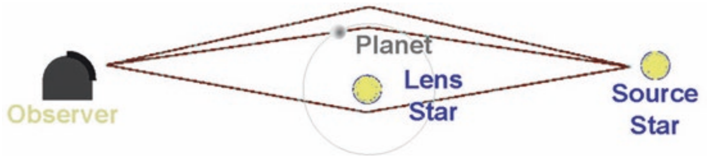


Fig. 15.2. Gravitational microlensing

the two stars are almost perfectly aligned, lensing events are short-lived, lasting for only a few weeks or days. Nevertheless, more than a thousand such events have been observed over the past 10 years. An example of success using this technique was the discovery of a low-mass planet in a very wide orbit, OGLE-2005-BLG-390Lb.

Direct Imaging

As the name suggest, this method of exoplanet detection involves, more or less, taking an image of the system in a way such that the planet can be resolved, separate from the host star. But it is not as easy as it sounds. The exoplanet will be extremely faint compared to the star and is often lost in its glare, so in reality what is measured is usually the thermal emission, or heat, from the exoplanet. But this is not always the case. In several cases multi-wavelength imaging is used so that one can distinguish between, say, a brown dwarf and a planet.

At the present moment, the method of direct imaging can be divided into two categories—exoplanets that are found around stars more massive than the Sun and are young enough to have protoplanetary disks, and those exoplanets that may be sub-brown dwarfs around dim stars, or brown dwarfs, both of which are at least 100 au away from their host stars. The above methods have met with considerable success—for example, the super-Jupiter exoplanet orbiting Kappa Andromedae and the exoplanets GQ Lupi b and AB Pictoris b.

Polarimetry

This technique uses the phenomena of the polarization of light in order to detect exoplanets. Simply put, the light from a star is generally not polarized, but if light is reflected from an atmosphere of a planet, the mechanism of reflection polarizes that light. By analyzing the combined light from the system, any polarization present can be measured, even though it may be only about 1 part in a million. This is achievable, as the measurement of polarization is very accurate and sensitive. The downside of this technique, which should be obvious, is that only those planets with atmospheres can be detected. What one needs is a large planet, with a substantial atmosphere that will reflect, and thus polarize, a lot of light. However, success in this area has been slow; as of summer 2014, only one planet had been detected by this method, and that was the confirmation of previously detected planet HD 189733 b. Nevertheless, it has great potential.

Astrometry

Our final method makes use of a technique that is actually very old. Basically, one measures, as accurately as possible, a star's position and noting any changes of that position over time. Historically, this was done by visual observation, and then later using photographic plates. The latter technique has the added bonus of an historical record that can be referred to. Now, if a star pos-

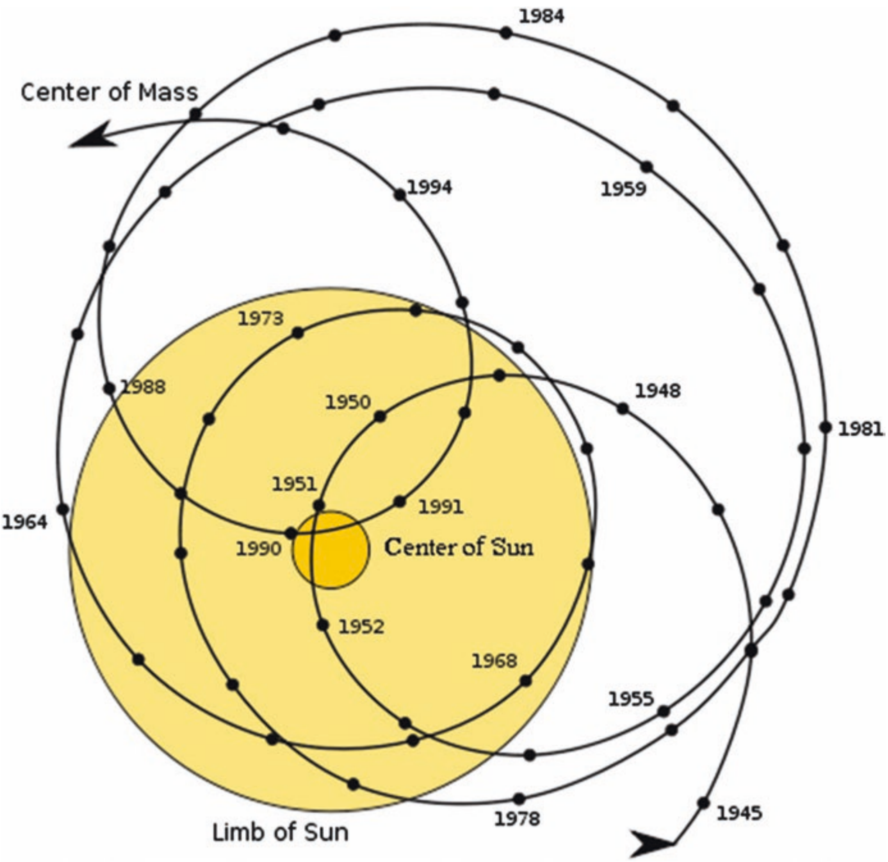


Fig. 15.3. Motion of the center of mass (barycenter) of the Solar System relative to the Sun

sesses an exoplanet, then this planet's gravitational effect will in fact cause the star to move in a very small circular, or sometimes elliptical, orbit. What is really occurring is that the star and exoplanet move about their common center of gravity—*the barycenter*. (See Fig. 15.3 for an example of this effect.

In most, if not all, cases the star will be far more massive than the exoplanet, and so its orbit will be much smaller. Usually, the common center of mass will lie within the radius of the star. Accordingly, it will be easier to find exoplanets that are around low-mass stars, especially brown dwarfs.

Even though astrometry is the oldest search method for extrasolar planets, originally popular because of its success in characterizing astrometric binary star systems, it must be stated that as yet, that only one exoplanet has been detected using this method. The Hubble Space Telescope did use the technique to confirm the existence of a previously known exoplanet, Gliese

876, and there is the detection of an exoplanet in the spectroscopic binary system HD 176051, called appropriately enough, HD 176051 b. The problem here lies in the fact that we don't know which stellar component the planet is orbiting around. However, all this will change profoundly with the space-based observatory GAIA that has the potential to discover literally thousands of exoplanets using the technique of astrometry. We eagerly await the news.

As of the time of writing, the percentage of confirmed planet detected using various techniques goes like this.

- Transit 75.6%
- Radial Velocity 19.5%
- Microlensing 2.4%
- Imaging 1.2%
- Transit Timing Variations 0.47%
- Eclipse Timing Variations 0.36%
- Pulsar Timing, 0.14%
- Orbital Brightness Modulation 0.14%
- Pulsation Timing Variations 0.05%
- Disk Kinematics 0.02%
- Astrometry 0.02%

The above listing of the various methods currently in use is by no means complete, as there are more techniques that have yet to be tested and validated, including transit imaging, magnetospheric radio emissions, auroral radio emissions and modified interferometry.

15.5. Observing Exoplanet Systems

Since the first extra-solar planet was discovered in 1989 literally thousands more have been discovered, and so to list them all would be impossible. But, more importantly, nothing is gained by viewing the innumerable star systems through a telescope, so we have just listed those stars that are visible to the naked eye and are known to harbor planets.[8] Nevertheless, it is a humbling experience to look at a star and to know that orbiting it are other planets, other worlds and possibly, other things as well!

A website is given in the appendix of this book that list up-to-date information on all the planetary systems discovered.

---

[8] For those of you that really want to see the exoplanet host stars through a telescope, there are many lists to be found on the web that will curb your craving.

The usual nomenclature is given in the details below, with the addition of the number of suspected planets in the system.

Pollux	β Geminorum	07^h 45.3^m	+28° 02′	January
1.14 m	1.09 M	K0 IIIvar	1 planet	33.78

Not a lot of people realize that this star hosts a planet, though for more than 20 years it was suspected that an extrasolar planet orbited Pollux. Its existence was confirmed in 2006. The planet, Pollux b, is estimated to have a mass at least 2.3 times that of Jupiter, and orbits Pollux with a period of about 590 days.

55 Cancri	HD 75732	08h 52.6m	+28° 20′	February
5.96 m	5.47 M	G8V	5 planets	41 l. y.

Also known as ρ[1] Cancri. This is one of the so-called “51 Peg” planets. The star is surrounded by a dust disc extending at least a 40 AU, with an inclination ~25°. Furthermore, there may be a hole with a radius of ~10 AU in the disc. The star has a companion, ρ[2] Cancri, about 1150 au away.

47 Ursae Majoris	HD 95128	10h 59.4m	+40° 25′	March
5.03 m	4.29 M	G0V	3 planets	46 l. y.

These planets are some of the few that appear to fit all the current models and theories about planetary formation—and may be the only one!

70 Virginis	HD 117179	13h 28.3m	+13° 46′	April
5.0 m	3.68 M	G2.5Va	1 planet	59 l. y.

This planet is a gas ginat with a mass of 7.5 Jupiters. The planet also has a very eccentric orbit, with an eccentricity of 0.4. A value of 0 would be a perfect circle, while a value of 1.0 is a long, flattened oval. Mercury and Pluto have the largest eccentricities in our Solar System, with values of around 0.2.

τ Boötis	HD 120136	13h 47.2m	+17° 27′	April
4.50 m	3.53 M	F6IV	1 planet	51 l. y.

This is another of the 51 Peg planets. It is the only system (so far) that has had a probable detection of the starlight reflected by the planet. This, the albedo, is claimed to be detected only in the wavelength range from 456 to 524 nm. The star has a companion, GJ 527B, about 240 AU away. There is, as yet unconfirmed, another planet but at a far distance from the star.

α Centuari	Rigel Kent	14h 39.6m	+60° 50′	April
1.33 m	4.38 M	K1 V	1 planet	4.37 l. y.

The closest system to the Solar System, this is a binary system, although to the naked eye will appear as a single star. The extrasolar planet α Centauri

B orbits the K-type star α Centauri Bb, and although it still needs to be verified it would be the closest planet to us as well as the lowest minimum mass planet detected so far, around a star that is similar to the Sun. Using a statistical approach to the system, there should be several other planets present, although these, as well as the original discovery, have yet to be verified, and the data is now in doubt.

ρ Coronae Borealis	HD 143761	16h 01.1m	+33° 18′	May
5.41 m	4.18 M	G0-2Va	1 planet	57 l. y.

Recent observations using infrared techniques have led astronomers to believe that there is a circumstellar disc of gas and dust around the star. From the disc inclination (46°) a planet with a mass of 1.5 that of Jupiter's can be inferred; however, this value differs from other results. The orbital period and amplitude imply a mass of around 1.1 Jupiter masses, and a semi-major axis (which is half the distance across the long axis of an ellipse and is usually referred to as the *average* distance of an orbiting object) of around .23 AU, or roughly half the distance between the Sun and Mercury. In situ formation of such a planet is thought to be unlikely. A more plausible scenario is that the planet formed at several AU from the parent star by means of gas accretion onto a rocky core and then migrated inward. This could have happened by interactions with another giant gas planet that was ejected in the process, through interactions with the protoplanetary gas disc or by interactions with planetesimals—the building blocks of planets, formed by accretion in the solar nebula.

14 Herculis	HD 145675	16h 10.4m	+43° 49′	May
6.67 m	5.32 M	K0V	2 planets	59 l. y.

14 Herculis (Gliese 614) is a star somewhat less massive than the Sun (80%), and 14 Herculis b has a slightly elongated orbit of 4.8 years. Its mass is about 4.64 times that of Jupiter, and it is at a distance of 2.77 AU from 14 Her. This giant planet is twice as close to 14 Her as Jupiter is to our Sun. The content in heavy chemical elements of 14 Her is rather large compared with that of the Sun, a discovery that reinforces the suggestion that giant planets are more frequently observed around metal-rich stars. Heavy chemical elements are needed to form dust or ice particles, and then, by agglomeration, planetesimals and the cores of giant planets. If the quantity of dust is large enough, this is certainly a factor in favor of the formation of giant planets. 14 Herculis c was detected in 2005, with an orbital period of 19 years and confirmed in 2021.

Gliese 777		20h 03.6m	+29° 54′	July
5.71 m	4.7 M	G5V	2 planets	51.7 l. y.

The star here is a yellow subgiant, in the constellation Cygnus. It is in actuality a binary system, with the possibility of a third component. The first planet was discovered in 2002, a long-period planet called Gliese 777b, and is classified as an eccentric Jupiter-type planet whilst the second, Gliese 777c was discovered in 2005, with a mass similar to Neptune. These are Jovian-type planets but with very eccentric orbits. They are also believed to affect the evolution of a solar system in that eccentric Jupiters can prevent a planetary system from having Earth-like planets because its eccentric orbit would remove all Earth mass planets from the habitable zone.

51 Pegasi	HD 217014	22H 57.4M	+20° 46′	September
5.49 m	4.51 M	G5V	1 planet	50 l. y.

This planet, 51 Pegasi b, was the first planetary-mass object discovered. It lies very close to its host star, with temperatures believed to be around 1200 C. It gave rise to the class of extrasolar planet called "hot Jupiters." However, the discovery gave rise to some problems. Hot Jupiters are a peculiar type of system characterized by orbital periods shorter than 15 days. The orbits are small, with radii less than 0.11 AU, which is about a tenth the distance between Earth and the Sun. Such an orbit is in fact much smaller than that of Mercury's (radius 0.38 AU, period 88 days). However, these planets are similar in mass to that of Jupiter and in some cases even larger. Research indicates that the planets have circular orbits. The 51 Pegasi-type planets were, and in some cases still are, a problem because they do not fit current planet formation theory. This predicts that giant planets like those in our Solar System (Jupiter, Saturn, Neptune and Uranus) should be formed in the colder, more distant parts of a protoplanetary disc, some 5 AU from a star. It seems that a possible solution to this problem may be what has been termed as *planetary migration*. This occurs when a planet or other stellar satellite interacts with a disk of gas or planetesimal, resulting in the alteration of the satellite's orbital parameters, especially its semi-major axis, with the inevitable effect that planets that were originally formed further out in a system move closer to the host star. In fact, this process is believed to have led to the observed structure of our own Solar System.

Formalhaut	α Pisces Austrini	22^h 57.6^m	−29° 37′	September
1.16 m	1.72 M	A3 V	1 planet	25.13 l. y.

This system is important in extrasolar planetary research for two reasons. The star is surrounded by a dusty torus of material, often referred to as Formalhaut's own "Kuiper Belt" and is believed to be a protoplanetary disc. In addition, a planet was located just within the dust ring and was the first ever to be seen in visible light, when the Hubble Space Telescope imaged it on November 2008. Its mass is estimated to be close to three times that of

Jupiter, with a lower limit of a mass similar to that of Neptune's. As yet, no other planets have been detected, but due to the structure of the dust disc, it is believed other planets may exist in the system, as yet undiscovered.

υ Andromedae	HD 982	01h 36.8m	+41° 24′	October
4.1 m	3.44 M	F8V	4 planets	44 l. y.

This was the first multiple-planet system discovered around a main sequence star. The first planet discovered, υ Andromedae b, is so close to the host star that it affects its chromospheric activity. Further analysis of the data then gave rise to another three planets, υ Andromedae c, υ Andromedae d, and υ Andromedae e. All 4 have very eccentric orbits, even more so than Pluto in our Solar System, and υ Andromedae d resides in the habitable zone. Research suggested that the eccentricity of the system could be due to an encounter between the outermost planet and an as yet undiscovered planets too distant and faint to be currently detected.

Tabby's Star	----------	20h 6.3m	+44° 24′	October
11.7 m	3.08 M	F3V	PLOONET	1470 l. y.

Located in Cygnus, roughly halfway between the bright stars Deneb and Delta Cygni and south of 31 Cygni, and northeast of the star cluster NGC 6866, is a star with a unique feature. It may host a ploonet, as suggested by the irregular dimming in brightness. Although the variations are slight, there have been dips of up to 5%, and is being constantly monitored, especially by AAVSO. Have a look at this one, especially at star parties, with a large telescope, and announce to the visitors…that star may have a moon!

Thought Question 15.4

Do you think amateur astronomers can participate in exoplanet research?

Problems

1. An exomoon is discovered orbiting an exoplanet at an average distance of 0.1 AU. Calculate the average time it would take to complete one orbit, in days.
2. An exocomet is orbiting a star with a perihelion distance of 60 AU, and an aphelion distance of 2 AU. Determine its period.
3. Astronomers find an exoasteroid orbiting a star using the transit method. They determine that is average distance from the star is 50 AU. Will the same astronomers see the same exoasteroid transit the star again in their lifetime?
4. Explain your answer.

Chapter 16

Galaxies

16.1. Introduction

We now discuss objects that every amateur astronomer has usually seen at least a handful of—*galaxies*.[1]

However, for the majority of amateurs, galaxies tend to remain faint and elusive objects, and perhaps only 15–20 galaxies are ever observed by 99% of amateur astronomers. It often comes as a surprise to know that with the proper optical system and under optimum seeing conditions (and a copy of this book!), there are in fact many more, in fact several hundred, that are within reach of even the smallest telescopes or binoculars. In fact a few are even visible to the naked eye if you know where to look!

Galaxies are vast, immense collections of stars, gas and dust. Indeed they are the source of all stars, because stars are not born outside of galaxies.[2] The number of stars in galaxies varies considerably, for instance in some giant galaxies, there may be over a trillion (10^{12}) stars—a number that staggers the mind. On the other hand, in small dwarf galaxies, such as Leo I, there may be only a few hundred thousand.

---

[1] The Milky Way Galaxy is often referred to as the "Galaxy," with a capital letter; any other galaxy is simply a "galaxy."

[2] Although a few stars may, after an immense amount of time, break free of a galaxy's grip and become intergalactic wanderers.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_16

16.2. Galaxy Types

Galaxies come in a variety of shapes and sizes, but the vast majority can be grouped into a few distinct classifications. When astronomers first began studying the galaxies, the most obvious characteristic that immediately became apparent was their shape, or morphology. Broadly speaking, they can be classified into three major categories:

Spiral galaxies appear as flat white discs with yellowish bulges at their centers. The disc regions are occupied by dust and cool gas, interspersed with hotter ionized gas, as is the case in the Milky Way. Their beautiful spiral arms are their most obvious characteristic.

Elliptical galaxies are somewhat redder, more rounded in appearance, like a football.[3] Compared with spiral galaxies, ellipticals contain far smaller amounts of cool gas and dust but larger amounts of hot ionized gas.

Those galaxies that appear neither disc-like nor rounded are classified as *irregular galaxies*.

Some spiral galaxies exhibit a straight bar of stars that cuts across the center, with spiral arms curling away from the ends of the bars. Galaxies with these features are known as *barred spiral galaxies*. Those galaxies that possess discs but not spiral arms are called *lenticular galaxies*, because they look lens-shaped when seen edge-on.

The classification system is further subdivided and specialized to take account of, for instance, the brightness of the nuclear region (the tight compact central region of the galaxy), the tightness of the spiral arms, etc.

Thought Question 16.1

Why do you think it was only at the beginning of the twentieth century that astronomers realized that galaxies were star systems beyond the Milky Way.

16.3. Galaxy Structure

At this point we will describe in a little more detail the structure of a galaxy as this will, at least in a small way, provide some insight into why galaxies appear the way they do. The books mentioned in the appendix of this book

---

[3] Or, at times, a rugby ball.

will have a much more detailed coverage of this topic, along with discussions on the origin and formation of galaxies.

Spiral galaxies have a thin *disc* extending outward from a central *bulge*. The bulge merges smoothly into what is called the *halo*, which can extend to a radius in excess of 100,000 light years. Both the bulge and halo make up what is called the *spheroidal component*. There are no clear boundaries between the constituent parts of this component, but a ballpark figure often used is that stars within 10,000 light years of the center can be considered to be bulge stars, whereas those outside this radius are members of the halo.

The disc component of a spiral galaxy cuts through both the halo and the bulge, and can, in a large spiral galaxy such as the Milky Way, extend 50,000 light years from the center. The disc area of all spirals contains a mixture of gas and dust, called the interstellar medium, but the amounts and proportions of the gas, whether atomic, ionized or molecular, will be different from galaxy to galaxy.

16.4. Stellar Populations

The stars contained within a spiral galaxy can also be classified depending on where they reside. Those that lie in the disc region are called *Population I* stars and are often young, hot and blue stars. Those in the bulge region are old red giants, called *Population II* stars. This is why photographs often show the spiral arms colored blue, owing to the Population I stars, with the bulge colored orange because of the Population II stars. The spiral arms may also be dotted with pink and red HII regions,[4] areas of star formation. Thus, new stars are usually formed in the spiral arms of galaxies, seldom in the bulge.

About 75% of large galaxies in the observable universe are apparently spiral or lenticular. Some spiral galaxies that can be found in a loose collection of other spirals. This is known as a *group*—spread over several million light years. Our own galaxy, the Milky Way, is a member of the *Local Group*.

Elliptical galaxies differ significantly from spirals in that they do not have a significant disc component. Therefore an elliptical has only the spheroidal component. The interstellar medium is also different in ellipticals; it is a mixture of low-density, hot, X-ray-emitting gas. Contrary to what you may have read in some books, ellipticals do possess a little gas and dust, and some have a small gaseous disc at their center that is believed to be the remains of spiral galaxies that the elliptical has consumed.

---

[4] See Chap. 2 for a description of HII regions.

The stars in the spheroidal population of elliptical galaxies give a clue to possible star formation, if there has been any. Such stars are orange and red, with an absence of blue stars, indicating that they are old, and that star formation occurred a long time ago.

Ellipticals are often found in large clusters of galaxies, usually located near their center. They make up about 15% of the large galaxies found outside clusters, but about 50% of the large galaxies within a cluster. Very small galaxies, called *dwarf elliptical galaxies*, are often found accompanying large spiral ones. A perfect example of such an arrangement, and one that is visible to the amateur, is the Great Andromeda Galaxy, M31, which is a classic spiral galaxy, and its attendants, M32 and M110, both dwarf ellipticals.

Several galaxies can be observed to not belong to either the spiral or the elliptical galaxy category. These are the irregular galaxies, and in fact more or less include all those galaxies that do not easily fall into the two previous classes. They include small galaxies such as the Magellanic Clouds[5] and those galaxies that are peculiar owing to tidal interactions. These systems of galaxies are usually white and dusty, as spirals are, though there the resemblance ends. Deep imaging has shown that the more distant galaxies are irregular, which indicates that this type of galaxy was more common when the universe was much younger.

16.5. Hubble Classification of Galaxies

The famous American astronomer Edwin Hubble was the first person to put the many disparate types of galaxies into some sort of order. He based his system solely on how the galaxy looked. The *Hubble classification*, as it is now known, is used as a means of categorizing a galaxy and the classic visualization of it is often referred to as the *Hubble Tuning Fork diagram* (see Fig. 16.1).

The classification system can appear cumbersome, especially when one delves deeper into its details, so we will outline them carefully here. Basically the classification is as follows.

An upper-case letter followed by either a number or a lower-case letter is assigned to the galaxy in question, and this identifies its morphology.

In the case of an ELLIPTICAL GALAXY, the letter *E* is used followed by a number. The larger the number, the flatter the galaxy. An E0 galaxy is

---

[5] Recent research has shown that the Large Magellanic Cloud, often classified as irregular, is in fact a spiral galaxy, even though it bears little resemblance to the classic spiral shape.

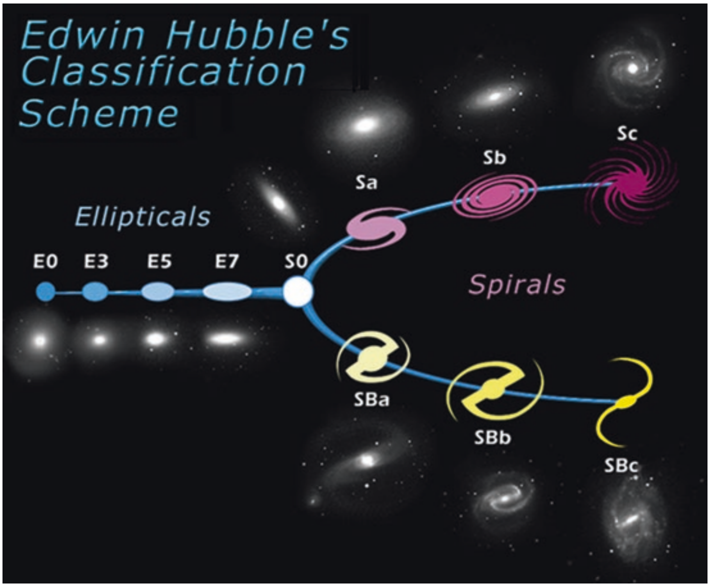


Fig. 16.1. Hubble's Tuning Fork diagram showing the main galaxy types. (Illustration courtesy of NASA and HST Heritage Project)

round, whereas an *E7* galaxy is very elongated. There exists a subgroup for the ellipticals where *D* signifies a diffuse halo, c is a supergiant galaxy, and d represents a dwarf galaxy. Thus, some of the largest elliptical galaxies have a *cD* classification.

A SPIRAL GALAXY is assigned the letter S, but can also be assigned *SA*, to signify that it is an ordinary spiral, or *SB*, where the *B* indicates it has a bar. It is then followed by a lower-case letter: *a*, *b*, *c* or *d*. Intermediate classes also exist for spirals, for example *ab*, *bc*, *cd*, *dm* and *m*. The lower-case letters a to d indicate the size of the bulge region, the dustiness of the disc, and the tightness of the spiral arms, while *m* denotes a stage where the spiral shape is barely discernible.

An Sa galaxy will usually have a large bulge, with a modest amount of dust, and tightly wound arms, whereas an Sd galaxy will have a small bulge and very loosely wound arms. An SBc galaxy will have a bar and also a small bulge.

There is also a classification for galaxies intermediate between spirals and ellipticals, the LENTICULAR GALAXIES, classified as *SO*. *SAO* is for those that are ordinary. *SBO* is for those that are barred. In addition, for galaxies intermediate between type *S* and *SB*, there is the classification *SAB*.[6]

---

[6] We will discuss this addition to Hubble's classification shortly.

An outer ring can surround both lenticular and spiral galaxies or perhaps the spiral arms will nearly close upon themselves, thus forming a pseudo-ring. These new features are classified as *R* and *R′*, respectively.

Finally, there are classifications for those galaxies that do not easily fall into any of the above three! These include:

- *Pec*, for peculiar galaxies, which may have a distorted form.
- galaxies that have an irregular morphology, and are classed as *Irr*. These can also be further classified as unstructured, *IA*, and barred, *IB*.
- dwarf galaxies, classified as *d*.

The difference between a galaxy classified as Pec and one classified as Irr can be very small, but a peculiar galaxy is one that is thought to have suffered considerable tidal distortion from the passage of another galaxy nearby.

The Hubble classification system can be represented by a simple diagram (Fig. 16.1). Note, however, that the diagram, and the classification scheme in general, do not represent an evolutionary sequence. Galaxies on this diagram do not start as ellipticals and then progress to be spirals, though there is some evidence that this type of evolution does take place.[7]

You will by now have noticed that elliptical galaxies have a classification that is associated with a numerical value. This is easily determined by a simple set of measurements (see Math Box 16.1).

For a spherical galaxy with a equal to b, the resulting number is 0, and the Hubble type is E0. Studies indicate that the limit is about E7, with the most common shape about E3. One can see that this shape classification depends both on the intrinsic shape of the galaxy, as well as the angle with which the galaxy is observed. Therefore some galaxies that are classified Hubble type E0 are in reality elongated, but are seen from one end, and not along its length, so to speak!

Math Box 16.1: Classification of an Elliptical Galaxy

The formula for determining the Hubble classification of an elliptical galaxy is:

$$n = 10 \times \frac{(a-b)}{a}$$

This can also be written as $n = 10 \times \left(1 - \frac{b}{a}\right)$.

Note that a and b are the major (a) and the minor (b) axes of the galaxy.

---

[7] This process is believed to occur when galaxies collide and meld into one another, forming a new, entirely different galaxy with a new morphology, possibly a giant elliptical.

If the major and minor axis of an elliptical galaxy is measured to be 260 and 140 units, respectively, what is that galaxy's Hubble classification?

We proceed as follows:

$$a = 260$$
$$\mathrm{b} = 140$$
$$n = 10\frac{(260 - 140)}{240}$$
$$n = 10\frac{120}{240}$$
$$n = 10 \times 0.5$$
$$n = 5$$

Thus, the classification of this particular elliptical galaxy is E5. Note that if both a and b are equal, the classification is E0.

16.6. Gérard de Vaucouleurs's Classification of Galaxies

The Hubble Tuning Fork classification system is mainly used these days by amateur astronomers and some professional astronomers, but it does have its limitations, as noted in the previous section. In fact, a study a new study using data from the longstanding Galaxy Zoo project and published in the journal *Monthly Notices of the Royal Astronomical Society*, used classifications of over 6000 galaxies and revealed that the accepted correlations between different features were not found in this sizeable and comprehensive sample.

Hubble's system classifies galaxies by type and shape and takes account two main features: the size of the central region (known as the 'bulge'), and how tightly wound any spiral arms are. His supposition was that galaxies with larger bulges tended to have more tightly wound spiral arms, lending vital support to the 'density wave' model of spiral arm formation.

But now, the new study finds no significant correlation between the sizes of the galaxy bulges and how tightly wound the spirals are. This suggests that most spirals are not static density waves after all.

What all this means is that perhaps more helpful methods to classify galaxies are needed. A system based upon Hubble's is the de Vaucouleurs system first described by Gérard de Vaucouleurs in 1959 and is an extension to the Hubble system. This classification system retains Hubble's basic

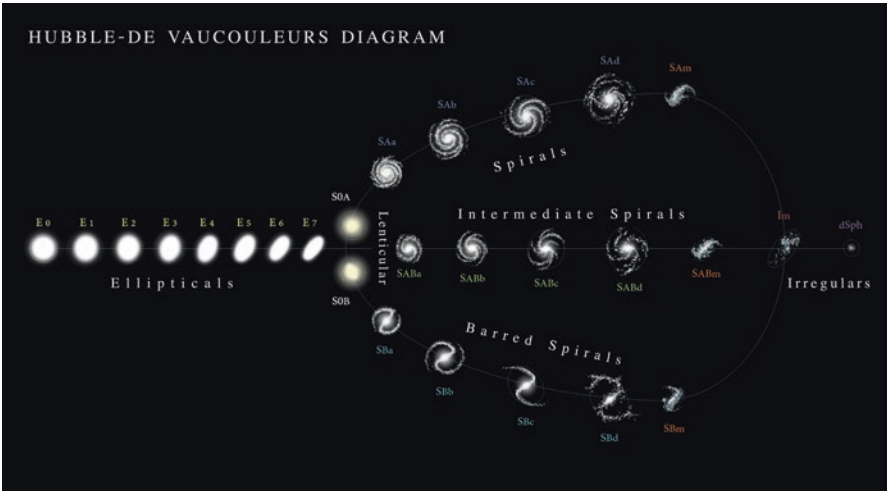


Fig. 16.2. Hubble—de Vaucouleurs diagram for galaxy morphology featuring ellipticals, lenticulars, spirals, intermediate spirals, barred spirals and irregulars. (Courtesy of Antonio Ciccolella and M. De Leo)

separation of galaxies into ellipticals, lenticulars, spirals, and irregulars, but expands it for spiral galaxies, centred on three morphological characteristics, see Fig. 16.2. These are.

- *Bars*. Galaxies are divided on the basis of the presence or absence of a nuclear bar, with the notation SA to denote spiral galaxies without bars, that matches Hubble's use of SB for barred spirals. It also includes an intermediate class, denoted SAB, containing weakly barred spirals. Lenticular galaxies are classified as unbarred (SA0) or barred (SB0), with the notation S0 now reserved for galaxies that are edge-on to the line-of-sight.
- *Rings*. Galaxies are divided into those possessing ring-like structures denoted (r) and those without rings, (s). Galaxies that seem to be a 'transition' are given the symbol (rs).
- *Spiral arms*. This is the same as Hubble's original scheme, with spiral galaxies being assigned a class based primarily on the tightness of their spiral arms.

The new scheme also extends the arms of Hubble's tuning fork to include several additional spiral classes:

Sd (SBd)—diffuse, broken arms made up of individual stellar clusters and nebulae, with very faint central bulge.

Sm (SBm)—irregular in appearance; but has no bulge component.
Im—very irregular galaxy.

Galaxies in the above three classes were mostly classified as Irr I in Hubble's original scheme. Furthermore, the Sd class contains some galaxies from Hubble's Sc class. In addition, galaxies in the classes Sm and Im are designated as "Magellanic" spirals and irregulars, respectively, named after the famous Milky Way neighbours, the Magellanic Clouds. So, with the de Vaucouleurs system, the Large Magellanic Cloud would be classified as type SBm, whilst the Small Magellanic Cloud is an irregular (Im).

By using all this new information one can adapt the many different elements of the modified classification schemes to give the complete classification of a galaxy (usually written in the order in which they are listed), i.e., a barred spiral galaxy with loosely wound arms and a ring is denoted *SAB(r)c*.

One way to think of this new system is to treat the de Vaucouleurs' system as a three-dimensional version of Hubble's two-dimensional tuning fork, with the stage (how spiral it is) the x-axis, family (how much of a bar) the y-axis, and variety (amount of ring) the z-axis.

This is by no means the only way to classify galaxies, as there are many more, but we will only briefly mention a couple of the more commonly used ones.

- *The Numerical Hubble Stage*. Numerical values are assigned to each class of galaxy in the De Vaucouleurs scheme. Values run from −6 to +10, with negative numbers corresponding to early-type galaxies (ellipticals and lenticulars) and positive numbers to late types (spirals and irregulars). Elliptical galaxies are now divided into three 'stages': compact ellipticals (cE), normal ellipticals (E) and late types (E^+). Lenticulars are similarly subdivided into early (S^-), intermediate (S^0) and late (S^+) types. Irregular galaxies can be of the Magellanic type irregulars, assigned a value of 10 or 'compact', value 11.
- *The Yerkes Scheme*. The Yerkes or Morgan scheme was devised by the American astronomer William Wilson Morgan, who, working with Philip Keenan, developed the MK system for the classification of stars through their spectra. Thus the Yerkes classification scheme uses the spectra of stars in the galaxy along with its shape, real and apparent, and the degree of the central concentration, to classify the galaxy.

Observationally, the Hubble classification system can be confusing (an understatement!), and as outlined above the system we have presented is by no means complete, even with the addition of further subdivisions to all the classes. But don't let that worry you. The complete system is only of relevance to those astrophysicists who study galaxies; to the observer the important point is whether the galaxy is a spiral, and, if so, is it barred, or whether the galaxy is an elliptical. In those few galaxies where the spiral or elliptical structure is very apparent, the subdivisions of, say, E1, E2 and Sa, Sb, SBa, etc., will be useful. Like most things in observational astronomy, it will all become easier with use.

There are some times when the classification system may be of no apparent use at all, namely when a galaxy is inclined to our line of sight. For instance, a galaxy such as M83, which is a nice spiral galaxy, is face-on to us, and thus the classic spiral shape is very apparent. However, a galaxy such as NGC 891, which is classified as a spiral, will appear as a thin streak of light, because the galaxy is edge-on to us, and the spiral shape will not be visible. However, finding galaxies that present an edge-on or nearly edge-on perspective adds another element to the pleasure of locating and observing these faint objects.

Thought Question 16.2

When we see an elliptical galaxy from its longer side, it could be classified as an E6, using Hubble's classification system, like this 

However, if we were to see more-or-less face on to its edge, it may look like this 

What is its Hubble classification now?

16.7. The Milky Way

Before we move onto actually observing galaxies, a small detour is necessary to discuss our home galaxy, the Milky Way, and especially, the center of the Galaxy, which is an observational joy.

The Milky Way, our home galaxy, as you are no doubt aware of, is a barred spiral galaxy, similar to many that have been mentioned earlier. It is about 1200 light years thick and 100,000–120,000 light years in diameter, containing over several hundred billion stars, and classified as a SAB(rs) bc-type galaxy. The Solar System is located around halfway from the center of the galaxy to the edge of the disc, about 27,000 light years, and around 70 light years above the plane of the galaxy. It is near the inner edge of a short spiral arm called the Orion Arm, which is around 15,000 light years long and contains the Cygnus Rift and the Orion Nebula (M42). The most recent estimate for the mass of the Milky Way is about 4.5 trillion solar masses, however, this includes the not only the mass from stars and dust but also the contribution from the strange material we call dark matter.

The center of the Milky Way can pose several problems to the astronomers, both amateur and professional. The reason for this? The vast amount of interstellar gas and dust that lies between us has made it nearly impossible to see it in visible light. However all is not lost, as gamma rays, hard X-rays, infrared radiation, and radio waves are able to pass through the interstellar medium, and so a picture can be built up of this inaccessible region.

What has been learned over the past 50 years is impressive. Located very near or even at the exact center is a strong radio source called Sagittarius A*[8] (pronounced "Sagittarius A star"), and this was the first cosmic radio source discovered. Measurements of the radio source indicate that it is no bigger than the diameter of Mars' orbit, and one of the (many) surprising results is that Sagittarius A* is stationary, which would indicate that it is very massive.[9] Combining all this information leads to the conclusion that at the center of our galaxy is a supermassive black hole, with a mass of about 4.1–4.5 million solar masses and a diameter of around 44 million km.[10]

When you look through a telescope at the region, the field of view will be full of numerous star fields, but these actually lie much closer to us than to the center. However when observing this region you may want to contemplate the almost certain fact that even though you cannot see the center, there in your eyepiece, ever invisible, lies a supermassive black hole, around which you, the Solar System, and the galaxy revolve.

Thought Question 16.3

What is the main reason that we are unable to see the actual centre of the Milky Way using either amateur or professional telescopes?

16.8. Observing Galaxies

To the amateur astronomer, observing galaxies can present something of a dilemma. In astronomy magazines and books, you are bombarded with images of galaxies, their spiral arms resplendent and multicolored, speckled

---

[8] Sagittarius A* is now believed to be made of two components—SgrA East and SgrA West. The former is a supernova remnant, and the latter is an ultra-compact, non-thermal source, i.e., a black hole.

[9] Recent analysis suggests that the density around the center of the Milky Way is about a million times greater than any known star cluster. It is probably made up of living stars, dead stars, gas and dust, and of course a black hole.

[10] To get a sense of scale, consider that the Mercury is 46 million km from the Sun at perihelion.

throughout with distinct pink HII regions. However, when you look at that same galaxy through your telescope, all you can see is a pale tiny blob!

It is true to say that in nearly every case, especially from an urban location, the galaxy you are looking at will be faint and indistinct. Only with the largest telescopes and the darkest possible skies can any real structure be seen. But take heart! Even with the naked eye, you will be astonished at what you can actually see with practice and from the right location.

This author recalls that during a visit to the wilder parts of Turkey, on several occasions under utterly dark skies, I was able to see M31 in Andromeda and M33 in Triangulum in such amazing detail that even today the memory takes my breath away. Using just my naked eyes, I was able to trace M31 to nearly 2½° across the sky, and M33 was a huge amorphous glow. Also, had I only known enough to look, I would have been able to see several other galaxies with the naked eye as well, but I was under the common misapprehension that the naked-eye limit is about sixth magnitude, whereas I now know that, with extremely dark skies and light-adapted vision, magnitude 8 is more like the limit.

The purpose of this anecdote is to remind you that, in order to see faint galaxies and the detail therein, dark skies are indispensable. With such dark skies, and armed only with a pair of binoculars, many galaxies will be within reach. If you have a telescope that number increases dramatically.

As usual, dark skies, dark-adapted vision and averted vision will all help in tracking down and seeing galaxies. Clean optics will also greatly aid you in your observations. Dust and smears of grease will reduce by a surprising amount the light that reaches your eyes, and in particular will reduce the contrast.

Generally, those galaxies that have a brightness greater than thirteenth magnitude are usually visible in telescopes of aperture 15 cm, and those of aperture 30 cm will see down to about 14.5 magnitude. There will of course be galaxies that will have much brighter magnitudes than these, and so will be visible in much smaller instruments. In some cases only the brightest part of a galaxy will be visible—perhaps its core (nuclear region), with the spiral part unobservable.

To be able to trace out the finer details of the spiral arms of galaxies, and to locate the bulge area, faint halo and HII regions, you will invariably require a large-aperture telescope. But if the purpose of your observing is just to locate these elusive objects, and to be amazed that the light that is entering your eye may have begun its journey over 100 million years ago, then there are a plethora of galaxies awaiting your visit.[11]

---

[11] Of course, I don't really have to mention that if you have a medium-to-large-aperture telescope, then the number of galaxies visible to you is vast, and the detail you will be able to see will astound you!

Some images of galaxies can be found at the end of this chapter.

The usual nomenclature applies in the following descriptions, but with these changes. Galaxies are extended objects, which means that they cover an appreciable part of the sky—in some cases a few degrees, in others only a few arc minutes. The light from the galaxy is therefore "spread out," and thus the quoted magnitude will be the magnitude of the galaxy as if it were the "size" of a star; this magnitude is often termed the *integrated magnitude*. This can cause confusion, as a galaxy with, say, a magnitude of 8 will appear fainter than an eighth-magnitude star, and in some cases, where possible, the surface brightness of a galaxy will be given. This will give a better idea of what the overall magnitude of the galaxy will be. For instance, Messier 64, the Black Eye Galaxy, has a magnitude of 8.5, whereas its surface brightness is 12.4. The surface brightness will be given in brackets after the quoted magnitude, and to use the above example of M64 will appear like this: 8.5 m [12.4 m].

Following on from the previous paragraph, the designation "easy," "moderate" or "difficult" takes into account not only the brightness of the galaxy but also the area of the sky the galaxy spans. Thus, a galaxy may be bright with—say—a magnitude of 8, which under normal circumstances would be visible in binoculars and designated as "easy." But if it covers a significant amount of the sky (and thus its surface brightness is low, making it more difficult to observe) it is designated as "moderate."

In addition, spiral galaxies can exhibit a variety of views, depending on their inclination to the Solar System. Some will appear face-on, others at a slight angle and a few completely edge-on. As an indicator of inclination, the following symbols will be used.[12]

Face-on: 
Slight inclination: 
Edge-on: 

Finally, the Hubble classification of galaxies as described earlier will also be used.

Here, now, are some spiral galaxies, without bars, you can observe.

Caldwell 7	NGC 2403	07^{H} 36.9^{M}	+65° 35′	January
8.5 m [13.9 m]	21.9′ × 12.3′		SAB(s)cd	Easy

This is one of the brightest galaxies that was somehow missed from the Messier catalog and is often left out of an observer's schedule. In binoculars it appears as a large, oval hazy patch with a brighter central region. With

---

[12] Be aware, however, that with small apertures, knowing the inclination won't really matter much, the galaxy will still look like a dim, diffuse blob.

averted vision, and an aperture of about 20 cm, faint hints of a spiral arm will become apparent. Larger apertures will of course present even further detail. It is not a member of the Local Group of galaxies[13] but believed to be a member of the M81/M82 group. It was the first galaxy outside the Local Group found to have Cepheid[14] variable stars discovered within it, and the current estimate of its distance is 11.5 million light years from us.

Messier 81	NGC 3031	09^{H} 55.6^{M}	+69° 04′	February
7.3 m [13.6 m]	26′ ׀ 14′		SA(s)ab	Easy

A spectacular object! In binoculars it will show a distinct oval form, and using high-power binoculars the nuclear region will easily stand out from the spiral arms. Using a telescope will reveal considerably more detail and show it to be one of the grandest spiral galaxies on view. With an aperture of about 15 cm, traces of several of the spiral arms will be glimpsed. A real challenge, however, is to see if you can locate this galaxy with the naked eye. Several observers have reported seeing it from dark locations. If you do glimpse it without any optical aid, then you are probably looking at one of the furthest objects[15] that can be seen with the naked eye, lying at a distance of some 4.5 million light years. M81 is partner galaxy to M82 (see the chapter on Active Galaxies), and both of these spectacular objects can be glimpsed in the same field of view.

Messier 96	NGC 3368	10^{H} 46.8^{M}	+11° 49′	March
9.3 m [13.1 m]	7.6′ ׀ 5.2′		SAB(rs)ab	Easy

This faint galaxy can be seen in binoculars as a faint hazy oval patch of light. But what you are observing is in fact just the bright central nucleus of the galaxy, as the spiral arms are too faint to be resolved. Telescopes will bring out further detail, and with good conditions the spiral arm features will be seen. There is some slight controversy over M96, as recent measurements of its distance place it at 38 million light years, which is 60% greater than the previous value. It forms a nice triangle with two other galaxies, M95 and M105, and as such is a member of the Leo I group of galaxies.

Messier 58	NGC 4579	12^{H} 37.7^{M}	+11° 49′	April
10.1 m [13.5 m]	5.9′ ׀ 4.7′		SAB(rs)b	Easy

---

[13] The Local Group is a cluster of several galaxies, including the Milky Way. It consists of M31, M33, M110 and M32, the Large and Small Magellanic Clouds and about 25 other dwarf galaxies, including Leo I and II, And I and II, the Draco, Carina, Sextans and Phoenix dwarfs.

[14] Cepheid variables are used as "standard candles," for measuring distances to other extra-galactic objects, and where discussed in Chap. 10.

[15] The galaxy M83 lies at the same distance and has reportedly been seen with the naked eye.

Through binoculars this galaxy will appear as a faint, hazy patch of light with a barely discernible nucleus. You may also glimpse in the same field of view the galaxies M59 and M60. A telescope of about 10-cm aperture will show some structure in the halo, along with faint patches of light and dark. There are some reports that a 20-cm telescope will allow the bar connecting the spiral arms to the nucleus to be resolved. This galaxy has about the same mass as the Milky Way and is about 95,000 light years in diameter.

Caldwell 12	NGC 6946	20^{H} 34.9^{M}	+60° 09′	July
9.6 m [14.1 m]	11.5′ǀ 9.8′		SAB(rs)cd	Moderate

A challenging galaxy to locate using binoculars, where it will appear, if at all, as a small round hazy patch of light with a barely perceptible increase at its center. What makes this galaxy difficult to locate and observe, even though it is close to us, is that it lies in the part of the sky near the plane of the Milky Way. This results in the light from the galaxy being dimmed by the intervening dust and stars. With telescopes of 20 cm and a dark location, the faint outer halo can be glimpsed. In order to observe further detail such as spiral arms, a very dark location along with large aperture is needed. Research indicates that there is a hectic period of star formation occurring in its inner nuclear region. Such an outburst is termed a *starburst*.

Caldwell 30	NGC 7331	22^{H} 37.1^{M}	+34° 25′	August
9.7 m [13.5 m]	10.5′ǀ 3.7′		SA(s)bc	Easy

This is the brightest galaxy in the constellation Pegasus, and with binoculars it will appear a faint patch of light that has a brighter core. Easily visible in a telescope of 20 cm, which will show its structure in a little more detail. Apparently this galaxy is similar to M31 but lies much further from us, at a distance of 40 million light years. There is also some debate as to whether it is linked with the famous Stephen's Quintet (see the section on groups and clusters of galaxies). Finally, in most spiral galaxies the central bulge typically co-rotates with the disk, but the bulge in Caldwell 30 is rotating in the opposite direction to the rest of the disk. As yet, no one knows how or why!

Caldwell 43	NGC 7814	00^{H} 03.2^{M}	+16° 08′	September
11.0 m [13.6 m]	5.5′ǀ 2.3′		SA(s)ab:sp	Moderate

This is no binocular object, but nevertheless a splendid sight, especially with larger aperture telescopes. It is a fine example of an edge-on galaxy, and bears many similarities to its better-known cousin, M104. Easily seen in a telescope of aperture 20 cm, it does however provoke debate among amateurs as to whether its dust lane can be seen with small telescopes. Some profess to have seen it with 20 cm, while others claim that at least

40 cm is needed. Try observing with as high a power as it can take, as this may help you to resolve this dilemma.

Messier 31	NGC 224	00^{H} 42.7^{M}	+41° 16′	October
3.6 m [13.6 m]	3°l 1°		SA(s)b	Easy

Also known as the Andromeda Galaxy. The most famous galaxy in the sky is probably also the most often visited one, and is always a first observing object for the beginner. It is visible to the naked eye, even on those nights when conditions are far from perfect. Many naked-eye observers claim to have seen the galaxy spread over at least 2½° of sky, but this depends on the transparency. In binoculars it presents a splendid view, with the galactic halo easily seen along with the bright nucleus. Large binoculars may even show one or two dust lanes. Using averted vision and with a very dark sky, several amateurs report that the galaxy can be traced to about 3° of sky in telescopes of aperture 10 cm. In larger telescopes a wealth of detail becomes visible. With an aperture of about 20 cm, a star-like nucleus is apparent, cocooned within several elliptical haloes. Another striking feature are the dust lanes, especially the lane running along its northwestern edge. Many observers are often disappointed with what they see when observing M31, as the photographs seen in books actually belie what is seen at the eyepiece. M31 is so big that any telescope cannot really encompass all there is to show. Patience when observing this wonderful galaxy will reward you with a lot of surprises. Spend several nights observing, and try to choose a dark night in a country location. This really is a spectacular galaxy. It contains about one trillion stars with a diameter of 130,000 light years and is among the largest galaxies known. It is the largest member of the Local Group. In older texts it is often referred to as the Great Nebula in Andromeda.

Here, now, are some barred spiral galaxies to observe.

Messier 95	NGC 3351	10^{H} 44.0^{M}	+11° 42′	March
10.0 m [13.8 m]	7.4′l 5.0′		SB(r)b	Moderate

This is a faint galaxy that shows little if any detail in binoculars, as it will just appear as a hazy patch; however it will be in the same field of view as M96. With a telescope of at least 15 cm some structure can be glimpsed, with larger apertures showing the distinctive bar feature. It is a member of the Leo group of galaxies. There is some debate as to the real magnitude of the galaxy, with some observers putting it at 9.2 m.

Caldwell 32	NGC 4631	12h 42.1^{M}	+32° 32′	April
9.1 m [13.0 m]	15.5′l 2.7′		SB(s)dsp	Moderate

An often neglected galaxy that is surprising, as it has a lot to offer. Visible in binoculars as a faint elongated object, it really needs a telescope to be appreciated. It is a very big galaxy, which, owing to its appearance, has led to it being unofficially nicknamed the Whale Galaxy. Its eastern end is appreciably thicker than its western, hence the name. This aspect can be seen with an aperture of 20 cm, and larger telescopes will show further details such as patches of light and dark, along with two prominent knots. On the northern side of the galaxy is a faint twelfth-magnitude star that, providing the seeing is good, will act as a pointer to a faint companion galaxy. Several theories have arisen as to the origin of its strange and disturbed appearance. The most probable is tidal interactions with several nearby galaxies.

Caldwell 72	NGC 55	00^{H} 15.1^{M}	−39° 12′	September
7.9 m [13.5 m]	32.4′ × 5.6′		SB(s)m:sp	Easy

Although this galaxy lies so far south as to make it invisible from the northern hemisphere, it still warrants inclusion. In binoculars it appears a faint spindle-shaped object, and large binoculars hint at some delicate structure. Telescopes show even more detail, and it is one of the few galaxies where an H-alpha filter will highlight its HII regions.

–	NGC 1365	03^{H} 33.6^{M}	−36° 08′	November
9.5 m [13.7 m]	9.8′ × 5.5′		(R)SB(s)b	Moderate

A very impressive galaxy, sometimes known as the Great Barred Spiral Galaxy, it is easily visible in binoculars as an elongated hazy object with a brighter center. In a telescope with an aperture as small as 8 cm, its galaxy origin is obvious, and larger apertures will show considerably more detail. Although not visible from the UK, it should be a nice observing target from the United States. Recent studies suggest that the Milky Way may in fact look the same as NGC 1365 looks, to an external observer.

Here, now, are some elliptical galaxies to observe.

Messier 49	NGC 4472	12^{H} 29.8^{M}	+08° 00′	March
8.4 m [12.9 m]	10.2′ × 8.3′		E2	Easy

This is the second-brightest galaxy in Virgo and easily spotted in binoculars as a featureless, oval patch of light. Although most ellipticals are rather featureless, M49 stands up quite well with higher power and large aperture, when some resolution can be seen in the nuclear area. It seems to have a bright nucleus surrounded by a diffuse core region, which in turn is surrounded by a rather diffuse halo. Some observers report that the nucleus shows a mottled appearance under magnification. The galaxy is at the center of a sub-cluster of galaxies called the Virgo Cloud, which in turn is part of the much larger Virgo Cluster. In addition it appears the elliptical galaxy is cocooned in an envelope of hot gas at a temperature of about 10,000,000 K. At such a high temperature, X-rays are formed, and it was with an X-ray telescope that this feature was detected.

Caldwell 35	NGC 4889	13^{H} 00.1^{M}	+27° 58′	April
11.5 m [13.3 m]	2.8′∣2.0′		E4	Difficult

This is well worth seeking out, as it is a very distant galaxy, about 350 million light years away. Excellent seeing conditions are needed to glimpse this tiny object, with a telescope of at least a minimum 20-cm aperture. It has a bright core, surrounded by the usual faint halo. It is a dominant member of the Coma Galaxy Cluster, which contains about 1000 galaxies (several of which can be seen in large-aperture telescopes of at least 40 cm). The cluster itself is made of many elliptical galaxies and S0-type galaxies. Apparently it is the result of a merger of two older clusters. Observing any of these galaxies is a feat indeed, but well worth the effort.

Caldwell 17	NGC 147	00^{H} 33.2^{M}	+48° 31′	September
9.6 m [14.6 m]	13.2′∣7.8′		dE4	Difficult

Located in Cassiopeia, this is classified as a dwarf elliptical (spheroidal?) galaxy. Although some distance from M31, the Andromeda Galaxy, it is in fact a companion to it. It is difficult to locate and observe, however, so dark skies are a prerequisite. It has been said that a minimum of 20-cm aperture is needed to see this galaxy, but there have been recent reports that under excellent conditions a 10-cm telescope is sufficient, although averted vision was needed. The moral of this story is that dark skies are essential to see faint objects. Increased aperture will help as well as higher magnification, when its nuclear region then becomes visible. A member of the Local Group, it is one of over 30 galaxies that are believed to be companions to either M31 or the Milky Way.

Messier 110	NGC 205	00^{H} 40.4^{M}	+41° 41′	October
8.2 m [14.1 m]	21.9′∣11.0′		E5P	Easy

The final entry in the Messier catalog, and added to the original list in 1967. It is the second satellite galaxy of M31, and although it has a brighter magnitude than M32, the first satellite, it has a much lower surface brightness. Consequently it is much harder to see. It is visible in large binoculars, but will only appear as a very faint, featureless glow, northwest of M31. In a telescope it shows a surprising amount of detail, and high magnification will bring out its mottled nucleus. In addition, it shows detail that is peculiar for an elliptical galaxy; furthermore, these details are visible to the amateur. Of course, exceptionally dark skies and perfect seeing and transparency will be needed, but in a telescope of even modest aperture, say 10 cm, with a high magnification, they are readily seen. Look for dark patches near a bright center. Strangely enough, the galaxy's features are reminiscent of those normally found in a spiral galaxy. A definite observing challenge!

Here, now, are the lenticular galaxies.

Caldwell 53	NGC 3115	10^{H} 05.2^{M}	−07° 43′	February
8.6 m [11.6 m]	7.2′ǀ 2.5′		SOsp	Easy

This galaxy is often overlooked, which is a shame because it is a fine example of its type, as well as being quite bright. In binoculars it will appear as a small, faint elongated cloud, while in large binoculars it displays the characteristic lens shape. It is easily located in telescopes because of its high surface brightness. In telescopes of aperture 20 cm, it will appear as a featureless oval cloud, with perhaps a slight brightening toward its center. As it is classed as an SO-type galaxy, it will not show any further detail even with larger aperture. It is a very big galaxy, some five times larger than the Milky Way. It is also one of the most favored objects that is purported to have a black hole at its center. Also known as the Spindle Galaxy.

Finally, here are some peculiar and irregular galaxies to observe.[16]

Caldwell 60/61	NGC 4038/9	12^{H} 01.6^{M}	−18° 52′	March
10.5/10.7 m [13.3 m]	7.6′ǀ 4.9′		Sp S(B)p	Moderate

Also known as the Antennae or Ring-Tail Galaxies. Together, these probably make one of the most famous objects in the entire sky, but few amateurs ever observe it, believing it to be too faint. A telescopic object, it will appear as an asymmetrical blur in apertures of about 20 cm. Large apertures will begin to hint at its detailed structure, and at 25 cm aperture it will begin to resemble the famous apostrophe shape. Using apertures of about 30 cm, along with medium to high magnification, will show you that there are two objects involved, and it would be a worthwhile project to see just how much detail is resolved with a different group of telescopes and observers. It is one of those celestial objects that is so familiar from photographs that your perception of what is seen will be tainted by what you expect to see. Nevertheless, it is a wonderful object. Sadly, it is very low down for northern observers, so perfect observing conditions will be necessary. The marvellous shape of the Antennae is the result of spiral galaxies passing close by each other, so that tidal interaction causes material to be dispersed. Witness the amazing long tails that can be seen on deep images of these galaxies. Furthermore, recent work has shown that the interaction has led to a vast bout of star formation. Also, it has encouraged astronomers to put forward the idea that spiral galaxies evolve into elliptical galaxies after such an encounter.

---

[16] Many irregular galaxies are also Active Galaxies, a topic discussed in Chap. 15. Therefore, a list of observable irregular/active galaxies will also be found therein.

Messier 85	NGC 4382	$12^{H} 25.4^{M}$	+18° 11′	March
9.2 m [13.1 m]	7.1′ × 5.5′		SA(s)OP	Moderate

This is a bright galaxy that can be glimpsed in binoculars on clear nights. With large binoculars it is even easier to see, where it will show a star-like nucleus surrounded by the faint glow of the halo. Using a telescope will just magnify the rather featureless aspect, though a few observers report that at high magnification some faint detail can be glimpsed to the south of the nucleus which may be a trace of some spiral structure. Also, there is an indication of a faint blue tint to the galaxy.

Caldwell 21	NGC 4449	$12^{H} 28.2^{M}$	+44° 05′	March
9.8 m [13.2 m]	6.0′ × 4.5′		IBm	Moderate

A member of the *Canes Venaticorum Group* of galaxies, this is a faint and frequently ignored object. Its irregular shape is often mistaken for a comet. Under good skies, a telescope of 20 cm aperture will easily discern its fan-shaped morphology, along with its faint nucleus. Larger telescopes will of course resolve the galaxy with a considerable amount of detail. An interesting point is that several HII regions are visible, especially one at the northern corner of the open fan shape. It is apparently a site of much ongoing star formation, and is similar in many ways to the *Large Magellanic Cloud.*

Caldwell 57	NGC 6822	19h 44.9m	−14° 48′	July
9.9 m [15.5 m]	15.5′ × 13.5′		IB(s)m	Moderate

This is a challenge for binoculars astronomers. Even though it is fairly bright, it has a low surface brightness and so is difficult to locate. Once found, however, it will just appear as a hazy indistinct glow running east-west. This is in fact the bar of the galaxy. Strangely enough, it is one of those objects that seems easier to find using small aperture, say 10 cm, rather than large. Nevertheless, dark skies are essential to locate this galaxy. Also known as Barnard's Galaxy.

Caldwell 51	IC 1613	$01^{H} 04.8^{M}$	+02° 07′	October
9.2 m [–]	11.0′ × 9.0′		dIA	Moderate

A very difficult galaxy to observe; a few reports state that it is visible in large binoculars as a very faint hazy glow, while others claim that a minimum of 20 cm aperture is needed. Whatever you choose, one thing is paramount, namely a dark sky. A member of the Local Group, it is similar in many respects to Caldwell 57. It is an old galaxy that is still forming stars.

Caldwell 24	NGC 1275	$03^{H} 19.8^{M}$	+41° 30′	November
12.0 m [13.2 m]	2.2′ × 1.7′		P	Difficult

The final galaxy in our list is a very important galaxy, even though visually it is not impressive. It nevertheless should be observed for several reasons. It is the main member of the Perseus Galaxy Cluster, also known as Abell 426. Several amateurs have stated that it can be seen in telescopes as small as 15-cm aperture, while larger apertures will make it easier to locate. If you can have access to a telescope of aperture 40 cm or larger then use the opportunity to look at this galaxy, as it will be surrounded by several fainter ones that are all part of the cluster. In many respects it is the most concentrated field of galaxies in the winter sky for the northern observer. Caldwell 24 is a strong radio galaxy and is believed to be the remnant of a merger between two older galaxies. It is one of over 500 members of the Perseus Cluster. The cluster itself forms part of an even larger super cluster, the Pisces-Perseus Supercluster.

Thought Question 16.4

What makes observing distant galaxies difficult for an amateur astronomer?

16.9. Clusters of Galaxies

Surprisingly, single galaxies are a rare breed. Most galaxies live in clusters that may contain just a few, to giant clusters that may have thousands of members. In addition small clusters occupy only a relatively small region of space, say, 1 Mpc, while the largest cover an immense 10 Mpc. The Milky Way is a member of a small cluster called the Local Group, with over 80[17] other galaxies of various types.

One can consider two types of clusters, rich clusters and poor clusters. The former may consist of over 1000 galaxies, lots of ellipticals and cover an area of over 3 Mpc in diameter. In this type of cluster, the galaxies are more often than not concentrated towards the cluster center. At the center itself, there may be one or two giant elliptical galaxies. A fine example of such is the Coma-Virgo cluster,[18] with the giant elliptical M87 at its center. This cluster has about 300 large galaxies and perhaps as many as 2000 smaller ones. It is the closest large cluster, lying at a distance of around

---

[17] Additional members of the Local Group are being found at regular intervals. These are small and indistinct, thus their difficulty in being observed!

[18] The Virgo supercluster is believed to be part of the even bigger Laniakea Supercluster, with around 100,000 galaxies.

55 million light years. It spans over 100 square degrees in both Virgo and Coma Berenices. Such is its influence that the Milky Way is actually gravitationally attracted to it.

Thought Question 16.5

Do you think that is possible for the Local Group to have a giant elliptical as a member?

Poor clusters, as the name suggests, contain less than 1000 members but can cover areas as big as that of a rich cluster, thus making the galaxies more spread out.

It seems that the rich clusters contain about 80–90% E-type and S0-type galaxies, with a few spirals, whereas poor clusters have a larger proportion of spiral galaxies. Furthermore, those galaxies that are in isolation, i.e., not in clusters, it appears that 80–90% of these are spirals. There is a large amount of evidence that suggests that large elliptical galaxies have been involved in many galaxy collisions, whereas spirals have not. In fact, it may well be that ellipticals are formed by the merger of spirals. The dwarf ellipticals, on the other hand, seem to follow a different evolutionary path. These are small galaxies that have lost their gas and dust due to interaction with several larger galaxies.

Although the evolution of galaxies is still not fully understood it is obvious that interactions between them is very important. Collisions, mergers and close encounters can all cause a burst of rapid star formation and very dramatic and spectacular tidal disruption. In fact, our own Milky Way is cannibalizing the Magellanic Clouds right now! Amazing!

Now let’s take a look at some groups and clusters of galaxies.

Hickson Group 68	NGC 5353	13^{H} 53.4^{M}	+40° 47′	April
11.1 m	→11.2′←	5[19]		Moderate

This is a very nice group of galaxies for amateur instruments. The brightest member can be seen in a telescope as small as 6 cm, and with a 15-cm aperture it will show a slight brightening of its center. The other galaxies will appear as faint patches of light, and to see the faintest member would most certainly require an aperture of about 25 cm.

Copelands Septet	NGC 3753	11^{H} 37.9^{M}	+21° 59′	March
13.4 m	→7.0′←	7		Difficult

---

[19] # Indicates number of galaxies in cluster/group as seen through a telescope.

This is a very small group of galaxies. Situated in the constellation Leo, all within about 7 arc seconds. Telescopes of at least 25-cm aperture will be needed, and even then you may not spot the fainter members of the group but just the four brighter galaxies. Larger apertures should of course allow you to spot them. Nevertheless, seeing conditions will determine what you observe, regardless of aperture. The group is a mix of barred spirals, ordinary spirals and lenticular galaxies. Also known as the Hickson Galaxy Group 57.

Stephen's Quintet	NGC 7320	$22^{H}\ 36.1^{M}$	+33° 57′	April
12.6 m	→4′←	5		Difficult

A very famous group of galaxies located in Pegasus, but one that has in the past proved strangely difficult for amateurs. Under perfect seeing conditions, the group is visible in a 20-cm telescope. However, the word *perfect* is not an exaggeration here. The largest member of the quintet is only 2.2–1.2 arc seconds in size, so it is very small, but it is the brightest. To actually see the group as a distinct unit and not a faint smudge of light will require a telescope of aperture 25 cm. This will show at least four of the group, but the fifth requires an aperture of at least 30 cm. Under high magnification and large aperture, structure can be seen within the brighter members. It is believed that four of the group are interacting with each other (also known as the Hickson Compact Group 92), and there is debate as to whether the fifth is in fact a line-of-sight galaxy. This is a challenge to the urban astronomer.

Seyfert's Sextet	NGC 6027	$15^{H}\ 59.2^{M}$	+20° 46′	May
13.3 m	→1.5′←	6		Very difficult

This is a real challenge! In all but the largest telescopes it is questionable if you will see anything at all, and even in apertures of around 40 cm the galaxies will barely be resolved. Nevertheless, it would be interesting to find out what would be the smallest aperture required to spot these faint galaxies.

Fornax Cluster	NGC 1316	$03^{H}\ 20.9^{M}$	−37° 17′	November
11.4 m	→12 + ′←	10+		Difficult

This is another large cluster of galaxies. Amateur telescopes should be able to pick out the brightest members with no difficulty. What makes this cluster so spectacular, however, is that with a modest aperture, say 25 cm, and clear dark skies, there are so many galaxies visible that identification is very difficult. The brightest member is visible even in an 8-cm telescope! A galaxy of note in the cluster is NGC 1365, which is a nice barred spiral 8

Photo 16.1. Black Eye Galaxy, Messier 64

arc seconds in length and visible in an 8-cm aperture telescope as a faint blur. The cluster also contains a galaxy known as the Fornax System, a very faint and small type of galaxy known as dwarf spheroidal galaxies (Photos 16.1 and 16.2).

Problems

1. The major and minor axis of an elliptical galaxy are 500 and 150 units, respectively, what is galaxy's Hubble classification?
2. The major and minor axis of an elliptical galaxy are 50 and 45 units, respectively, what is galaxy's Hubble classification?
3. What is the approximate shape of the galaxy in question 2.

Photo 16.2. Andromeda galaxy, Messier 31 (inner regions)

Chapter 17

Active Galaxies

17.1. Active Galactic Nuclei (AGN's)

On class of object that seems increasingly important for galaxy evolution is known as Active Galaxies.

The story begins in the 1950s, when radio astronomers began to detect galaxies that were emitting vast amounts of radio energy, in some cases as much as 10 million times more radio energy than a normal galaxy would. Later on, when space-borne telescopes made an appearance, it was found that there exist galaxies that emit incredible amounts of energy in the infrared, ultraviolet and X-ray. These then are the active galaxies. Subsequent observations have shown that most of this "extra" energy is emitted from the central regions of the galaxies, and that is why they are called active galactic *nuclei.*

The different types of active galaxies are legion! In fact, there was a time early on when it seemed as if every time an active galaxy was discovered it could be put into its own individual class. But we now know different. For instance there are[1]:

- Seyfert Galaxies, Type 1 and 2 [and 1.6, 1.7. 1.8 & 1.9!]
- LINERS [(low-ionization nuclear emission line regions]
- LLAGN [low-luminosity AGN]

---

[1] And this is not a complete list either!

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_17

- Radio-loud AGN
- Radio-quiet AGN
- BLAZERS, consisting of BL Lacs and OVV's [optically violent variables]
- Flat Spectrum Radio Quasars (FSRQ's)
- Steep Spectrum Radio Quasars (SSRQ's)
- Starbursts
- QSO's and quasars, these two being the most extreme type of AGN's.

This is a lot!

So, what is really going on?

17.2. Origin of Nuclear Activity

The answer, if correct, is quite simple. At the center of the galaxy is a supermassive black hole surrounded by an accretion disc. This disc is very hot close to the black hole, but cooler further out. Theoretical studies suggest that the inner part of the disc can be very narrow, and that the black hole is hidden from us deep within this narrow central area. The outer part of the disc is believed to be a large, dense torus-shaped object consisting of dusty gas. Material falls into the black hole via the accretion disc, and prestigious amounts of energy is emitted. Sometimes this energy is focused into jets that we can, as in Messier 87. Sometimes the emitted energy causes nearby clouds of fast-moving hydrogen to emit very brightly in the hydrogen alpha wavelength; other circumstances give rise to clouds that emit light further out from the center.

Broadly speaking, what type of active galaxy [or AGN] we see depends on how the accretion disc is inclined to our line of sight. If we can see the inner regions, we may have a Seyfert 1 active galaxy. If the inner region is obscured from view, we then may see a Seyfert Type 2. It is important to realize that even though we may see a galaxy face-on, it doesn't necessarily mean we will see the active nucleus face-on. The accretion disc may be inclined at a very steep angle to the plane of the galaxy, as in Centaurus A. Figure 17.1 attempts to show this.

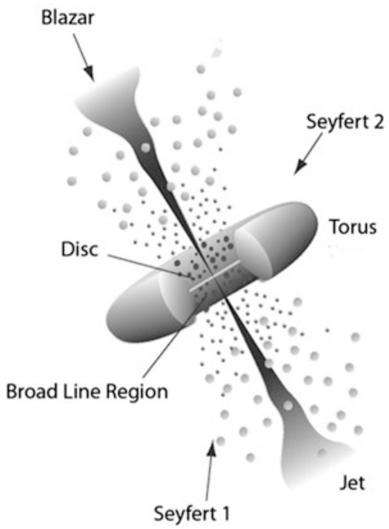


Fig. 17.1. Diagrammatical form of the unified theory of active galactic nuclei

17.3. Classification of Active Galaxies

We now have a better way to determine what type of active galaxy we are observing, and fully characterize it. Active galactic nuclei can now be classified by three parameters:

- Optical variability.
- Radio emission.
- Spectral line width.

With this in mind one can describe the galaxies in this way:

- SEYFERT galaxies are mostly radio-quiet and not so strongly variable as other types of AGN.
- SEYFERT-1's have broad and narrow spectral lines. SEYFERT-2's have only narrow lines.
- Some SEYFERT-1's exhibit broadened lines that are relatively narrow and are designated *Narrow Line Seyfert 1's* (NLS1's).
- QUASARS are broad-lined and some of them are variable. They can be further divided between *radio-quiet quasars* and *radio-loud quasars*.
- RADIO GALAXIES have strong radio emission and are not variable.
- BLAZERS are highly variable, and some of them have strong radio emission and can be divided (roughly) into two classes, *narrow-lined BL Lac*

objects (some of which show no lines at spectra at all, and that's why the redshift of many BL Lac's remains unknown) and *broad-lined OVV's* (Optically Violently Variable quasars).

As mentioned earlier, the source for all the energy in active galaxies is believed to be supermassive black holes that lurk at their centers. Further observations suggest that most active galaxies are interacting or merging with nearby smaller galaxies, and these events provide a source of material that can feed the central supermassive black hole. This leads nicely to the idea that those galaxies that are not interacting with a companion, or haven't done so in the recent past, will not have material flowing into the black hole, and thus will not be active. And indeed, this is what we see.

A certain type of AGN that has profound consequences for both galaxy evolution and cosmology are the quasars, also known as quasi-stellar objects, or QSO's. Thus, it is important that we briefly mention them here. The story begins in 1963 when Maarten Schmidt at Hale Observatories managed to identify some previously unknown spectral lines in a supposedly stellar spectrum. He discovered that the unknown lines were in fact hydrogen lines, but with a large redshift. In the object he was looking at, 3C273, he measured a redshift of 15.8%. This isn't particularly large, as redshifts go, but soon after his discovery quasars with larger redshifts began to be discovered.

So what, you may ask? Well, the significance is that these quasars are very far from us, as evident from their redshifts; however, they appear bright on photographs, in fact, star-like, hence the quasi-stellar nomenclature. To be so easily seen,[2] and yet at immense distances from us, they must be very luminous, perhaps as much as 10–1000 times as luminous as a galaxy. Thus, quasars must be superluminous.

Another important factor now appears; some quasars fluctuate in brightness in only a few months, and as an object cannot alter its brightness appreciably in less time than it takes light to cross its diameter, then these quasars must be very small objects, perhaps not more than a few light-months in diameter. What can possibly make nearly 1000 times more energy than all the stars in a galaxy yet be so small? You guessed it—a supermassive black hole.

We now know that quasars are the nuclei of galaxies[3] that lie at tremendous distances from us, and thus must be objects that formed in the early

---

[2] By easily, I mean photographically.

[3] The quasar HE0450-2958 is very odd as it was the first quasar found to apparently lack a host galaxy, and often called the "naked quasar". Currently, the scientific consensus is that it probably does have a host galaxy, but it is difficult to see because of the bright quasar light. However, recent research questions the idea after a study by the European Southern Observatory.

universe. Deep images suggest that these young galaxies are distorted, and many have close companions suggesting that quasars contain supermassive black holes. The activity we observe is possibly initiated by interactions between the host and companion galaxies.[4]

Thought Question 17.1

What would be easier to observe with an amateur telescope, say 200 mm aperture, an active galaxy or a quasar?

17.4. AGN Variability

As we have discussed, active galaxies contain a small radiation-emitting core embedded in an otherwise typical galaxy. Several active galaxies have been found with a core that is highly variable, as well as very bright when compared to the rest of the galaxy. This variability can last months, weeks, days and even minutes, and it allows us to determine the size of the emitting region by using a very clever technique that involves knowing the speed of light and being able to set limits on how quickly an object can change its brightness.[5]

The best way to describe this concept is like this: Imagine an AGN that measures 4 light-weeks across (2 light-week in radius) as shown in Fig. 17.2. Now suppose that the entire AGN emits a brief flare (~1 s) of light coming from a volume of hot gas or plasma. Photons from the part of the AGN nearest to Earth arrive at the telescope first.

Photons from the middle of the AGN (the largest part if spherical) arrive on Earth, and telescopes, sometime later.

Finally, light from the far side of the AGN arrives after a measurable time difference from the arrival of the first photons.

Although the object emitted a sudden flash of light, what is observed is a gradual increase in brightness that lasts, in this case, a full 4 weeks from the first recorded incident. In other words, the flare is stretched out over a time interval equal to the difference in the light travel time between the nearest and most remote observable regions of the AGN.

---

[4] As an interesting aside, in 1965, Soviet astronomer Nikolai S. Kardashev declared that the quasar CTA-102 was sending coded messages from an alien civilization.

[5] It is important to stress that this is just a model and a rather over-simplification of what really happens. Nevertheless, it works reasonably well.

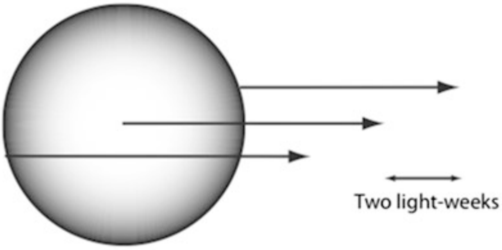


Fig. 17.2. Light time variability from an AGN. (Idealized case. It won't look like this in reality)

For example, if an AGN is 1 light year in diameter, it will take a year longer for the signal from the *far* side of the AGN to be detected than the signal from the *near* side.

Thus, we can calculate the diameter of the light-emitting region just by measuring the length of time it takes for the light received to vary in brightness. Once we have done that we simply multiply this measured length of time by the speed of light. An example of such a calculation is shown in Math Box 17.1.

The method outlined above will give a good estimate of the size of the emitting region, but it will not be exact. Research has shown that AGN's are not regular in the number of short-duration outbursts. In fact, there can be many shorter variations within longer, more intense variations. Normally we assume that the *shortest* duration variation is the one that tells us the size of the emitting region because the speed of light is the *maximum* speed the variation can travel across the object, but there could, and possibly do, exist several slower ways for the light to "move" across the region. With this in mind we can use the fastest light variation to estimate an upper limit to the size of the region. Remember however that the region that emitted the outburst could well be smaller than the calculated size, but it won't be bigger.

In addition, if you expand the depiction so that the emitting region is very large, then the variations in brightness will tend to be "smeared out," reducing the appearance of variability. On the other hand, if the emitting region is very small, the variability can be extremely short—maybe as small as a few hours.

Thought Question 17.2

Can you think of two differences between Quasars and Active Galaxies?

Math Box 17.1 Determination of Size of Active Galactic Nuclei

The light from an AGN varies in brightness over a 1-week period. Determine an estimate of its size, and compare this to the size of the Solar System, ~10^{13} m.

Size of emitting region (meters) = (speed of light) (meters/second) × (variation time) (seconds)

$$d = c \times \Delta t$$

First convert the 1-week into seconds, as the speed of light is given in meters per second.

$$7 \text{ days} \times 24 \text{ hours / day} \times 3600 \text{ second / hour} = 604{,}800 \text{ seconds}$$

Then,

$$d = c \times \Delta t$$

$$d = 3 \times 10^8 \text{ ms}^{-1} \times 6.05 \times 10^5 \text{ s}$$

$$d = 1.82 \times 10^{14} \text{ m}$$

The Solar System is approximately 10^{13} m in diameter, so the AGN flare is about 18 times the diameter of the Solar System.

Recall that a typical AG can be several million, or even a billion, times the mass of the Sun, and have about 1000 times the luminosity of the entire Milky Way Galaxy, making this distance, 1.82×10^{14} m, incredibly small in comparison.

17.5. Starburst Galaxies

There is one type of galaxy that leads many astronomers to debate whether it should be included under the heading of active galaxies at all, and this is the starburst galaxy. This is a galaxy that is currently (as we see it now) undergoing an unusually high rate of new star formation, when compared to the average rate of star formation observed in most other "normal" galaxies. Such a high rate of star formation will inevitably lead to the galaxy consuming all of its available gas that is usually used in forming stars, but on a timescale that is far shorter than the age of the galaxy. As a consequence, the starburst period of a galaxy is a relatively short phase in a galaxy's total evolution.

From a purely astrophysical point of view, starbursts have a spectrum that originates, for the most part, from new stars, or from material that has absorbed and re-radiated the light from new stars. However, the mechanism that gives rise to the spectra of an active galaxy originates from the very many (very!) transitions that occur near the nucleus of the galaxy, often referred to as *non-thermal radiation*.

Although there is no one definition that astronomers can agree on that defines a starburst galaxy, an approximate definition can be obtained if the galaxy possesses the following three characteristics:

1. Star formation rate, or SFR, simply put, the rate at which the galaxy is currently converting gas into stars.
2. The amount of gas that can be used to form stars.
3. Comparing the star formation timescale as it consumes the available gas with the age or rotation period of the galaxy.

Note that all the above parameters are interrelated.

Of course, a question that is often asked is "What causes this new star formation in the first place?" This is where a trigger is needed to initiate star formation, and it is one that is, surprisingly, very common.

Many starburst galaxies (although not all) appear to have tidal tails, streams of material arcing away from the galaxy that can be taken as an indication of a close encounter with another galaxy. Some starburst galaxies are themselves actively in the midst of a merger. Furthermore, some starburst galaxies that are not merging, but merely passing each other closely, may interact gravitationally, with the result that the rotation rates of the stars in the spiral arms and inter-arm gaps, along with any bar, can funnel gas inward towards the nucleus and provoke bursts of star formation close to the galactic nucleus.

It is difficult to classify a starburst galaxy as a certain type, as the period of star formation we are discussing can be very brief and thus does not really represent the galaxy but rather a phase of the galaxy's life. Nevertheless, one can roughly assign a type to the galaxy during its starburst phase. A few of the types that have currently been agreed upon by astronomers are:

- *Luminous infrared galaxies (LIRG's)*. These are galaxies that, as the name suggests, have an excess of radiation emitted in the infrared part of the spectrum, emitting more energy in the infrared than at all other wavelengths combined. For example, a LIRG's luminosity can be 100 billion times that of our Sun. These can be further subdivided into two groups. (1). First there are the *ultra-luminous infrared galaxies (ULIRG's)*. These are very dusty galaxies. The newly formed stars in them emit copious amounts of ultraviolet radiation that is absorbed and then reradiated in the infrared spectrum.[6] This obviously explains the observed extreme red color of ULIRG's. (2). The second group are the *hyperluminous infrared galaxies (HLIRG's)*. As the name suggests, these are galaxies with an infrared luminosity even higher than that of the ULIRG's.

- *Blue compact galaxies (BCG's)*. These are galaxies that have a low mass, a low metallicity and are dust-free objects. Observationally these are often blue in the optical part of the spectrum along with a strong ultraviolet component, exactly because they are dust-free and also they contain, naturally, a large number of hot young stars. There are two subclasses of BCG's. (1). The *blue compact dwarf galaxies (BCD's)* are, as the name suggests, small compact galaxies. (2). The *pea galaxies* are small compact galaxies that resemble primordial starbursts. What is particularly interesting about these galaxies is that were found by non-astronomers who took part in the Galaxy Zoo project.
- *Wolf-Rayet (WR) galaxies*. These are galaxies where a significant number of bright stars are Wolf-Rayet stars. Since the Wolf-Rayet phase is relatively shortlived, around 10% of the star's total lifetime, they are very luminous. This, along with the fact that a galaxy will only contain few of these, allows for the spectra of these stars to be identified in the spectra of the galaxies and thus determine the star formation, and thus starburst, characteristics of the galaxy.

Well-known starburst galaxies include M82 and NGC 4038/NGC 4039 (the Antennae Galaxies).

Thought Question 17.3

Would a starburst galaxy always remain classified as a starburst galaxy?

17.6. Observing Active Galaxies

It is a nice aspect of observational astrophysics that several active galaxies can be easily observed by amateur astronomers, as the list that follows will show.[7] The usual definitions used before for galaxies apply. Also, as is always the case, there are many more that can be observed, if one has a

---

[6] Not all astronomers are convinced that the UV radiation is produced by star formation alone but may be powered by active galactic nuclei (AGN). However, there is considerable X-ray evidence that ULIRGs are powered by star formation triggered by mergers.

[7] It is, of course, not a full list, as there are many AGN's visible in amateur telescopes. These just represent the brightest.

larger aperture telescope; the limiting factor applied here is to list only the brighter ones.

Caldwell 29	NGC 5005	13^{H} 10.9^{M}	+37° 03′	April
10.3 m [13.2 m]	6.3′| 3.0′		SAB(rs)bc	Easy

This galaxy is not a binocular object, and so telescopes of about 15-cm aperture will reveal it only as an oval patch with a bright nucleus. The galaxy doesn't have any conspicuous spiral arms so that even with large-aperture telescopes further detail will be sparse, and only some slight irregularity in overall brightness will be resolved. Although it is similar to the Milky Way, what makes this galaxy special is that it is an active galaxy of the LINER class. In the center of the galaxy is some sort of mechanism that gives rise to both the observed spectral lines and a radio source. It may be due to massive stars called *warmers*, or an accretion disc around a black hole.

Messier 77	NGC 1068	02^{H} 42.7^{M}	−00° 01′	November
9.1 m [13.0 m]	7.1′| 6.0′		(R)SA(rs)b	Easy

This is a famous galaxy for several reasons. In binoculars it is visible just as a hazy patch of light, and under excellent seeing a faint star-like nucleus may be glimpsed. In telescopes of about 10 cm and greater, and providing that dark skies are available, then the spiral arms can be glimpsed. But what makes this galaxy so special is that it is the archetypal active galaxy of the class known as Seyferts. Its uniqueness was discovered in the middle of the twentieth century by Carl Seyfert, who noticed that it had very prominent emission lines. These are due to the high velocity of gas close to the nucleus of the galaxy. The high speed of the gas, in the order of 350 kilometers per second, is believed to be due to the influence of a massive black hole. M77 is in fact classified as a Seyfert II galaxy, which indicates that it has only narrow emission lines. Seyferts are distant cousins of the famous quasars. It is one of the brightest active galaxies visible to the amateur astronomer.

Caldwell 67	NGC 1097	02^{H} 46.3^{M}	−30° 16′	November
9.4 m [13.76 m]	9.3′| 6.6′		SB(s)b	Easy

This is a nice galaxy, and its bar can easily be seen. With a 20-cm aperture telescope, the core is resolved and easy to see, as well as a faint elongated glow, which in fact is the bar. Larger apertures will resolve this feature quite well, along with the spiral arms which emanate from the bar's end. It is classified as a Seyfert galaxy of Type I. This means that gas close to the nucleus is moving at extremely fast speeds, maybe in excess of 1000 kilometers per second. The most likely cause of this motion is the influence of a supermassive black hole. A Seyfert I galaxy has both broad and narrow emission lines, the width of the line being a measure of the velocity of the gas that produced the emission line.

Messier 87	NGC 4486	12^{H} 30.8^{M}	+12° 24′	March
8.8 m [13.1 m]	8.3′| 6.6′		E0.5P	Easy

This is a very special galaxy, bright and easily seen in binoculars. Telescopes show little. But this rather bland appearance is deceiving. This is a monster of a galaxy, with a mass estimated to be that of 800 billion Suns, making it one of the most massive galaxies known in the entire universe. It is an active galaxy, and lurking at its core is a black hole with a mass of 3 trillion Suns. Another feature some observers report seeing with telescopes of aperture 50 cm is the famous “jet” that streams out from M87. It would be a challenge indeed, and a triumph, were this ever to be observed from the light-polluted skies of northern climes. The jet is a stream of plasma (hot ionized gas) several thousand light years in length that is believed to be due to some sort of interaction between the black hole and its surroundings. It is, however, relatively easy to photograph and image with a CCD camera. M87 lies at the heart of the Coma-Virgo Cluster and most of the surrounding galaxies are influenced by its tremendous gravitational attraction. The cluster has about 300 large galaxies and perhaps as many as 2000 smaller ones. It is the closest large cluster, lying at a distance of around 55 million light years. It spans over 100 square degrees in both Virgo and Coma Berenices. Such is its influence that the Milky Way is actually gravitationally attracted to it.

Messier 82	NGC 3034	09^{H} 55.8^{M}	+69° 41′	February
8.2 m [12.9 m]	11.0′| 4.6′		IOsp	Easy

A very strange galaxy, the strangeness of which becomes readily apparent when it is seen through a telescope. It can be glimpsed with binoculars, where it will appear as an elongated pale glow. Large binoculars will begin to hint at some detail, and with averted vision the dark dust lane may be seen. In even a small telescope of 10-cm aperture, it is evident that something strange has happened to M82. The western part is obviously brighter than the eastern, and the core region appears jagged and angular. Throughout the length of the galaxy, starlight appears to stream through the gaps in the dark dust lanes. It is a galaxy that will repay long and detailed study, especially at large aperture and high magnification. It is an active galaxy of the starburst type and is undergoing an immense amount of star formation. This may have been caused by the close passage of its companion M81. During that time, which was about 40 million years ago, the gravitational effect of M81 caused the interstellar material within M82 to collapse and form new stars. Subsequently, material that was dragged from M82 is now believed to be falling back onto it, which gives rise to both its appearance and the new era of star formation. Although classified as an irregular galaxy in 2005, more recent near-infrared observations detected spiral arms in M82. Both M81 and M82 can be seen in the same field of view and are a stunning sight.

Caldwell 77	NGC 5128	$13^{\mathrm{H}}\ 25.5^{\mathrm{M}}$	−43° 01′	April
6.8 m [12.9 m]	18.2′× 14.5′		SOpec	Easy

Although this galaxy is too far south for the UK and some US observers, it nevertheless warrants inclusion because it is so spectacular. Photographs show it as a nearly circular object bisected by a very prominent dark dust lane. Visible in binoculars as a hazy star, with larger binoculars the famous dark lane can just be glimpsed. In small telescopes, aperture 15 cm, the dark lane is easily seen. Larger aperture will of course give a more detailed view, with the dark lane showing some structure. The late writer and astronomer Iain Nicolson[8] said this about the galaxy, "Centaurus A (NGC 5128) is a magical object, one of those rare extragalactic objects that, when it swims into the field of view, looks just like the photographs that grace the pages of astronomy books. The first time I saw it, it seemed almost to fill the field of view. It's a beautiful object in its own right: a near-spherical elliptical galaxy with a pronounced dark lane right through the middle. It is especially intriguing because of its status as the nearest active galaxy. To know that this object has a compact core that probably houses a supermassive black hole makes Centaurus, for me, one of the most exciting objects in the sky." Its peculiar morphology is believed to be the result of a merger between two smaller galaxies. This famous galaxy is also known as Centaurus A.

Caldwell 24	NGC 1275	$03^{\mathrm{H}}\ 19.8^{\mathrm{M}}$	+41° 30′	November
12.0 m [13.2 m]	2.2′× 1.7′		P	Difficult

This is a very important galaxy, even though visually it is not impressive. It nevertheless should be observed for several reasons. It is the main member of the Perseus Galaxy Cluster, also known as Abell 426. Several amateurs have stated that it can be seen in telescopes as small as 15-cm aperture, while larger apertures will make it easier to locate. If you can have access

---

[8] Iain Nicolson, a wonderful friend of mine of over 50 years friendship, sadly, passed away in September 2020. He was a giant of British amateur astronomy, and sorely missed.

to a telescope of aperture 40 cm or larger use the opportunity to look at this galaxy, as it will be surrounded by several fainter ones that are all part of the cluster. In many respects it is the most concentrated field of galaxies in the winter sky for the northern observer. Caldwell 24 is a strong radio galaxy and is believed to be the remnant of a merger between two older galaxies. It is one of over 500 members of the Perseus Cluster. The cluster itself forms part of the even larger Pisces-Perseus Supercluster.

3C 273	12^{H} 29.1^{M}	02° 03′	March
12.9 m	Redshift (z) 0.158		2.6 billion l. y.

This quasar is the brightest in the sky and within reach of medium-to-large-aperture telescopes. There are even reports that it has been glimpsed in telescopes of 20 cm and thus is well within reach of most amateurs. Averted vision will also help locate this distant object. It is situated about 3.5° northeast of Eta Virginis and 3° southeast of the galaxy M61. First find the galaxy NGC 4536 (magnitude 10.6, surface brightness 13.2, at position R.A.12^{h} 34.5^{m} Dec. 02° 11′). At about 1.25° east of this galaxy is the quasar. In the immediate vicinity is a double star, arranged east-west with 3 arc seconds separation. The double has magnitudes 12.8 and 13, and the quasar is a bright, blue-tinted stellar object east of the double system.

PKS 405-123	MSH 04-12	04^{H} 07.8^{M}	−12° 11′	November
14.8 m		Redshift (z) 0.57		6 billion l. y.

This is another quasar that should be within the reach of amateur astronomers and has been glimpsed in telescopes of 20 cm. It is located in the constellation Eridanus. The quasar lies about 3° to the northeast of Zaurak ([Gamma] g Eri). When seen through an eyepiece, you may spot a tiny green dot to the left. This is the planetary nebula NGC 1535 (Cleopatra's Eye). If you do manage to see the quasar, and you will need detailed star maps to confirm the observation, then you will be a member of a very small and elite group of observers. It is also incredible to note that the light that enters your eye from this quasar started its journey some 1.5 billion years *before* the Solar System was formed!

Thought Question 17.4

How would you determine that the star-like object you observe is a star or a quasar?

Problems

1. The first Quasar identified was 3C273, with a redshift of 0.158. What is its recessional velocity in km?
2. If the Hubble constant is 67.8 $\mathrm{kms^{-1}Mpc^{-1}}$, determine the quasars distance. (Ignore the effect of the expansion of the universe).
3. The light from an AGN varies in brightness over a 3.5-day period. Determine an estimate of its size in km.
4. How many times larger is this than the Solar System, ~10^{13} m.

Chapter 18

Cosmology

To discuss cosmology in detail would entail a complete book, along with an appreciable background knowledge of serious mathematics—otherwise known as an astrophysics textbook. This book will not begin to delve that deeply. However, do not dismay, as there is a fair amount that can be covered, without recourse to any mathematics or background knowledge in cosmology. This will mean of course, that several topics will only be briefly covered, such as dark matter and dark energy. Some of the more esoteric topics, such as Branes and the multiverse, will be covered, somewhat briefly, in the final chapter. But we can cover quite easily the (probable) beginnings of the universe and its evolution to the present day. So let's begin our final journey with a definition of cosmology:

> Cosmology is, to put it simply, the study of the universe, taking into account all the phenomena that occur both in time and space. It deals with all the possible origins, evolution and fates of the universe, and all the physical laws that govern the aforesaid topics.

This is a big concept, and that is an understatement. It is only in the last 25 years or so that cosmology has become an observational science, as before that it was more or less in the domain of the theoreticians, but advances in instrumentation, along with several space-borne experiments, have literally led to a revolution in our understanding, or rather attempts at understanding, the universe we live in and observe.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_18

One of the problems we encounter in cosmology, with most of us being mere mortals and not theoretical or mathematical physicists, is that, at times, the concepts cannot easily be explained using spoken or written language. In fact, it has been said that the closer we get to a fundamental truth in science, the more difficult it is to explain with words. In such cases it can only truly be understood using the language of science, that is to say, mathematics, and very complex mathematics at that!

Let's try to give you an idea of what we mean by this.

The first descriptions, or models, of the atom were those of a nucleus surrounded by electrons in various orbits, and the picture that was used to describe this was likened to the Solar System, with the Sun representing the nucleus and the planets the electrons in their orbits. Later on, when it was realized that this description was inadequate, the description, or model, was changed and this time the electron was likened to a fuzzy area surrounded the nucleus. You can see its already starting to get complicated. Once more as particle physics advanced, the idea of an electron and its nucleus changed again, and now fundamental particles can be likened to strings that vibrate, the ubiquitous *string theory*.

But here's the important point. Notice how we use the words to describe the concepts—solar system, planets, waves and lastly, strings. We are relating terms and words that we are familiar with to give an idea of the science. However, what do you do when there is no comparable term, or word, or model that you can use to describe something, because there is nothing to compare it to?

Here's an example. We are all familiar with the three-dimensional world of length, breath, and width, and we can add time as another dimension. Now imagine a cube in three dimensions, existing in time. Easy isn't it. Now picture in your mind an object with five dimensions.

See how difficult it is. In fact, it isn't difficult at all—it's impossible. We cannot picture or describe such an object using language. However, it is very easy indeed to describe it with mathematics. But we're not going to.

The moral of the above paragraph was to gently introduce you to the fact that some of the concepts covered in this chapter will just have to be taken as given, as the mathematics that describe the phenomena will not be discussed. For those of you that want more information there will be several books listed in the appendices of this book that will satiate your appetite.

Now that you have a foretaste of what's to come, let's begin at the beginning.

18.1. The Big Bang

If ever there was a term that nearly everyone has heard about but may not realize how inaccurate it is, then it is what we call the *Big Bang*—the event that led to the creation of the universe. It wasn't big, and there was no bang. The term Big Bang was first used by the famous British astrophysicist Fred Hoyle,[1] in a BBC radio broadcast, and was said as a way of discriminating it from his own cosmological model, which was the *Steady State*[2] model.

The Big Bang is currently the accepted model for the creation[3] and ensuing evolution of the early universe, with its core idea of an expanding universe. If the concept is correct, then naturally it follows that:

(i) The universe was hotter, and much denser, in the past.
(ii) At a certain moment, all matter and energy was at a single point—the singularity.
(iii) This moment, and point, is the beginning of everything.

Before we go any further, let's just clarify something is always portrayed incorrectly in all documentaries that deal with the Big Bang. The image that is presented in such televisual feasts is of an expanding globe of energy radiating outward from a single center. It couldn't be more wrong. It wasn't expanding into anything; there was no boundary between the edge of the expansion and "outside." There was no outside. The event we call the Big Bang created the universe, and with this model there was nothing to expand into before the Big Bang. Everything that became the totality of space and time was created in this event. Furthermore, the expansion was not from a central point but occurred everywhere!

Now you can begin to see how difficult it is to picture this in your mind, or explain with words. And it doesn't stop here either, because the problem we now have is how do you describe everywhere, when we are talking about a single point, or singularity, that is indescribably tiny. Or may have no dimensions at all. Even time didn't exist, and if it did, it may not have had the characteristics of time that we are familiar with today.

---

[1] The idea of the Big Bang model was first proposed by the Belgian priest Georges Lemaître in 1927.

[2] The Steady State theory was an alternative to the Big Bang theory of the universe's origin. In the steady state, new matter is continuously created as the universe expands. The theory is now obsolete.

[3] There are a few new ideas that suggest that the Big Bang wasn't the beginning of everything, but rather just an event that took place in an even larger universe. This will be discussed in the Chap. 19.

Well, there is one way to deal with all these mind-blowing ideas that seem to contradict all we are familiar with in our everyday life, and that is to accept them and move on.

But before we move on to the next section that deals with the events that occurred after the Big Bang, there is a question you are all probably waiting to be answered—"What caused the Big Bang?"

This is tricky to answer, as there is no answer that can be given—plenty of speculation, but no real concrete explanation.[4] Here's why. Recall that all of space and time was created at the event we call the Big Bang; thus the cause of the Big Bang cannot have occurred with these parameters. It lies beyond them, and it may seem very strange to you to have an event that seems to have no cause. Furthermore, it couldn't have started at a particular time, because time itself was created at the moment the Big Bang occurred. This of course leads to another question—"What was there before?"

Easy answer—there was no *before*. There was nothing.[5]

All this will probably leave you with in unsatisfied state of mind, but don't let this bother you, as it shows us how much more there is to be learned about the universe, and whenever (if ever) we do finally find out the cause of the Big Bang, it will probably be stranger than we could have ever imagined.

Thought Question 18.1

Is there a centre to the universe?

18.2. Hubble and Humason

Before we go any further, let's look at some history of cosmology, early in the twentieth century. It will come as a surprise to many of you when we say that at the beginning of the century, many astronomers actually believed that the Milky Way *was* the universe! There was nothing else, and when astronomers saw these tiny spirals or blob-shaped objects in telescopes, these seemed to actually be within the confines of the Milky Way. Of course some people had other viewpoints, and many debates as to the true nature of the universe were held, but it took the work of a remarkable man to settle the debate once and for all.

---

[4] That's not to say there aren't any ideas as to what gave rise to the Big Bang. There are, a lot of them, but even to just begin to describe them would require having chapters on introductory quantum physics and general relativity.

[5] See Chap. 19, as this idea is now seriously being questioned.

The astronomer Edwin Hubble made some measurements on the spiral galaxy, M31, more commonly known as the Andromeda Galaxy. He was observing Cepheid variable stars, and as we know, one can use these as distance indicators. With the technique, he showed that the galaxy actually lay outside of the Milky Way, an idea that wasn't initially accepted. He made several other measurements on other galaxies that eventually led to a fundamental change of our scientific view of the universe. He is often incorrectly credited with discovering the redshift of galaxies, but in fact that had been done a decade earlier by the American astronomer Vesto Slipher. However, it is his distance measurements, along with his work with fellow astronomer Milton Humason, that gave rise to his fame.

Milton Humason, a former mule-team driver who dropped out of school when he was 14 but went on to became a famous astronomer, was a meticulous observer and made several measurements of the redshift of galaxies. Working with Hubble, they discovered that there was a rough relationship between the distance of a galaxy with its redshift, and recall that a measurement of the redshift can give an object's radial velocity. There was a considerable amount of scatter, but nevertheless they were able to plot a graph with the data from 46 galaxies and obtained a value for what is now referred to as the *Hubble constant* of 500 km/s/Mpc, a value that is far higher than the currently accepted value due to errors in their distance calibrations.

In 1929 Hubble formulated the redshift distance law, nowadays termed simply *Hubble's law*. Simply put, the law states that the farther away a galaxy is, the faster it is moving away from us, with the velocity increasing linearly with distance, i.e., the plot is a straight line. Mathematically, it looks like this:

$$v = \mathrm{H}_0 d$$

where v is the velocity [km s^{-1}].

H_o, the slope of the line, is called Hubble's constant [km s^{-1} per Mpc].
d is the distance [Mpc].

An estimate of the Hubble constant is derived from data obtained by the Planck mission,[6] and is 67.80 ± 0.77 km s^{-1} per Mpc. What this work really did was provide the first observational evidence to support the idea of an expanding universe, proposed in a paper written 2 years earlier in 1927 by

---

[6] Measurements from the Planck mission published in 2018 indicate a lower value of 67.66 ± 0.42, although, in March 2019, a higher value of 74.03 ± 1.42 was determined using an improved procedure involving the Hubble Space Telescope. The two measurements disagree at the 4.4σ level, that means it is beyond a plausible level of chance.

Georges Lemaître. An example of a calculation using redshift and the Hubble law is given in Math Box 16.1. Note, however, that the formula here is for galaxies that are moving considerably slower than the speed of light. For galaxies with relativistic velocities, a modified formula is used.

Having now stated that a value of the Hubble constant, H_o, has been established, a slight problem has recently appeared. Or rather, quite a large problem. Basically, it appears that using a particular method to determine $H_{o,}$ will give a different value to all the other values determined, that use a different method.

There are two basic approaches to determining the value of $H_{o.}$ One is to measure parameters for objects using calibrated distance ladder[7] techniques and is called the "Late Universe" method(s) and so far, the results seem to converge on a value of approximately 73 km/s/Mpc. The second method is the "Early Universe" method and is largely based on measurements of the cosmic microwave background and these tend agree on a value near 67.7 km/s/Mpc, (taking into account the any change in the expansion rate since the early universe).

Over the past few years, the techniques have improved, so that the estimated measurement uncertainties have shrunk. However, the range of measured values has not, to the point that the disagreement is now statistically significant, see Fig. 18.1. This discrepancy is called the ***Hubble Tension***.

In fact, in April 2019, it was reported that there were further substantial discrepancies across different measurement methods in Hubble constant values that may suggest the existence of a new realm of physics not currently well understood. By November 2019, even renowned physicist Joseph Silk referred to it as a "possible crisis for cosmology", as the observed properties of the universe appear to be mutually inconsistent. In February 2020, the Megamaser Cosmology Project published independent results confirming the distance ladder results which differed from the early-universe results. Then, in July 2020, measurements of the cosmic background radiation by the Atacama Cosmology Telescope predicted that the Universe should be expanding more slowly than is currently observed. As of 2021, the cause of the discrepancy is still not understood.

---

[7] The cosmic distance ladder (also referred to as the extragalactic distance scale) is a succession of methods by which astronomers determine the distances to celestial objects. A real direct distance measurement of an astronomical object is possible only for those objects that are "close enough" (within about a thousand parsecs) to Earth. Techniques for determining distances to more distant objects are all based on various measured correlations between methods that work at close distances and methods that work at larger distances. Several methods rely on a standard candle, which is an astronomical object that has a known luminosity.

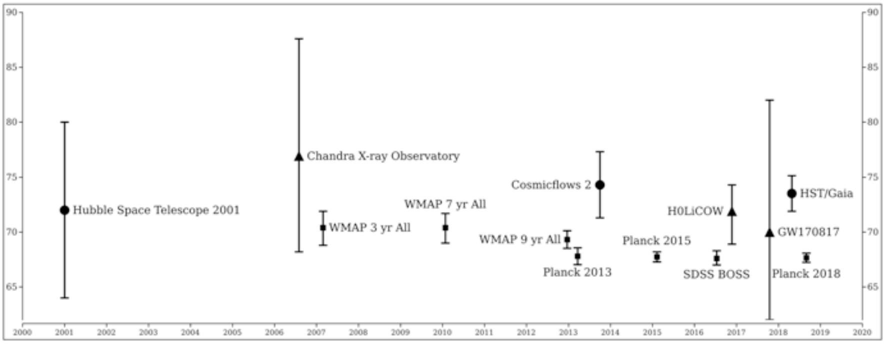


Fig. 18.1. Estimated values of the Hubble constant, 2001–2020. Estimates in black represent calibrated distance ladder measurements which tend to cluster around 73 km/s/Mpc; red represents early universe CMB/BAO measurements showing good agreement on a figure near 67 km/s/Mpc, while blue are other techniques, whose uncertainties are not yet small enough to decide between the two

From all these measurements taken over several years, it appears that increasing the accuracy of the measurements seems to exacerbate the tension problem and has led some scientists to even suggest that there's something about the cosmic expansion of the universe we clearly don't understand at its most fundamental level.

Although a completely rigorous discussion of this topic is beyond the scope of this book, suffice it to say it is very easy [within limits] to determine both the redshift and *recessional velocity*, as it is called, of a galaxy, even from an amateur astronomy point of view, as you will see later.

As an aside it is important to mention that to be completely accurate it is really galaxy clusters, and thus the galaxies contained therein, that are moving away from each other. Galaxies, say, within the Local Group actually have random motions. Witness the fact that we are moving toward M31 and that the Large Magellanic Cloud is moving toward us (Math Box 18.1)!

Math Box 18.1: Redshift

The redshift of an object is the difference between the observed wavelength of a spectral line and its rest wavelength:

$$\text{Redshift} = z = \frac{\lambda_{observed} - \lambda_{rest}}{\lambda_{rest}}$$

Example:
A galaxy has an observed Hα line at 662.9 nm. The rest wavelength of Hα is 656.3 nm. Calculate the redshift of the galaxy and its velocity of recession

$$z = \frac{662.9 - 656.3}{656.3} = 0.010$$

The redshift of the galaxy is 0.010.

For nearby galaxies, when z is much less than 1, its velocity can be calculated by

$$v = c \times z$$

where c = speed of light, 3.0×10^8 meters per second. Thus

$$v = c \times z = \left(3.0 \times 10^8 \text{ m/s}\right) \times 0.01 = 3000 \text{ km per second}$$

The galaxy's velocity of recession is 3000 km/s.

Hubble's Law

Hubble's law allows us to determine the distance to a galaxy if we know how fast it is moving away from us. This velocity is called the *recessional velocity*.

$$v = H_0 \times d$$

H_0 is the Hubble constant and is generally quoted in kilometers per second per megaparsec. The value (take your pick!) appears to be about 67.7 km/s/Mpc.

Example:
Estimate the distance to the galaxy discussed above.

$$d = \frac{v}{H_0}$$

$$d = \frac{v}{H_0} = \frac{3000}{67.7} \approx 44.3 \text{ Mpc}$$

The galaxy is approximately 44.3 Mpc away. As 1 Mpc = 3.26 million light years, this is equivalent to 140 million light years.

Thought Question 18.2

Is the space between the Milky Way and the Andromeda galaxy getting larger due to the expansion of space?

18.3. After the Big Bang

Now that we have all accepted that the Big Bang occurred, we can describe very accurately what happened afterwards, using the physics that's been developed, experimentaly up to around 10^{-10} s after the Big Bang and the rest being theoretical. We can break the events down into stages, or eras, to describe the events.

Planck Era (t < 10^{-43} s)
This era, the "first instant," lasted for just 10^{-43} s, but, as we cannot yet link quantum mechanics, the successful theory of the very small, to general relativity, our successful theory of the very large, we are powerless, sadly, to describe what happened in this era.[8] The Planck Era, 10^{-43} s after the Big Bang, is as far back as our current science allows us to go.

At the current time in the universe, there are four very disparate forces that govern all phenomena in the physical world, and you will be familiar with at least some of them. They are:

- *Gravitational force*—Surprisingly a weak force, but very long-ranged. In addition, it is always attractive and acts between all of matter in the universe.
- *Electromagnetic force*—Long-ranged and is responsible for electric and magnetic effects such as the interaction of magnets.
- *Strong nuclear force*—Strong but short-ranged, it is responsible for keeping nuclei together.
- *Weak nuclear force*—Very weak force, responsible for both radioactive decay and neutrino interactions.

During the Planck Era, the temperature was incredibly high, greater than, 10^{32} K, and at this time these four forces were unified. That is to say, the forces were not separate and distinct but behaved as one force. Also, as the universe began to cool, subatomic particles began to form. Almost everyone is familiar with Einstein's famous theory, $E = mc^2$, which states that matter can be converted to energy, but as with all physics equations, it works in both directions, and energy can also be converted to matter.

If the temperature is hot enough, above 10^{12} K, then gamma rays can form protons, neutrons and electrons. Thus:

---

[8] Even though enormous strides have been made with string and M theory, recall from Chap. 1 that any theory has to have experimental proof for it to be accepted, and as yet, there isn't any definite proof for string theory. Admittedly, it does look as if it will be the way forward, but it has some way to go. With that in mind, I suppose it should be called the string hypothesis.

$$\gamma \rightarrow p^{+} \;\; n \;\; e^{-}$$

and

$$\gamma \rightarrow p^{-} \;\; n \;\; e^{+}$$

where γ is a gamma ray, and p, n and e are protons, neutrons and electrons. Notice however that the first set consists of a p^+ and an e^-; this is matter, whereas p^- and e^+, is referred to as antimatter, with p^- the antiproton and the e^+ a positron.

GUT Era ($10^{-43} < t < 10^{-36}$ s)

As the temperature of the universe continues to fall, the unified force divides, for want of a better word, into two forces: gravity and something called the *GUT* force, GUT standing for the *grand unified theory* force. Think of this GUT force as the unification of the electromagnetic, strong nuclear and weak nuclear forces. During this incredibly brief moment in time, the temperature fell to about 10^{29} K, and this cooling lasted until the universe was 10^{-38} s old. Then another major event occurred; the strong force "froze out" of the GUT force. The effect was dramatic, with an unfathomable amount of energy being released as this happened. This led to something that is still happening today—the universe dramatically expanding in a process we refer to as *inflation*. The physical principles that caused inflation are not fully understood, but it does lead to observable properties that we can see today and which we will discuss later in the chapter, although briefly.

Incredibly, inflation caused the universe to expand from a radius of 10^{-50} m in radius at 10^{-35} s to almost 1 m in radius at 10^{-34} s. This is considerably faster than the speed of light (an understatement). However, you are no doubt aware that nothing travels faster than light, and this is correct, but the laws of physics are not violated. Nothing is moving through space here, as you would imagine, say, as a rocket or galaxy does, but rather *it is space itself* that is expanding. So the laws of physics hold true.

Electroweak Era ($10^{-36} < t < 10^{-12}$ s)

At the beginning of this era, the universe contained three natural forces—gravity, the strong nuclear force, and the electromagnetic and the weak nuclear force (combined as one). However, the temperature continues to fall, and when it gets to about 1015 K, the electromagnetic force and the weak nuclear force separate. Thus we see that all four forces are now distinct, and the temperature of the universe at the end of the Electroweak Era is about 100 million times hotter than the center of the Sun.

It is with experiments carried out at these energies that have successfully detected the electroweak W and Z (weak) bosons—the carriers of the electroweak force—that has led us to have confidence about predicting the state of the universe from 10^{-10} s after the Big Bang to the present time.

Era of Particles (10^{12} s < t < 10^{-3} s)

At the beginning of the Particle Era, the spontaneous creation and annihilation of particles manages keep the number of photons and subatomic particles roughly in balance. But as the temperature of the universe falls, fewer and fewer photons will have sufficient energy to create particles. This results in the photon-to-particle ratio increasing with time, and the variety of particles settles into the more stable forms. These could be protons, neutrons, electrons, neutrinos and perhaps even the elusive and undetected WIMP[9]'s, the speculated particles responsible for dark matter.

There is however a slight imbalance of matter over antimatter, and this resulted in the universe we see today, where there is considerably more matter rather than antimatter. At the present time, the current ratio of photons-to-protons is around 10^9:1, and so, for every billion protons and antiprotons that annihilated each other during this era, one proton was left over by the time the Particle Era came to an end, at 10^{-3} s, with a temperature of about 10^{12} K. But consider this: Some of those protons and neutrons that formed at this point, and later, would have eventually made their way into the atoms that make up your body.

Era of Nucleosynthesis (10^{-3} s < t < 3 min)

During this era, protons and neutrons started fusing, but not for long, as these new nuclei are torn apart by the high temperatures. This time period lasts for about 3 min, and as the temperature falls to around 10^9 K, fusion ceases. Note that this temperature is the same, roughly, as in the center of the Sun, where fusion occurs. However the density of material in the universe was too low to allow fusion to continue, whereas in the center of the Sun the density is high enough for fusion. Afterwards, the ordinary (baryonic) matter leftover in the universe was 75% hydrogen nuclei (i.e., individual protons) and 25% helium nuclei, along with trace amounts of deuterium (H isotope) and lithium nuclei, and except for the small percentage of heavier elements (metals) forged in stars, the composition of the universe has not changed much since the end of the Nucleosynthesis Era, with the 75%/25% ratio more or less the same as is observed today.

Era of Nuclei (3 min < t < 3.8×10^5 year)

At the beginning of this era, the universe is a hot plasma of hydrogen and helium nuclei along with free electrons. The photons are bouncing from electron to electron, changing direction and not traveling very far, and thus we say that matter and photons are *coupled*. The result is that the universe

---

[9] Weakly Interacting Massive Particles.

is opaque, and if there is any ionized hydrogen or helium about that acquires an electron they are quickly ionized again.

However, the temperature of the universe continues to fall, and when it falls to about 3000 K, which occurred around when the universe was 377,000 years old, the low temperature allows electrons to combine with nuclei to form stable atoms of hydrogen and helium, But, in addition to this nuclei formation, the universe now becomes transparent, as the photons are free to pass by nuclei without being absorbed, as photons do not have enough energy to lift electrons to higher atomic orbits. This is sometimes referred to as *decoupling*.

These photons, that were around at the time of decoupling, are in fact the same photons that we see today in the cosmic microwave background (CMB), but after they have been greatly cooled by the expansion of the universe.

Era of Atoms ($3.8 \times 10^5 < t < 10^9$ year)

The universe is now filled with atomic gas, and during this time the only radiation emitted is the 21-cm spin line of neutral hydrogen. This time is sometimes referred to as the *Cosmic Dark Ages*, and current ideas estimate that it lasted between 150 million to 800 million years after the Big Bang.

During this time, small density enhancements in the gas, along with gravitational attraction by dark matter, tends to clump the material that eventually forms protogalactic clouds. In addition, the first star formation lights up the universe, which in turn is believed to have induced the formation of galaxies.

Era of Galaxies ($t > 10^9$ year)

It was originally thought that the first galaxies came into existence about 1 billion years after the Big Bang; however, recent observations suggest otherwise. The galaxy UDFj-39546284 is thought to have been formed some 480 million years after the Big Bang, or about halfway through the Cosmic Dark Ages.

This is the current era of the universe.

18.4. Evidence for and Against the Big Bang Theory

In Chap. 1 we discussed the scientific method, which states that any good scientific model should make predictions that can be verified. If we were to accept the Big Bang model as the correct description of the creation of the universe then it would make sense to make predictions from it, which we hope we can observe. One of these predictions is the existence and characteristics of the *cosmic microwave background* (CMB).

In 1965 Arno Penzias and Robert Wilson accidentally discovered the CMB, predicted by theory to have originated from the decoupling of photons from matter at the end of the Era of Nuclei, and it appeared to come from every direction. Their experiment showed that it had a perfectly thermal spectrum with a temperature of 2.73 K, the temperature one expects after expanding the universe 1100 times.

The spectral distribution of this radiation is the same as radiation from a 3000 K object (recall Wien's law from Chap. 1), just like the surface of a red giant, but since then the universe's size has expanded 1100 times, and the cosmological redshift has turned this radiation into microwaves.

Space-borne experiments such as COBE (Cosmic Background Explorer) and WMAP (Wilkinson Microwave Anisotropy Probe) have confirmed the spectrum to a very high accuracy, but they have also shown that across the sky there exist subtle variations of a few parts in 100,000 (with motion of spacecraft relative to CMB removed). These small variations were believed to reflect increased ordinary baryonic matter densities, that is to say, the kernals on which galaxies will eventually form. However, since the first galaxies formed within a billion years of the Big Bang, the implied ordinary matter (baryonic) densities from CMB enhancements suggest that there wasn't enough to create galaxies that fast. But if we take WIMPs into consideration, since they do not interact with photons and thus will not enhance the CMB, then they must have formed much higher density concentrations before baryons appeared, and these WIMP concentrations would allow galaxies to form faster. This then tells us that the CMB enhancements must be echoing the WIMP density.

So far, we have considered the evidence that supports the Big Bang theory, but prior to 1980, cosmologists had already identified three major questions that the theory was unable to answer:

1. Where does the structure that we observe on the largest scales come from? Galaxies are thought to have been seeded by the mass density variations implied by variations in the CMB. These variations in turn imply dark matter variations. But where do these dark matter variations come from?
2. Why is the large-scale universe so uniform (also known as the *horizon problem*)? The CMB implies that the density of the universe at the end of the Era of Nuclei was no more than 0.01%. But this is remarkably smooth considering the vast expanse of the universe, even then.
3. Why is the total density of matter and dark energy so close to the critical density (also known as the *flatness problem*)? Is it really just a coincidence? Or is there some underlying reason?

All of these above questions can be answered to some degree by accepting an idea that was first proposed by a young American physicist and cosmologist in the early 1980s. So let's take a look at this idea.

18.5. The Inflationary Model

In 1981, Alan Guth proposed that grand unified theories could solve the problems of the Big Bang model as accepted at the time, if one accepts the premise that when the GUT force separated into the strong and electroweak forces, a tremendous amount of energy was released. This sudden and tremendous expansion is called *inflation*.

Although we won't go deep into the processes involved here, we can make some general statements and at the same time try to answer the problems stated above, but first let's make a few assumptions about the universe that are collectively known as the *cosmological principle*.

The cosmological principle is a series of fundamental statements we make about the universe on the largest scales.[10] They are:

- *Homogeneity*—On the largest scales, the local universe has the same physical properties throughout the universe. Every region has the same physical properties (mass density, expansion rate, visible vs. dark matter, etc.)
- *Isotropy*—On the largest scales, the local universe looks the same in any direction that one observes. You should see the same large-scale structure in any direction.
- *Universality*—The laws of physics are the same everywhere in the universe.

With these ideas in mind, let's know look at how the problems of the Big Bang were addressed.

---

[10] As you would expect, recent discoveries have brought the cosmological constant into question. The Clowes-Campusano Large Quasar Group, discovered in 1991, has a length of 580 Mpc, slightly larger than the consistent scale. The Sloan Great Wall, discovered in 2003, has a length of 423 Mpc, only just consistent with the cosmological principle. U1.11, a large quasar group discovered in 2011, has a length of 780 Mpc, two times larger than the upper limit of the homogeneity scale. The HUGE Large Quasar Group, discovered in 2012, is three times longer than that, and twice as wide as is predicted possible according to current models, and so challenges the understanding of the universe on large scales. And at the time was thought to be the largest structure in the observable universe. However, in November 2013, an even larger new structure, 10 billion light years wide was discovered, the Hercules-Corona Borealis Great Wall. All these discoveries have been influential in providing further doubt concerning the validity of the cosmological principle.

18.5.1 Inflation and the Cosmic Microwave Background Variations

Quantum mechanics and the uncertainty principle state that at any point in space there are random energy fluctuations—even in a perfect vacuum, and these energy fluctuations could have been enormous at the beginning of the universe, but remember that they will still be microscopic in physical extent, as the universe itself was small. What inflation does is increase the physical size of these energy and mass fluctuations to, say, Solar System size in just 10^{-36} s after the Big Bang. As a consequence, the variations seen in the CMB that we discussed in a previous section are therefore just the result of these quantum fluctuations that occurred at the very beginning of time.

18.5.2 Inflation and the Horizon Problem

Other than being a coincidence, for any two regions in space to have identical properties, such as the temperature measured by COBE and WMAP, they must be able to communicate with each other by physical contact, or the transmission of information. With this in mind, no matter what direction we observe the CMB, for it to have a near identical temperature, cannot be a coincidence. However, if we accept the standard Big Bang with its nearly linear expansion rate, the model will not have allowed regions today that are separated by immense distances to have communicated any time in the past. However, inflation can overcome such an obstacle by having any two regions being in close contact before the onset of inflation. Let us emphasize once again that inflation does not violate the speed-of-light limitations since objects in the inflationary era are moving apart due to the expansion of space and not by their motion through it.

18.5.3 Inflation and the Flatness Problem

Einstein's theory of General Relativity predicted three categories for the shape of the universe:

(i) *Flat* (or critical) with the spatial geometry like that of Euclid.
(ii) *Spherical* (or closed) with a spatial geometry similar to that found on the surface of a sphere.

(iii) *Saddle Shaped* (or open) with a spatial geometry similar to that found on the surface of a saddle.

Which shape the universe takes will depend on the overall density of the universe relative to the *critical density* of the universe. This critical density is the value at which the universe is in balance (neither expanding or contracting), and expansion has stopped. In addition, if at the end of the Era of Nuclei, the universe deviated by only a few percent from the critical value, then it would have collapsed or it would have expanded too fast to create the present universe we observe today.

So it would appear that inflation's enormous expansion rate actually forces the universe to appear flat within the observable universe, thus solving the flatness problem. However, it does lead us onto another slight problem, and that is the dark matter needed to bring the density up to critical. We'll leave a discussion of dark matter to the next section.

Thought Question 18.3

Is the Big Bang still occurring?

18.6. Dark Matter and Dark Energy

There are few topics in cosmology that can arouse as much debate and speculation as dark matter and dark energy. Even their names conjure up exotic scenarios in the vast distances of space. And yet, we still do not know what they are, what they are made of, or how they formed.

Let us go back to the twentieth century when astronomers were attempting to determine the masses of galaxies by measuring the velocities of stars around the galactic nuclei. It was suspected that the stars would orbit the galactic centers in a manner similar to how the planets orbit the Sun, that is to say, the further away from the Sun, the slower the planet's velocity (this is referred to as *Keplerian motion*). The reason for this is that most of the mass in the Solar System resides in the Sun. It was assumed that a similar situation would occur in the galaxy, with most of its mass, stars, gas and dust closer to its center than further out in the spiral arms, and a resulting graph, known as a *rotation curve*, would show this. Nothing could be further from the truth.

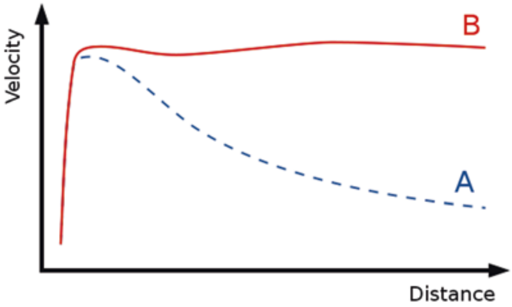


Fig. 18.2. Galactic rotation curves. (Illustration courtesy of Phil Hibbs)

What was observed showed stars revolving around the center of the galaxy with almost the same speed, over a sizable range of distances from its center. An example of such a graph is shown in Fig. 18.2 with the curve A, the expected velocity, and curve B, the observed velocity.

This discrepancy between the theoretical prediction and the observed result had huge implications that there was some sort of matter extending far out beyond the luminous parts of the galaxy, that is, stars and clouds of gas, well into the galactic halo. The problem lies in the fact that we cannot see it visually but only detect it by its gravitational effect on the motions of the stars, hence the term *dark matter*.[11] It should be mentioned here that the possibility of dark matter was actually predicted in 1933 by the astronomer Fritz Zwicky when he calculated the mass of the Coma Galaxy Cluster. Using a method now referred to as the *mass-to-light ratio* he determined the mass of the cluster from the amount of light it emitted (from all the galaxies in the cluster with their constituent stars and gas clouds, etc.). However, the result he obtained gave a cluster mass that was 400 times greater than expected from the luminosity of the cluster, implying that most of the mass of the cluster was dark.[12]

Since Zwicky's work, and the relatively recent galaxy rotation curve observations, many other types of observations have provided evidence for the existence of dark matter, including the velocity dispersions of galaxies, Type Ia supernovae distance measurements, galaxy clusters and gravita-

---

[11] A more expansive discussion on Dark Matter is given in Chap. 19. There will, of course, be some duplication of material.

[12] His result was somewhat high, and recent observations give a smaller but still significant result, implying the existence of dark matter.

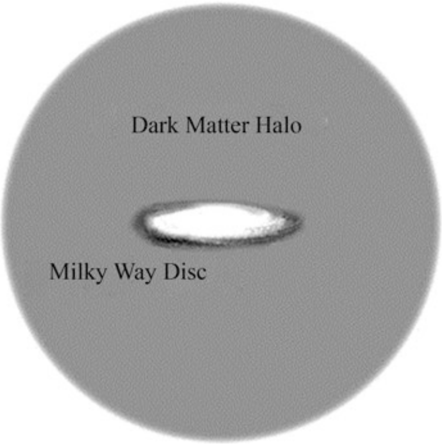


Fig. 18.3. The Milky Way and its Dark Matter halo

tional lensing, and the large-scale structure formation of the local universe.[13]

What is dark matter made of?

That is a very good question, and as yet there is no 100% positive answer, but there are a few candidates, one being WIMPs, or weakly interacting massive particles, and massive compact halo objects (MACHOs).

To get an idea of what the Milky Way would look like if we could see the dark matter, see Fig. 18.3.

Dark energy however presents a very different viewpoint of the universe. Research data suggests that the universe is expanding at an ever-increasing rate, and the mechanism that drives this expansion is what we refer to as Dark Energy. A detailed discussion of Dark Energy is in the next chapter.

At the current time, the make-up of the universe is believed to be thus.

Thought Question 18.4

We see evidence of Dark Energy and its consequential acceleration of the expansion of space, at cosmological distances from us. Will this acceleration of space ever occur in the Milky Way?

---

[13] There are several alternatives to the dark matter idea, the most famous being one that modifies gravity without invoking dark matter, or MOND (Modified Newtonian Dynamics), as it is known. Mordehai Milgromin originally proposed it in 1983.

25.8% ± 1.1% dark matter
69.2% ± 1.2% dark energy
4.84% ± 0.1% ordinary matter

Dark matter and dark energy are fundamental constituents of our universe, and an understanding of their influence, composition and evolution has profound consequences for the future of our universe, and that is what we shall now look at.

18.7. The Future of the Universe

For the past hundred years of so, there were three models that described possible scenarios for the future of our universe, and these differing models depended on critical density.

We know that gravitational attraction between galaxies can overcome the expansion of the universe in localized regions. An example of this would be the gravitational attraction between the Andromeda Galaxy and the Milky Way. But now consider how strong the gravitational force from all the galaxies in all the clusters would have be in order to stop the entire universe from expanding. Obviously it would depend on the total mass density of the entire universe.

It helps here if we think of the energy of the expanding universe as a kinetic energy; the mass density mentioned above, which is the energy required for the gravitational pull to equal this kinetic energy (of the expanding universe) is called the *critical density* ρ_c. This leads us to state the following:

1. If total mass density < critical density, the universe will expand forever.
2. If total mass density > critical density, the universe will stop expanding and then contract.

Now, the value of the Hubble constant, H_o, gives us the current kinetic energy of the universe, and along with other data allows us to determine that the critical density is about 5 atoms of monatomic hydrogen per cubic meter, which isn't much at all, whereas current estimates of the density of ordinary matter is about 0.2 atoms per cubic meter. Thus the conclusion we arrive at is that all the luminous matter that we observe accounts for less than 1% of the critical density. Even if we take dark matter into consideration, there still wouldn't be enough of the combined ordinary and dark matter to give a value greater than the critical density. This leads to the inevitable conclusion that the universe will expand forever.

These ideas were developed before the discovery of the accelerating expansion of the universe due to dark energy, and so taking this discovery into account we can now produce four possible scenarios for the future of the universe:

- (i) *Recollapsing universe*—The expansion will someday halt and reverse.
- (ii) *Critical universe*—It will not collapse but will expand more slowly with time.
- (iii) *Coasting universe*—It will expand forever with little slowdown.
- (iv) *Accelerating universe*—The expansion will accelerate with time.

It is the accelerating universe model that is currently favored.

Before we leave cosmology, there are a few final points to be made. Firstly, it is possible, using the Hubble constant, to derive an age for the universe, but it is important to remember that the constant may not have been "constant" throughout the age of the universe. It may have had a different value in the past, and we also have to take into account the effect of dark matter and dark energy, discussed previously. With that in mind, the current best estimate for the age is 13.787 ± 0.020 billion years.

Secondly, if we use this age, you might assume that the size of the universe is 13.787 billion light years in radius. But you would be wrong, because that would be the size for a static universe—one that isn't expanding. But we know it has been expanding, and so objects that were close to us in the distant past are now much further away. The current best estimates now put the edge of the observable universe at about 46–47 billion light years away. Finally, this *observable horizon* isn't the edge of the universe but just the observable limit. It carries on far beyond the horizon, perhaps to infinity.[14]

We have now, briefly, covered the main topics of cosmology. So without further ado, let's move on to the few observable objects that encompass cosmology.

18.8. Cosmology and the Amateur Astronomer

Let us state right now that, alas, there isn't much in the way of observational cosmology that can be attempted by an amateur astronomer, but that's not to say there isn't any. There is, but it's not a lot.

---

[14] Although the spatial size of the entire universe is currently unknown, the cosmic inflation model suggests that it must have a minimum diameter of at least 23 trillion light years.

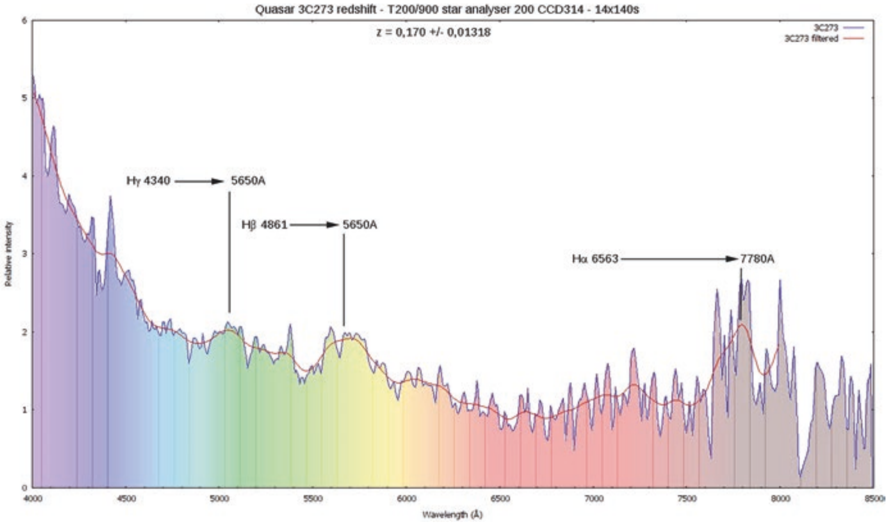


Fig. 18.4. The spectra of the quasar 3C 273 showing the redshift due to expansion of the universe. It was taken with a Star Analyzer 200. Several comparable spectra have also been taken with an 8-in. telescope.

Quasar Spectroscopy

Obviously one of the easiest observations to make is to take spectra of close galaxies and measure their redshift, thus allowing you to determine the recessional velocity of the galaxy. An example of such a spectra is given in Fig. 18.4. This was taken using equipment mentioned in the caption that is well within the reach of all amateur astronomers, and the results are spectacular.

Gravitational Lensing

It is possible, under the right conditions, to see one of the most fascinating consequences of both dark matter and Einstein's general theory of relativity—*gravitational lensing*.

A simple outline of what happens in general relativity will help you to understand this. Gravity has the ability to "bend" light, if the gravitational force is strong enough. The first experimental justification of Einstein's theory was in fact a measure of this light bending, when on May 29, 1919, the British astronomer Arthur S. Eddington measured the amount that starlight was deflected by the Sun. He took his measurements during a total eclipse of the Sun so that any faint stars would not be rendered invisible by the glare of the Sun. The accuracy of the measurements was about 20%, but it was enough to vindicate the theory. Subsequent measurements using radio waves have managed to confirm the predictions made by Einstein to within 1%.

The Sun is not the only thing that can bend a ray of light. Any object that has sufficient mass can deflect light waves. Calculations show that when light rays from a distant object pass close to a compact but massive galaxy, the bending of the light can result in the appearance of multiple or twisted images. It is as if the galaxy were acting like a lens, and so any object emitting light from *behind* the galaxy has its light bent as it passes close to the galaxy. This bizarre effect is called gravitational lensing.

In 1979, astronomers noticed that a pair of quasars known as QSO 0957 + 561 had identical spectra and redshifts, and it was suggested that these two quasars might in fact be one, the two images produced by an intervening object. This was subsequently proved to be the correct explanation, whereby the light from the distant quasars was being lensed by an intervening cluster of galaxies.

You may have seen images of such objects in various books and magazines and always thought that it would be nearly impossible to see these through a telescope. The examples always given are usually of quasars so distant that the Hubble Space Telescope or at least the world’s largest ground-based telescopes are needed to image them.

This is (just) more or less true, but there are one or two quasars that can be and have been seen by amateurs. It isn’t easy. Good seeing conditions are essential to observe these faint objects, and a detailed star atlas is required to confirm the observation, not forgetting a large aperture telescope.

Twin quasar	Q0957 + 0561A/B	10^{H} 01^{M}	55° 53′	February
16.8 m (17.1 17.4 A/B)		Separation 6″	8,000,000,000 l. y.	

The quasar is in the constellation Ursa Major, and so is a fine target for northern hemisphere observers. The starting point for the quasar is the bright edge-on galaxy NGC 3079 (8.1′ × 1.4′, magnitude 11.5, within reach of a 20-cm telescope; several fainter galaxies lie nearby). The galaxy points to the quasar, which is to the southeast, about two galaxy lengths away near a parallelogram of thirteenth- and fourteenth-magnitude stars. The quasar lies off the southeastern corner. The two components are 17.1 and 17.4 magnitude, separated by 6″. Observers with very large instruments of aperture 50 cm have reported seeing the two objects cleanly split. Like most quasars, Q0957 + 0561 is slightly variable in brightness. With small telescopes, the two images will appear as one but slightly elongated. In this case, the lensing is done by a cluster of galaxies, which lie 3.5 billion light years away and is splitting the light of the more distant Q0957 + 0561 into multiple images. Two of these images are much brighter than the others, and this is what is observed. It may be wise to try as high a magnification as possible. This is a good observing challenge for CCD owners. The quasar lies at a distance of almost 8 billion light years and may well be the most distant object visible to the amateur astronomer.

Leo Double Quasar	QSO 1120 + 019	11^{H} 23.3^{M}	01° 37′	March 13
15.7 20.1 m (A/B)	Redshift (z) 1.477			

Now for a challenge, as this is an extremely difficult quasar to resolve. The brighter A component is easily seen in large telescopes, but the fainter B component is very difficult.

Cloverleaf Quasar	H 1413 + 117	14^{H} 15.8^{M}	11° 29′	April 25
17 m (A/B/C/D)	Redshift (z) 2.558			

An exceedingly difficult object to observe, and only with perfect conditions and very large-aperture telescopes will it be seen. The greatest separation among the four is about 1.36″. It has probably never been visually observed from the UK, but US observers report seeing just an asymmetric, faint, hazy and tiny blob of light, although it has been imaged by CCD.

18.9. Olber's Paradox

The final observational topic concerned with cosmology isn't so much a telescopic observation but more of a naked-eye one, and it involves asking a question. This is perfect for a star party where you can spring it onto unsuspecting members of the public before you give the surprisingly simple but profoundly deep answer.

The question is often referred to as *Olber's paradox*, named after the German astronomer Heinrich Wilhelm Olbers,[15] and goes something like this: *Why is the sky dark at night?*

This isn't as daft as it sounds.

If we lived in an infinite universe, one that had no beginning and thus no limit, then wherever we look, we should have, in our line-of-sight, a star. If the stars and galaxies are distributed throughout all of infinite space, then in whatever direction one looked, one should see a star. The night sky would be ablaze. But it isn't, hence the paradox.

You can try thinking about it like this: If you were in a very, very big forest, then in no matter in what direction you looked, you would see trees, and behind them more trees, and so on and so forth. You wouldn't see to the

---

[15] Although named after him, others had posed the paradox long before—Kepler and Thomas Digges, a sixteenth century English astronomer.

edge of the forest, but wherever you looked your line of sight would impinge on a tree.

So, the answer to the paradox is: *The observable universe isn't infinite. It has a beginning.*

The solution to Olbers' paradox, the reason why the sky is dark at night, has to do with the fact that the universe had a beginning—the Big Bang, and that the age of the universe is finite. As we look outward into space we are looking back in time, and only the light from galaxies that has been traveling for 13.7 billion years (approximately) or less, has reached us. We cannot see the light from an earlier time because there was no universe before that time! There are of course galaxies further away, *but the light from them hasn't gotten to us yet*. It's still on its journey. Consequently, we can only observe a finite number of galaxies (and stars).

Just from asking a simple question, we can speculate about the origin of the universe.

We now move onto the final chapter that deals with the most esoteric questions that are currently being researched, or rather discussed, in astronomy. Prepare yourself for a most interesting time.

Problems

1. A galaxy is 100 Mpc distant. Using a Hubble constant of 70 km s^{-1} Mpc^{-1}. Determine its recessional velocity.
2. A galaxy has a recessional velocity of 10,000 kms^{-1}. Using a Hubble constant of 68.9 kms^{-1} Mpc^{-1}. Determine its distance (ignore effects of the expansion of space).
3. Convert the above distance to light years.

Chapter 19

The Speculative Universe

19.1. Introduction

We are all familiar with the topics and nature of star birth, star evolution and star death, with all its accompanying aspects such as white dwarfs, planetary nebulae, pulsars neutron stars, and supernovae. So, it is fair to say that we have a good knowledge of how stars are born, live and die. That is not to say that we understand everything, but rather we have a decent grounding in the basic mechanisms of stellar evolution. In addition, the same could be said about galaxies, the interstellar medium, and the early universe. Research is, of course, continuing, as we strive to understand more of the fundamental details.

However, many theoretical astrophysicists are now coming up with ideas that may have no current observational evidence, but are nevertheless intriguing, highly speculative and in some cases, mind-blowing, that, if correct may result in us having to completely change our ideas about the universe.

It may be that some of these ideas will be shown to be utterly wrong, and perhaps a few may be correct, but as with most research, some ideas, although not completely accurate may contain an inkling of some new astrophysics pointing us toward completely uncharted territory.

Let us begin our journey into the unknown.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5_19

19.2. Dark Matter

19.2.1 The History of Dark Matter

If there is one topic that everyone has heard of, it is Dark Matter, that nebulous for want of a better word, “stuff” that pervades all of space, but whose nature remains as yet, undetermined. It may be a surprise to learn that the initial ideas about what we now call Dark Matter, came about over 140 years ago, albeit in a much-simplified form.

In 1884, Lord Kelvin, a British scientist, estimated the number of “dark bodies” in the Milky Way by observing the velocity dispersion[1] of the stars orbiting around its centre. He was then able to estimate the mass of the Galaxy, and in doing so showed that it was different from the mass of visible stars. He concluded “many of our stars, perhaps a great majority of them, may be dark bodies”.

Them in 1906, Henri Poincaré used the French term *matière obscure* (“dark matter”) in discussing Kelvin’s work.

The first person to suggest the existence of dark matter using the aforementioned stellar velocities was the Dutch astronomer Jacobus Kapteyn in 1922. And fellow Dutchman and radio astronomy pioneer Jan Oort also hypothesized the existence of dark matter in 1932 by studying stellar motions in the local galactic neighbourhood and found the mass in the galactic plane must be greater than what was observed. Alas this measurement was later determined to be erroneous.

In 1933, the Swiss astrophysicist Fritz Zwicky, working at the California Institute of Technology, whilst studying galaxy clusters made a similar inference. Zwicky applied the virial theorem[2] to the Coma Cluster[3] and obtained evidence of an unseen mass he called *Dunkle Materie* (‘dark matter’). He estimated its mass based on the motions of galaxies near its edge comparing that to an estimate based on its brightness and number of galaxies and estimated that the cluster had about 400 times more mass than was

---

[1] Think of this as the spread of different velocities of many objects. It is a statistical approach to measuring the velocity of a group of astronomical objects, such as an open cluster, globular cluster, galaxy, galaxy cluster, or supercluster.

[2] In science the virial theorem is a general equation that relates the average over time of the total kinetic energy of a stable system of discrete particles.

[3] The Coma Cluster (Abell 1656) is a large cluster of galaxies containing over 1000 identified galaxies, and along with the Leo Cluster (Abell 1367), is one of the two major clusters comprising the Coma Supercluster. It’s located in and takes its name from, the constellation Coma Berenices.

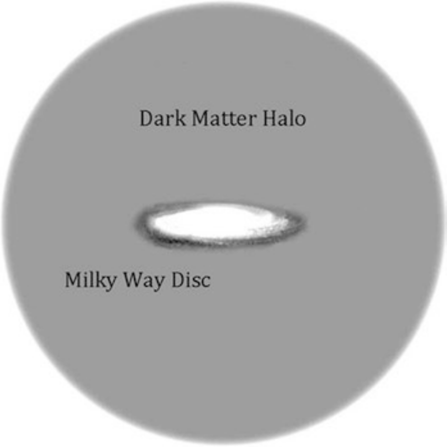


Fig. 19.1. The Milky Way and Dark Matter Halo

visually observable. With these results, he inferred some unseen matter provided the mass and associated gravitation attraction to hold the cluster together.[4]

During the 1970s many astronomers, including Vera Rubin, Kent Ford, and Ken Freeman's postulated further strong evidence for the existence of Dark Matter and in 1980, an important paper showing Rubin and Ford's results revealed that most galaxies contained around six times as much dark as visible mass, and so the obvious need for Dark Matter was widely recognized as a major unsolved problem in astronomy.

To get an idea of what the Milky Way would look like if we could see the dark matter, I refer you to Fig. 19.1, and recent research suggests that as much as 95% of the Milky Way is composed of dark matter, with a mass of perhaps 6×10^{11} to 3×10^{12} solar masses.

19.2.2 *The Observational Evidence for Dark Matter*

There is, without a doubt, a plethora of observational evidence for the existence of Dark Matter, and although we won't go into any details here, it is

---

[4] Zwicky's estimates were off by more than an order of magnitude, mainly due to an obsolete value of the Hubble constant; nevertheless, he did correctly conclude from his calculation that the bulk of the matter was dark.

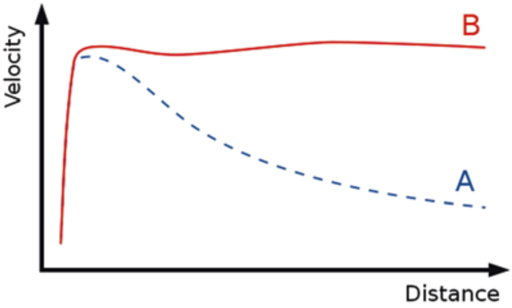


Fig. 19.2. The rotation curve of a typical spiral galaxy: predicted (**a**) and observed (**b**). Dark Matter can explain the “flat” appearance of the velocity curve out to a large radius

noteworthy to briefly mention a few of the more famous and easily understood methods.

19.2.2.1 Galaxy Rotation Curves

It is well known that the arms of spiral galaxies rotate around the galactic centre, and from Kepler’s Second Law, it was expected that the rotation velocities would decrease with distance from the centre, similar to the planets in the Solar System. This is not observed; however, instead, the galaxy rotation curve remains flat as distance from the centre increases. See Fig. 19.2.

If we assume that Kepler’s laws are correct, then the obvious and only way to resolve this discrepancy is to conclude that the mass distribution in spiral galaxies is not the same as that of the Solar System. In particular, there is a lot of non-luminous matter (Dark Matter) in the outskirts of the galaxy.

19.2.2.2 Velocity Dispersions

Stars in bound systems like a galaxy must obey the virial theorem and can be used to measure the mass distribution in systems such as elliptical galaxies or globular clusters. Observations of the velocity dispersion of elliptical galaxies do not match the predicted velocity dispersion from the observed mass distribution, and so, as with galaxy rotation curves, the obvious way to resolve this discrepancy is to assume the existence of a non-luminous matter, i.e. Dark Matter.

19.2.2.3 Galaxy Clusters

Studies of galaxy clusters have shown to be of particular importance because their masses can be estimated in three independent ways:

(i) From the scatter in radial velocities of the galaxies within clusters
(ii) From X-rays emitted by hot gas in the clusters.
(iii) Gravitational lensing (of very distant galaxies) can measure cluster masses.

Usually, these three methods are in good agreement that Dark Matter outweighs visible matter by approximately 5 to 1.

19.2.2.4 Gravitational Lensing

One of the consequences of general relativity is massive objects such as a cluster of galaxies lying between a more distant source, such as a quasar, and an observer should act as a lens to bend the light from this source. The more massive an object, the more lensing is observed.[5]

By measuring the geometry of the observed distortion, the mass of the intervening cluster can be obtained. Note however that the Dark matter does not bend light itself; it is the mass, in this case the mass of the dark matter bends spacetime. Light follows the curvature of spacetime, resulting in the lensing effect (see the chapter on Relativity for more details).

19.2.2.5 Cosmic Microwave Background

Both dark matter and ordinary matter are, simply matter, but ironically do not behave in the same way. In the early universe, ordinary matter was ionized and interacted strongly with radiation. However, Dark matter does not interact directly with radiation, but it does affect the CMB by its gravity (mainly on large scales), and by its effects on the density and velocity of ordinary matter. Perturbations, therefore, of both ordinary and dark matter, evolve differently with time and leave different imprints on the cosmic microwave background (CMB).

---

[5] In May 2021, a new and detailed dark matter map was revealed by the Dark Energy Survey Collaboration, which revealed previously undiscovered filamentary structures connecting galaxies.

19.2.2.6 Structure Formation

The term structure formation refers to the period after the Big Bang when density perturbations collapsed to form stars, galaxies, and clusters. If there were only ordinary matter in the universe, there would not have been enough time for density perturbations to grow into the galaxies and clusters that we observe around us.

19.2.2.7 Bullet Cluster

The Bullet (1E 0657–56) consists of two colliding clusters of galaxies. To be completely accurate, the name *Bullet Cluster* refers to the smaller subcluster, moving away from the larger one, and lies at a distance of 3.72 billion light-years.

Gravitational lensing studies of the Bullet Cluster are often cited as the best evidence to date for the existence of dark matter and provides a challenge for alternate gravity theories because its apparent centre of mass is far displaced from the baryonic centre of mass.[6] The standard Dark Matter models can easily explain this observation, but alternative theories of gravity have a much harder time. If dark matter does not exist, then the next most likely explanation must be that general relativity—the prevailing theory of gravity—is incorrect and should be modified.

Other methods, the details of which we won't go into include Type Ia supernova distance measurements, Sky surveys and baryon acoustic oscillations, Redshift-space distortions and the Lyman-Alpha Forest.

19.2.3 What Is Dark Matter Made Of?

Dark matter refers to any material which interacts mainly via gravity with visible matter such as stars and planets. Consequently, it need not be composed of a new type of fundamental particle, as many think, but could, at least in part, be made up of standard baryonic matter, such as protons or neutrons. However, most astronomers and particle scientists believe that dark matter is dominated by a non-baryonic component, which is probably composed of an as yet unknown fundamental particle or comparable exotic state of matter. Let's discuss the main arguments for both baryonic and non-baryonic matter origins of Dark Matter.

---

[6] This is matter composed of protons and neutrons; ordinary matter, as distinct from exotic forms.

Baryonic Matter

Stars and planets are made of Baryons, but this baryonic matter also comprises less common non-primordial black holes, neutron stars, faint old white dwarfs and brown dwarfs, collectively known as massive compact halo objects (MACHOs), which can be hard to detect due to their relatively small sizes and low luminosities.

However, there is a large amount of evidence to suggest that most of the dark matter is not made of baryons. For instance.

(i) Baryonic gas or dust would be visible when backlit by stars., if there was enough of it.
(ii) The theory of Big Bang nucleosynthesis predicts the observed abundance of the chemical elements. In order to agree with the observed abundances requires that baryonic matter makes up between 4–5% of the universe's critical density. In contrast however, large-scale structure and other observations suggest that the total matter density is about 30% of the critical density.
(iii) Observations of microlensing in the Milky Way have found at most only a small fraction of the dark matter that may be in dark, compact, conventional objects (MACHOs, etc.).
(iv) Detailed analysis of the small irregularities (anisotropies) in the cosmic microwave background measurements by WMAP and Planck indicate that around five-sixths of the total matter is in a form that interacts significantly with ordinary matter or photons only through gravitational effects.

Non-baryonic Matter

There are many hypothetical particle candidates for non-baryonic dark matter such as axions, sterile neutrinos, weakly interacting massive particles (WIMPs), gravitationally interacting massive particles (GIMPs), supersymmetric particles, geons, or primordial black holes.

And what does a WIMP actually do? Well, the WIMP interacts only through the weak nuclear force, discussed earlier, and via gravity. It also has a large mass compared to standard particles, thus its observed characteristics of having a substantial gravitational effect.

In addition, there are three neutrino types that have already been observed and that are abundant, and dark, and matter, but due to their individual masses—which are at the current time not known with complete accuracy—are almost certainly too tiny, so that they can only provide a small fraction of dark matter.

Furthermore, unlike baryonic matter, nonbaryonic matter did not contribute to the formation of the elements in the early universe and thus its presence can only be revealed by its gravitational effects, or weak lensing.

The current Dark Matter candidates, both baryonic and non-baryonic are

I. Light dark matter—Dark matter weakly interacting massive particles candidates with masses less than 1 GeV.
II. Mirror matter—An hypothetical counterpart to ordinary matter.
III. Exotic matter—Any kind of non-baryonic matter.
IV. Neutralino—Neutral mass eigenstate formed from superpartners of gauge and Higgs bosons.
V. Dark galaxy—A hypothesized galaxy with no, or very few, stars.
VI. Scalar field dark matter—Classical, minimally coupled, scalar field postulated to account for the inferred dark matter.
VII. Self-interacting dark matter—Hypothetical form of dark matter consisting of particles with strong self-interactions.
VIII. Weakly interacting massive particles (WIMP)—Hypothetical particles that are thought to constitute dark matter.
IX. Weakly interacting slim particle (WISP)—Low-mass counterpart to WIMP.
X. Strongly interacting massive particle (SIMP)—Hypothetical particles that interact strongly with ordinary matter but could form the inferred dark matter despite this.
XI. Chameleon particle—Hypothetical particle that couples to matter more weakly than gravity.

It is worth stressing at this point that as of summer 2022 ***none*** of the above have been detected.

It is important to point out here that at the present time, these WIMP's are still hypothetical, and although there are many experiments going on in an attempt to detect and classify a WIMP, none have yet reported success. Recent research suggests that as much as 95% of the Milky Way is composed of dark matter, with a mass of perhaps 6×10^{11} to 3×10^{12} solar masses.

19.2.4 Alternative Hypothesis of Dark Matter

Our discussion of Dark Matter will be incomplete if we fail to mention a few of the alternative hypothesis that exist because Dark Matter has not, as yet been successfully identified. Probably the most common technique is to modify general relativity, and although General relativity is well understood on solar system scales, surprisingly its validity on galactic or cosmological scales is not well proven. Therefore, an appropriate modification to general relativity could conceivably eradicate the need for dark matter.

The best-known hypotheses of this type are Modified Newtonian Dynamics, or MOND, as it is more commonly known, and its relativistic generalization tensor-vector-scalar gravity (TeVeS), as well as f(R) gravity, and negative mass,[7] and dark fluid, and entropic gravity.

The main problem with many of the alternative hypotheses is that the observational evidence for dark matter comes from many independent observational techniques, and although it is possible to explain a single observation without recourse to Dark Matter, to explain them all in the absence of dark matter is problematical. There has, nonetheless, been some successes for a few alternative hypotheses, for instance a 2016 test of gravitational lensing in entropic gravity, and a 2020 measurement of a unique MOND effect.

But it must be stressed that the predominant opinion among most astrophysicists is that whilst alterations to general relativity could possibly explain a small part of the observational evidence, there is probably more than enough data to confidently conclude that there must be some form of Dark Matter, whatever its constituents, present in the Universe.

19.3. Dark Energy

Dark Energy is a mysterious and enigmatic topic. Once again, we travel back to the end of the twentieth century, and a research initiative that was attempting to discover if the Hubble constant had change in any way over the history of the universe. Using *type Ia supernovae*, as a method of determining its distance, and luminosity, the team attempted to see what changes, if any, had occurred to the Hubble constant over time. The data the research team collected surprised everyone, as it suggested that in fact the universe was expanding at an ever-increasing rate, that is to say the expansion was accelerating. Since then, other observations, using differing methods have substantiated this work, and the mechanism that drives this expansion is called *Dark Energy*.

It must be admitted that what dark energy consists of is not yet known. We do however know what its effects are, more-or-less—dark energy is a hypothetical form of energy, which permeates all of space and tends to

---

[7] Currently, the closest known real representative of such exotic matter is a region of negative pressure density produced by the Casimir effect.

accelerate the expansion of the universe. I am sure that having read this, you now understand all there is to know. The problem arises in that this topic is highly theoretical and immerses itself in ideas and mathematics that are very complex and at times, even more obtuse, and really does seem to border on the philosophical, but I shall try to make it clear without recourse to any mathematics or jargon.

Basically, it goes something like this. There are, currently, two proposed forms for dark energy. One is referred to as the *Cosmological Constant*, a constant energy density that fills space, equally everywhere, always. The other form is often referred to as *Quintessence*, a dynamic quantity whose energy density can vary in time and space. Although we really don't know much about the nature of dark energy, we do have evidence for its existence from the redshift data that suggest that the universe has expanded more in the last half of its life and that some sort of energy, that isn't either ordinary matter or dark matter, is necessary for the existence of the flat universe we observe. It is also believed to only interact with our universe via gravity, and no other force.

To give you an idea of our current picture of the universe, the latest data suggest that our universe has the following ingredients.

25.8% ± 1.1% dark matter
69.2% ± 1.2% dark energy
4.84% ± 0.1% ordinary matter

This gives us the astonishing result that we don't actually know what 95.1% of our universe is made of.

Furthermore, if we use Einstein's idea of energy and mass being equivalent, then we find that the density of dark energy is very low, about (1.67×10^{-27} kg/m^3). What this means is that in the solar system, there would only be 6 tons of dark energy within the radius of Pluto's orbit. But on a larger scale, even though its density is low, it will come to dominate the energy density of the universe and the implications of this are provoking.

As the acceleration increases with time, then distant galaxies will *move* further away from us, until their velocities exceed the speed of light. Remember that they are not violating any rules of physics, as it is the space between them and us that is expanding. The light from these galaxies will thus never reach us, and as their velocities increase, so does the redshift, resulting in the wavelength of the light becoming too large to detect. Either way, they're going to disappear from view. Assuming that the amount of dark energy is constant, the current estimate of the distance to this cosmological horizon is 16 billion light years. This means that a light signal from any event that is happening at the present moment would eventually be able to reach us in the future, only if the event were less than 16 billion light years away, but if the signal from the event were more than 16 billion light years away, it would never reach us.

However, there are differing scenarios about the future of the universe, that are extremely speculative, and thus are ideal for this chapter. They are,

- The phantom energy[8] model of dark energy results in a divergent expansion, which implies that the effective force of dark energy continues growing until it overshadows all the other forces in the universe, tearing apart all gravitationally bound structures, and that means galaxies, solar systems, planets, and us. The Hubble constant increases to infinity in a finite time, and even the advancement of time itself will stop. Eventually it would even overcome nuclear forces to tear apart atoms themselves, ending the universe in what is termed the "Big Rip".[9]
- Alternatively, it may be that the dark energy might dissipate over time or even change its behaviour completely and become attractive. This leaves open the possibility that gravity will eventually prevail leading to a universe that contracts in on itself in a "Big Crunch".[10]
- There could even be a dark energy cycle, which implies a cyclic model of the universe in where every iteration of a Big Bang followed by a Big Crunch taking about a trillion (10^{12}) years.

It is important to note however, that none of the above three ideas are supported by observations, and thus remain highly theoretical.

All attempt to directly observe dark energy in a laboratory have failed to detect a new force. But recently, it has been suggested by researchers at Cambridge University, that the currently unexplained data observed in the XENON1T[11] detector in Italy may have been caused by a chameleon model[12] of dark energy.

---

[8] Phantom energy is a hypothetical form of dark energy.

[9] Recent observations of galaxy cluster speeds by the Chandra X-ray Observatory seem to suggest the Big Rip will not happen.

[10] The Big Crunch is a hypothetical scenario where the expansion of the universe eventually reverses and the universe collapses, followed by another Big Bang. Most of the evidence indicates that this theory is not correct.

[11] The XENON dark matter research project, operated at the Italian Gran Sasso National Laboratory, is a series of experiments aiming to detect dark matter particles.

[12] The chameleon is a hypothetical particle that couples to matter more weakly than gravity, and has been suggested as a dark energy candidate.

19.4. The Multiverse

We now discuss a topic in cosmology that is very speculative, and indeed has caused quite a lot of discord amongst theorists—the Multiverse.

This is one of those subjects that really does beggar belief, and for most of us just has to be accepted[13] without recourse to any of the highly mathematical astrophysics one would need to gain even an inkling of the science behind the ideas of a multiverse. Basically, it assumes that there exist an infinite number of other universes, that themselves exist in an infinite "larger" universe. This where to starts to get strange, as we now say that one universe is within a much larger universe. However, before we delve even deeper into the idea of a multiverse, let's say at the outset that these "other" universes may, or may not be anything like the one we reside in.

19.4.1 Basically…

The Multiverse idea suggests that our universe, with all its hundreds of billions of galaxies traversing hundreds and probably thousands of billions of light-years,[14] may not be the only one. There may be an infinite of universes, each of them with their own unique laws of nature, that may have stars and galaxies of stars and galaxies, providing of course that stars and galaxies can exist in these galaxies, and it may even be that there is the possibility of intelligent civilizations in these other universes.

But remember, that at the moment, this is just a theoretical model in cosmology, and is still controversial among many cosmologists and physicists. Some regard the multiverse idea as not worthy of scientific inquiry and even suggest that attempts to exempt the multiverse from experimental confirmation would erode public confidence in science and ultimately damage the study of fundamental cosmology. Others have argued that the multiverse should be regarded as a philosophical notion rather than a scientific hypothesis because it cannot be analytically falsified. Recall from earlier in the book that the ability to disprove a theory by means of scientific experiment is a critical criterion of the accepted scientific method. The famous Princeton cosmologist Paul Steinhardt has famously argued that no experiment can rule out a theory if that theory provides for all possible outcomes.

---

[13] Or not, as the case may be!

[14] It is now believed by some cosmologists that our universe is infinite. This topic is covered later in the chapter, but for, now, just accept it.

19.4.2 How & Why…

The most popular origin for the existence of the Multiverse arises from inflation theory, that describes an event that occurred when our universe was very young—less than a second old. In what was an extremely brief amount of time, the universe experienced a period of rapid expansion, “inflating” to become many orders of magnitude larger than its earlier size.

This period of inflation in our universe is thought to have ended about 14 billion years ago, and Heling Deng, a cosmologist at Arizona State University and an expert in multiverse theory, suggested that inflation did not end everywhere at the same time, so that as inflation ended in some regions, it continued in others.

This implies that whilst inflation ended in our universe, there could exist very distant regions where inflation continued—and may even be continuing. The scenario then emerges of individual universes that “pinch off” of larger inflating, still expanding universes, thus creating an infinite vista of eternal inflation, filled with copious individual universes.

Within this framework of eternal inflation, each individual universe would appear with its own unique laws of nature, forces, fundamental particles, and fundamental constants. Heling Deng further states this could explain why our universe has the properties it does—particularly the properties that are currently awkward to explain with fundamental physics, such as dark matter or the cosmological constant. He says, “If there is a multiverse, then we would have random cosmological constants in different universes, and it is simply a coincidence that the one we have in our universe takes the value that we observed”.

19.4.3 The Many Different Types of Multiverses

If the idea of a multiverse amazes you, then it will come as no surprise that there are equally a variety of Multiverse models that can exist. For instance, there are

1. **Bubble Universes** created from other Big Bang events, so distant we can’t conceive of the distances involved. If we consider our universe a collection of galaxies fashioned by a Big Bang, forever expanding outward, we can envisage our universe encountering another universe created in a similar fashion. On the other hand, distances are so vast these multiverses would never interact.

2. **Multiverse from Repeating Universes** where multiverses are based on an infinite space-time, so that if it is infinite, then eventually the arrangement of particles will repeat themselves.
3. **Braneworlds or Parallel Universes** discusses the idea that the universe we perceive isn't all that exists as there are additional dimensions beyond the three spatial dimensions we perceive, plus time. Thus, there could be other three-dimensional "branes"[15] co-existing in higher-dimensional space, thus acting as parallel universes.

It gets even more complicated as the cosmologist Max Tegmark and theoretical physicist Brian Greene[16] have devised classification schemes for the various theoretical types of multiverses and universes that they might comprise. To discuss them fully is beyond the scope of the book, but I will mention Tegmark's only to indicate the complexity of Multiverse ideas.

The four levels of Tegmark's classification are arranged such that subsequent levels can be understood to encompass and expand upon previous levels.

Level I: An extension of our universe. Space in our universe goes on far beyond what we can see, and perhaps goes on forever—which would mean that infinitely many other regions exist in our own universe, regions similar to our observable universe, where the laws of physics are the same.

Level II: Universes with different physical constant. Infinitely many other regions exist in the same space-time as that of our universe, but they are disconnected permanently from our own universe, and within each of them the laws of physics are different.

Level III: Many-worlds interpretation of quantum mechanics. A kind of space different from the space-time of our universe exists called "Hilbert space," which is infinite-dimensional and abstract, and where the laws of quantum mechanics spawn multiple universes. The universe branches into different realities with every tick of time, whether at every Planck time,[17] which is 10^{-43} s, or at every instant of time when an observation is made. These other whole worlds would not be far away in terms of our

---

[15] A brane is a physical object that generalizes the notion of higher dimensions. In cosmology, the central idea is that the visible, three-dimensional universe is restricted to a brane inside a higher-dimensional space,

[16] Brian Greene's multiverses are far more numerous than Max Tegmark's, and consists of nine types; Quilted, Inflationary, Brane, Cyclic, Landscape, Quantum, Holographic, Simulated, Ultimate.

[17] The time it takes a photon, traveling at the speed of light, to travel one Planck length, 10^{-35} m.

kind of space—so in a sense they are right here—but branched out immensely into this different kind of Hilbert space. This is the "many-worlds" interpretation of quantum mechanics, devised by the then relatively unknown physicist Hugh Everett in 1957 and now enjoying a resurgence.

Level IV: Ultimate ensemble.[18] Every consistent system of mathematics describes some kind of existing world or universe. Max Tegmark says "… every mathematical structure which mathematicians can study is on the same footing and describes some kind of physical universe. I think that the reason that nature is so well-described by mathematics is because in a very deep sense, nature really is mathematics."

19.4.4 The Search for Proof

All these amazing ideas about the Multiverse are all well and good, but, like any other scientific hypotheses, to be accepted in the scientific world, one needs actual physical proof.

There have been a couple of instances, where observations have been put forward as evidence for the Multiverse model.

One such piece of evidence was an explanation of the anomalies in the current structure of the universe that was explained as the gravitational tug exerted by other universes. This idea was put forward by American cosmologist Laura Mersini-Houghton who is a proponent of the multiverse hypothesis. Predictions of her theory were stated as having been successfully tested by data gathered by the European Space Agency's Planck spacecraft that mapped the cosmic microwave background radiation left behind when the universe began 13.8 billion years ago. The findings implied that the universe was just one of billions, or even an infinite number, and the map showed anomalies that cosmologists believe could only have been caused by the gravitational pull of other universes outside our own.

Further evidence was presented when data from the Wilkinson Microwave Anisotropy Probe (WMAP) was analysed that suggested our universe collided with other (parallel) universes in the distant past. However, a more thorough analysis of data from the WMAP and from the Planck satellite, which has a resolution three times higher than WMAP, did not reveal any statistically significant evidence of such a bubble universe collision. In

---

[18] The ultimate mathematical universe hypothesis is Tegmark's own hypothesis that considers all universes to be equally real and can be described by different mathematical structures.

addition, there was no evidence of any gravitational pull of other universes on ours.

A final suggestion as to the possibility of the Multiverse is somewhat more esoteric, and is based simply on the fact that we, as humans exist. It goes like this.

Life, particularly intelligent life capable of making cosmological observations, exists in our universe, and certain, very particular characteristics of our universe appear unique for the development of life, the longevity of stars, the abundance of carbon, the availability of light for photosynthesis and the stability of complex nuclei. However, if our universe is just the only one, then it would not be likely that all these unique features would occur. In other words, so many disparate things had to be just right in our universe that the existence of life seems improbable. Thus, if there was only one universe, it would be unlikely to have life in it. Whereas in a multiverse, there would be enough probabilities for life to occur in at least one universe. This idea is known as the Anthropic Principle.

Finally for this section, there is the question of the size of the multiverse. According to astrophysicist Alan Guth, if we start with the size of the complete Universe in which we find ourselves, then this Universe may be at least 10^{23} times larger than our observable universe. "*Thus, the vast expanse of our visible universe*," says Guth "*is but an insignificant speck within just our own inflating pocket universe, and this universe itself is only one pocket universe among an innumerable or even an infinite number of other pocket universes*".

At the moment, there is no definite experimental or observational evidence as to the validity of the Multiverse, but, over the next few years, as more detailed observations are made as well as advances in theoretical studies, there will be, no doubt, be a few surprises.

19.5. The Infinite Universe

This is a topic that can be described in just a few sentences, but trying to understand it, or even picture it, is very difficult indeed. What an infinite universe means is that the universe has no edge or centre, but extends spatially to an infinite size, and always has. I think all of us are familiar with the idea that at the event we call the Big Bang, the universe was incredibly tiny, and then grew larger as time progressed. This is what we read in textbooks or presented on television, but it is a somewhat simplistic picture. In actuality the *UNIVERSE*[19] wasn't concentrated into a miniscule point at the time of the Big Bang, but the observable universe was concentrated into a point. Let me explain. What we perceive as our universe has always depended on the age of the universe and the speed of light, which is finite, at 300,000 km per second. At the moment of creation, our observable universe was more-or-less infinitely small, but after the event we often call the Big Bang, it expanded, and so we could "see" more of it; after say, 1 s, we could see a distance of 1 light second, and after 1 year, we could see to a distance of about 1 light year, Thus, our observable universe is getting bigger as we see more of it as the universe ages, and can be thought of as an expanding sphere. *This is our observable universe.*

Thus, there is a subtle distinction between the whole *UNIVERSE*, and the part we can see.[20] Current estimates of the age of the universe put it at about 13.8 billion years old. Now, this may lead you to say that the radius of the universe is thus 13.8 billion light years, but no, we need to take into account that the universe has been, and still is, expanding, so in the past 13.8 billion years, the observable universe has grown to a radius of 46.5 billion light years in all directions. Therefore, our observable universe is about 93 billion light years across. But remember that this is just the part of the complete Universe that we have been able to see since out universe was created. As Professor Marilena Loverde, at StonyBrook university puts it, "*The distance between distant objects is increasing, but also the amount of time that light has had to travel has increased with the passage of time*".

Another way to picture this is from Dr. Karim Malik, at University College, University of London, when he says "…*I prefer the idea of the visible universe being inside a cube, and if you run time backwards the cube shrinks, but remains a cube. To get an idea of the whole universe, just add more cubes to the cube on all sides. If you can imagine that, just add more and more, and you get an idea of an infinite universe*".

Having completely confused you about the size of our universe, or the complete Universe, it's now time to move on and completely bewilder you with the next section.

---

[19] What is meant by the *UNIVERSE* here is the total and complete universe, within which may be many other small (!) bubble universes, as discussed in the section on the Multiverse. Therefore, whenever the word is in uppercase and italicized, I mean the complete and total Universe, not the one we live in.

[20] The famous physicist and Nobel laureate Steven Weinberg stated… "The word 'universe,' I suppose, should properly mean the whole thing—everything. But when we think of 'universe,' we sometimes use the word to mean just our Big Bang, the things we can see out to almost 14 billion light-years in all directions".

19.6. What Came First, Inflation or the Big Bang?

I think it is true to say that our familiar description of the universe and its beginning goes something like this…." The moment of the universe's creation is called the Big Bang, and then, after that comes a period called Inflation, then stars, galaxies etc., were formed". This is what is in textbooks and taught at school. But a new more radical interpretation has now come forward, Cosmic Inflation.

Leaving out the more esoteric and mathematical aspects of cosmic inflation, the story goes like this.

Physicist Alan Guth at the Massachusetts Institute of Technology devised cosmic inflation to solve several deep problems in the cosmology of our universe, for instance, why was the early universe extremely consistent, or homogenous, even though separated regions were causally disconnected, what cosmologists call the "horizon problem". Or to put it more simply, why is one region of the universe so like another, even though they are tremendously far apart.

He described inflation this way: "Our universe began with a startlingly brief period of enormous, exponential expansion driven by a peculiar state of matter, called a 'false vacuum,' that actually creates a gravitational repulsion.", and according to quantum field theory, the word "vacuum" indicates a sector of space that has local-minimum energy, but not the lowest possible energy, and the word "false" is used to mean ultimately unstable, though the vacuum can remain stable for a very long time. The significance is that a "false vacuum" can "tunnel" into a lower energy state, releasing or "creating" enormous energy.

In this scenario, Guth continues, the exponential expansion ends, because the false vacuum that's driving the repulsive gravity is unstable, and so it decays, much like radioactive elements decay.

Now, this event, at the end of this initial inflationary period, and after space had expanded exponentially in this short-lived fraction of a second, is what we refer to as the traditional Big Bang, or the Hot Big Bang, as it is sometimes referred to, in that the vast energy that was originally locked up in the false vacuum was released and thus converted into the energy and matter (sub-atomic nuclei, atoms) of the early universe. This energy is what produces an unimaginably hot, uniform soup of plasma and particles, which is the starting point in the more familiar, traditional Big Bang theory. To put it in perhaps, a simpler way, during the period of cosmic inflation,[21] the universe—

---

[21] The question you are all asking is "what occurred to start inflation off? "The answer is that although there's a lot of research and speculation nobody knows.

- filled with energy that was inherent to space itself
- this caused a rapid and exponential expansion
- this stretched the universe flat
- gave it the same properties everywhere
- but with small-amplitude quantum fluctuations
- the quantum fluctuations get stretched to all scales
- inflation comes to an end.
- the energy previously inherent to space itself, is converted into matter and radiation
- this leads to the hot Big Bang, arising from an inflationary state, not a singularity

But it doesn't end just there, because according to this new model of cosmic inflation, whilst the decay of the false vacuum triggered the Big Bang in one part of expanding space, this decay did not happen in all expanding space—because not all expanding space decays at the same time, and each decay could generate a single so-called "pocket universe," Guth's term for a connected region of space-time.

The cosmologist Alex Vilenkin, at Tufts University in Massachusetts, stated "because the space between these pocket universes is expanding very fast, room is being made for new bubbles to form, so there will be an unlimited number of pocket universes formed in the course of inflation. "If this picture is right, Guth said," we see no end to it."

We have, in effect, described Multiverses, and Sir Martin Rees, the U.K.'s Astronomer Royal, calls the multiverse "speculative science, not just metaphysics.", and would be amazed, "if the universe didn't extend thousands of times beyond what we can see. He also asks the question, "if there are many Big Bangs, are they all governed by the same laws of physics?" The "fascinating option," said Rees, is that different physical laws govern the other universes—"space may be different, gravity may be different, atoms may be different. This would mean that reality would consist of all these universes, governed by different laws, and only some tiny subset of them would be governed by laws that would allow complexity to evolve. Most universes would be sterile because, for example, gravity would be too strong to allow complex structures, or atoms would not be stable."

To end this section and chapter, it may well be prudent to say that there are a few physicists who have another viewpoint. In 2021 Bruno Bento presented new research that implied that the universe had no beginning and challenged the generally accepted idea that the universe was born in the Big Bang about 14 billion years ago. Simply put, the conclusion was that the universe may not have had a beginning, and therefore, has always existed, and what we think of as the Big Bang was only “a special moment in the evolution of the universe, and not its true beginning.”

As you can see, there is a lot of speculative astrophysics, and even though most of it may be shown to be incorrect, as with all things in science, no matter how strange an idea may be there is usually a kernel of truth hidden in the hypothesis. I have no doubt that over the next few years some answers may be forthcoming as well as many more questions. We live in strange, yet wonderful times.

Thought Question 19.1

A lot of speculative topics have been discussed in this chapter. Which one, do you think, will be resolved within the next 50 years?

19.7. Endnote

We have now arrived at the end of our spectacular journey, and I hope you have enjoyed the trip as well as being amazed and sometimes astounded by what you have read and hopefully, observed. But this isn’t the end—it is just the beginning because you have only seen a handful of the plethora of celestial delights that await you.

The next time you observe the night sky, just think; you will have an inkling of what those objects are, how they formed, how they could die and what they are made of, whether they be stars, exoplanets, clusters, nebulae, galaxies or even the universe itself.

Incredible!

Happy Observing!

Thought Question Answers

Chapter 1

1.1 The Sun is much closer to Earth than Sirius, thus, its apparent brightness is far greater.
1.2 Brighter stars will have smaller magnitudes than fainter stars.
1.3 It must be very far away in order to be fainter than the Sun, even though it has a far higher luminosity.
1.4 The now smaller aperture reduces the brightness of the image, but it will still look the same. In addition, the resolution, or amount of fine detail, will also be reduced.
1.5 Yes and No. Light pollution will mean that you never fully realise the potential of the telescope even if it has a large aperture. In addition, such a telescope will be very heavy, and one may not be inclined to use it as often as one would, say, with a smaller aperture telescope. On the other hand, if it is possible to transport it to a dark location, the benefits of its large aperture will become apparent, but it may still incur a backache. If you have a large telescope, be careful with it and don't pull a muscle moving it.

Chapter 2

2.1 Stellar parallax was not observed which led people to surmise that the Earth did not orbit the Sun. Also, any motion of the Earth was not felt such as a "Great Wind", or similar, which would indicate the Earth moving.
2.2 Copernicus assumed circular orbits and not elliptical, which is the true description.
2.3 Your weight would be zero, but your mass will still be 100 kg.

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5

2.4 Although they both have around 95% CO_2 in their atmospheres, Mars' atmosphere is around 100 times thinner, thus there is far less of a Greenhouse effect.

2.5 Saturn. Although easily visible to the naked eye, no features (the rings for example) or moons can be seen without a telescope of some sort.

Chapter 3

3.1 An absence of red colour would indicate that the conditions necessary to produce the hydrogen alpha transition are absent. This would mean that hot, blue stars are not in the nebula or even close enough to it for the emission line to be produced.

3.2 Surface temperatures of brown dwarfs can range from about 700 K to 2400 K. Due to the different surface temperatures distinctive characteristic spectral lines will be formed. For instance, class L dwarfs exhibit lines of metal hydrides along with alkali metals, T dwarfs display methane lines, whereas Y dwarfs show ammonia lines.

3.3 Spectral lines are due to the photospheric temperature of the star. With this stars' temperature, the hydrogen lines are weaker than the calcium in the visible spectrum. Most stars have a similar composition of mostly Hydrogen and Helium.

3.4 The lines are probably that of the Sodium (Na) doublet, 579 and 577 nm respectively and Mercury (Hg) 436 and 546 nm respectively. These are due to high-pressure sodium (HPS), metal halide (MH) and mercury vapor (MV) streetlights.

3.5 At such a high redshift and consequently, large distance, it cannot be a star even though it looks like one. Therefore, it can only be an object that has a very high luminosity in order to be observed at such a distance. It is probably a Quasar.[1]

Chapter 4

4.1 Mostly. It lies in the middle of the middle of the range for Mass, Temperature and Luminosity. In addition, it is a main sequence star, along with about 90% of all stars. Note that the vast majority of stars are low mass, low luminosity, low temperature red dwarf, main sequence stars.

4.2 They are Giant and Supergiant stars. Main Sequence red stars are of class M, and very low luminosity, thus, not visible to the naked eye.

---

[1] It is in fact he highest-redshift quasar known (as of December 2017) ULAS J1342+0928.

4.3 Luminosity. The mass of stars varies from $0.08M_\odot$ to about $100M_\odot$,[2] a factor of about 1000. The luminosity however ranges from 10^6 $L_\odot$to around $0.001L_\odot$. This is a factor 10^9. Thus, luminosity varies by a much larger amount than mass.

Chapter 5

5.1 Absolutely not! Aldebaran is a red supergiant and has a surface temperature of only 4000 K which is far too low to excite the hydrogen atoms in order to produce an HII region.
5.2 An emission nebula would consist mostly of the Balmer lines of Hydrogen, along with some helium and lighter elements. The continuum spectra would be more-or-less absent.
5.3 Stars are formed in regions where the gas and dust grain density are high. However, the dust grains are very good at scattering and absorbing visible radiation. Nevertheless, infrared radiation is able to pass through the dust clouds, and therefore an infrared telescope would be needed.
5.4 Its blue colour.
5.5 With the naked eye, the night sky would be utterly black, because the dust grains would block out all the starlight.
5.6 The star with the larger mass, as more massive stars are able to progress through the various stages of star formation faster than lower mass protostars.

Chapter 6

6.1 No. Protostars, although highly luminous are usually hidden deep with a molecular cloud and can only really be seen with infrared telescopes, and in some cases with radio telescopes, such as protostars class 0.
6.2 One could observe it visually during the last stage as most of the dust would have been driven away with the star is on the main sequence.
6.3 Recall that massive stars do everything faster, thus the $15M_\odot$ star will reach the main sequence first.
6.4 Gravity.

Chapter 7

7.1 Very unlikely as open clusters are mostly confined to the disc of the Milky Way
7.2 It is a recently formed cluster, as blue stars are high mass, and consequently have short lives. Thus, for it to have blue stars, it must be relatively young.

---

[2] There are reports from time to time of higher mass stars, but these are extremely rare and of dubious origin. Time will tell.

7.3 No. As open clusters age, the stars are dispersed, so over time individual stars of the cluster spread throughout space, just as the Sun has done

7.4 No. OB associations contain O and B stars, that are hot and blue, but also have very short lives.

Chapter 8

8.1 Depending on the circumstances, the Sun will have to increase its rate of fusion to maintain the balance, otherwise gravity will become dominant. This leads on the next stage of the life of the Sun, a Sub Giant, and then a Red Giant.

8.2 The Planets and the Moon would not be visible as they are only seen due to the light from the Sun reflecting off of them.

8.3 Sunspots can be seen at moderately high latitudes on the Sun, at the beginning of the 11-year cycle.

8.4 No, we can only estimate future solar activity based on past cycles, and even then, the estimate can often be wrong.

8.5 The X-ray and most of the, UV light is absorbed by the Earth's atmosphere.

Chapter 9

9.1 In order for the separation to be seen visually, i.e., through a telescope, the distance between the two stars must be quite large. Thus, using Kepler's Laws, a large separation would result in a long period. Conversely, spectroscopic binaries would need to have smaller separations, and thus shorter periods for the motion of the stars to be detected spectroscopically, for instance using the Doppler effect.

9.2 Observations of the system taken over a suitable period – months or years, would show the two systems orbiting their common centre of gravity, with a change in Position Angle and Separation. A visual double would not show such behaviour.

9.3 At a large distance, it may be impossible to measure the separation of the stars visually as the resolution may not be sufficient to see two distinct stars. Spectroscopic binaries are not affected by the distance.

Chapter 10

10.1 None, their main sequence lifetimes are longer than the current age of the universe**.**

10.2 100%. Red stars on the main sequence are low mass, low luminosity and not visible to the naked eye.

10.3 Even though it has a surface temperature less than the Sun, it has a much larger surface area, and thus in total, emits more energy

10.4 Most globular clusters orbit the central bulge of the Milky Way, thus, due to the Suns location in the disc we see the clusters arranged around the centre of the Milk Way. This part of the sky is often best observed during the summer months

10.5 A low luminosity Cepheid, as they have shorter periods and thus take less time to determine its characteristics.

10.6 RR Lyrae stars have the same intrinsic luminosities, around $50L_{\odot}$. Thus, if we already know its luminosity and compare it to its apparent brightness, we can determine its approximate distance.

Chapter 11

11.1 No, stars of this low mass will take about 18 billion years to evolve into red giants, which is longer than the current age of the universe. Therefore, there would be very little carbon around to initiate life as we know it.

11.2 The difficulties were caused by all the atmospheric carbon hiding the absorption lines that are normally used as temperature indicators for the stars.

11.3 It is probably because we are looking at the basic shape, that of a thick torus around the star along with outflows ejected through the hole in the torus, but from varying points of view.

11.4 You shouldn't be. Massive stars, especially ones with a mass of 12 $M_{\odot}$, result in either a neutron star or a black hole.

11.5 White dwarfs exist for immense periods of time, possibly billions of years. Whereas planetary nebulae are more ephemeral, and at most last only a few tens of thousands of years.

Chapter 12

12.1 Any massive stars that once existed in globular clusters would have exploded as supernova a long time ago. Whereas open clusters could still contain several high mass stars that would end up as supernovae. Thus, you would concentrate your search in open clusters.

12.2 In order for a type I supernova to occur, mass as needs to be transferred from Sirius A to the dwarf star Sirius B. However, in this particular system, the distance between them is too great for this to occur.

12.3 A type I supernova occurs when mass is transferred from a star onto a white dwarf. The white dwarf consists of carbon, oxygen, neon, and magnesium, as well as hydrogen, thus its spectrum will show evidence of these elements, with little hydrogen. A Type II supernova results from a high mass star that is mainly hydrogen and helium, along with some other elements. Hence its spectra would be dominated by the lines of hydrogen.

12.4 No, to be defined as a pulsar the beam, whether radio or some other form of electromagnetic radiation, needs to sweep over the Earth. Even though the neutron star may have such beams, if it does not sweep over the earth, it is not defined as a pulsar.

12.5 Probably not, as a compact object of mass of 8 $M_\odot$ is more likely to be a black hole, and not a neutron star.

Chapter 13

13.1 Newton's equations are far simpler, and for most phenomena on the Earth, the solutions differ very little to those of Einstein's solutions. This is because, believe it or not, the gravity on Earth is quite weak. It is only in the presence of very strong gravitational fields do Newton's laws fail and Einstein's laws dominate.

13.2 The Earth's orbit would remain the same, as the mass of the earth/black hole doesn't change, and there would be no planets for you to observe as they only reflect the light from the Sun.

13.3 As you and the weighing scale are falling with the same acceleration, it would read zero because both you and the scale are essentially in free-fall.

13.4 What relativity really tells us about the speed of light is that it is a limit on the speed at which information can be transmitted. However, while the laser dot can carry information from you to each of the individual stars, it clearly cannot be transmitting any information from one star to the other. (After all, you held the laser, not someone at either star.) As a result, there's no violation of relativity.

Chapter 14

14.1 No, the Earth does not have enough mass to warp spacetime sufficiently to create the horizon.

14.2 Three times larger

14.3 Massive stars end their lives as either neutron stars or black holes, whereas lower mas stars end up as white dwarfs. Recall however that massive stars ae very rare, whereas low mass stars are in abundance. Therefore there are more white dwarfs than either neutron stars or black holes.

Chapter 15

15.1 The methods used in exoplanet detection needed to be very accurate and precise, and these have only been developed in the past 30 years.

15.2 Planets in the solar system are visible because they reflect the light from the Sun. If the solar system was to be observed at a distance, say, of several light years, the Sun would be a faint star, whereas the planets would be invisible as the glare from the Sun would prevent the planets from being seen.

15.3 Yes, star formation is still ongoing in the spiral arms of the Milk Way, therefore it is only reasonable to assume exoplanet formations is also continuing.

15.4 Yes, there are several projects where amateur astronomers can participate in exoplanet research using only commercially bought telescope and equipment. Several universities as we as NASA have programs dedicated to this. Search the web for information.

Chapter 16

16.1 Prior to this the methods of distance measured were inaccurate end limited to objects within the Milky way. Indeed, many thought that the galaxies observed were within the Milky Way. It was only when new techniques of distance determination were developed, such as using Cepheid variables, that it was discovered that galaxies actually lie beyond the confines of the Milky Way.

16.2 Probably an E0, hence the difficulty and limitations of using the Hubble Tuning Fork classification.

16.3 Between us and the centre lies an immense amount of interstellar dust. This is very efficient at scattering and absorbing any visual light. If on the other hand we used an X-ray, or Infrared telescope we would be able to image the central regions.

16.4 The further away a galaxy lies, the fainter it becomes.

16.5 Giant elliptical galaxies tend to be in the centre of a large cluster of galaxies, consisting of several thousand galaxies of many types. They are believed to be formed from smaller galaxy mergers and collisions. The local group has only 80 known members at the present time.

Chapter 17

17.1 An active galaxy would be easier to see in a 200 mm telescope as they are much brighter than quasars. For instance, Messier 77 (NGC1068) is a Seyfert type 2 active galaxy, of magnitude of 8.9, whereas a bright quasar is 3C273 at 12.8 magnitude.

17.2 Quasars are point sources with very high redshifts. Active galaxies tend to look like a galaxy, with lower luminosities (but still higher than "normal" galaxies). You could also mention that quasars have very high redshifts whereas active galaxies have lower redshifts.

17.3 No. In a relatively short time, about 10^8 years, due to the high levels of star formation it is believed that the supply of gas and dust in the galaxy would be exhausted.

17.4 Borrow a spectroscope and obtain the spectra of the star-like object. If it is a star, the redshift will be relatively small, whereas a quasar redshift is considerably larger.

Chapter 18

18.1 This is tricky. Currently, it is believed the UNIVERSE[3] is infinite in extent, and our universe is just a part of it. However, the part of the universe we can see is a sphere some 46 billion light years in diameter, and we are at its centre.

18.2 No. The Andromeda galaxy and Milky Way are moving toward each other due to their mutual gravitational attraction. This mutual gravitational force is stronger than the force which causes the expansion of the universe.

18.3 Another tricky question. It all depends on definitions. Some astronomers refer to the Big Bang as the start of our universe at the moment in time when we can measure and understand the physics, and that's all, and others refer to it as the very moment of the creation of the universe and its continuing expansion. You choose.

18.4 There are several speculative ideas about the future of the universe. The phantom energy[4] model of dark energy results in an expansion that implies that the dark energy force will continue grow until it dominates all other forces in the universe. In this scenario, dark energy would ultimately tear apart all gravitationally bound structures, including galaxies, the Milky Way, the solar system, Earth and humans, and eventually overcome the electrical and nuclear forces to tear apart atoms themselves, ending the universe in a "Big Rip".

Chapter 19

19.1 I have no idea! But I expect some will be proved false, some correct and some will remain unsolved. There may even be new ideas even more amazing. We live in interesting times.

---

[3] The UNIVERSE in upper case, is the complete and infinite universe, of which our universe is but a part. See Chap. 19 for more information.

[4] A hypothetical form of dark energy with negative kinetic energy.

Answers to End of Chapter Problems

Chapter 1

1. 10 pc
2. 32.6 ly
3. 33,000 K
4. O
5. $2.5R_{\odot}$
6. (a) 200X, (b) .232 arcsec., (c) 0.25 degree
7. 2.5

Chapter 2

1. 0.25 AU
2. 0.2
3. 1.9×10^{20} N

Chapter 3

1. 486.135 nm
2. Hydrogen Beta
3. 0.029, 8776 $\mathrm{km^{s-1}}$
4. 3.327,000 $\mathrm{km^{s-1}}$
5. No, at this distance and redshift one must consider the expansion of the universe and use a different redshift formula.

Chapter 4

1. $\sim 10^2$
2. $\sim 3.5 \times 10^{-3}$ $L_{\odot}$

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5

Chapter 5

1. ~1.3×10^{15} m.
2. .04 pc.

Chapter 6

1. ~9 million years

Chapter 7

1. ~2.5×10^{2}.
2. ~1×10^{6}.

Chapter 8

1. 4.3 days
2. ~1.10^{-11} joules.
3. ~1×10^{11} years

Chapter 9

1. ~3.7 $M_{\odot}$
2. 10 $M_{\odot}$
3. 10 AU

Chapter 10

1. 10^{5} years.
2. ~5×10^{12} years
3. Cannot be determined as it is not on main sequence.

Chapter 11

1. ~12.3 $M_{\odot}$
2. 2.8 $M_{\odot}$
3. Another white dwarf perhaps.

Chapter 12

1. 300 kms^{-1}.
2. Blue.
3. 0.066.
4. 6.383%.

Chapter 13

1. 0.714.
2. 17.87.
3. 1.41 light years.
4. 709 kg.

Chapter 14

1. 30 km.
2. 900 km.
3. 3339 km.
4. 27 km.

Chapter 15

1. 11.5 days
2. 172 years.
3. No.
4. Its period is 356 years.

Chapter 16

1. 7.
2. 1.
3. Nearly circular.

Chapter 17

1. 47,400 kms^{-1}.
2. ~700 Mpc.
3. 9.1×10^{10} km.
4. ~9 times.

Chapter 18

1. 7000 kms^{-1}.
2. 145 Mpc.
3. 525 million ly.

Appendices

Appendix 1: Degeneracy

Degeneracy is a very complex topic but also a very important one, especially when discussing the end stages of a stars life. It is however, a topic that sends quivers of apprehension down the back of most people. It has to do with quantum mechanics, and that, in itself is usually enough for most people to move on, and not learn about it. That said, it is actually quite easy to understand providing that the information given is basic, and not peppered throughout with mathematics. This is the approach I shall take.

In most star the gas of which the star is made up of, will behave like an ideal gas, i.e., one that has a simple relationship between its temperature, pressure, and density. To be specific, the pressure exerted by a gas, is directly proportional to its temperature and density. We are all familiar with this. If a gas is compressed it heats up, likewise, if is expands, it cools. This also happens inside a star. As the temperature rises, the core regions expand and cool, and so it can be thought of as a safety valve.

However, in order for certain reactions to take place inside a star, the core is compressed to very high limits, which allows very high temperatures to be achieved. These high temperatures are necessary in order for, say, helium nuclear reactions to take place. At such high temperatures, the atoms are ionised so that it becomes a soup of atomic nuclei, and electrons.

Inside stars, especially those where the density is approaching very high values, say, a white dwarf star or the core of a red-giant, the electrons that make up the central regions of the star will resist any further compression,

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5

and themselves set up a powerful pressure.[1] This is termed degeneracy, so that in a low-mass red giant star, for instance, the electrons are degenerate, and an electron-degenerate pressure supports the core. But a consequence of this degeneracy is that the behavior of the gas is not at all like an ideal gas. In a degenerate gas, the electron degenerate pressure is not affected by an increase in temperature, and in a red giant star, as the temperature increases, the pressure does not, and the core does not expand as it would if it were in an ideal gas. The temperature therefore continues to increase, and further nuclear reactions can take place.

There comes a point however when the temperatures are so high that the electrons in the central core regions are no longer degenerate, and the gas behaves once again like an ideal gas.

Neutrons can also become degenerate, but this occurs only in neutron stars.

For a fuller and more rigorous description of degeneracy, then I recommend any of the astrophysics books mentioned in the Appendix 2. Be warned however, that mathematics is used liberally.

Appendix 2: Book, Magazines, Organizations, and Equipment

There are many fine astronomy and astrophysics books in print, and to choose among them is a difficult task. Nevertheless the few mentioned here are I beleive to be amongst the best on offer. You do not need to buy or even read them all, but it would be in your best interests to check at your local library to see if they have some of them.

In addition, with the advent of the Internet, searching for and ordering books is very simple. Thus it is up to you to find the books, and their associated publishers, in whichever country you happen to reside.

Star Atlases and Observing Guides

A Field Guide to Deep Sky Objects, M. D. Inglis
Observer's Guide to Star Clusters, M. D. Inglis
Astronomy of the Milky Way, Volume I, M. D. Inglis
Astronomy of the Milky Way, Volume II, M. D. Inglis
The Sun and How to Observe It, J. L. Jenkins
Venus and Mercury and How to Observe Them, P. Grego

---

[1] This is a consequence of the Pauli exclusion principle, which states, that two electrons cannot occupy the same quantum state. Enough said I think!

The Moon and How to Observe It, P. Grego
Mars and How to Observe It, P. Grego
Jupiter and How to Observe It, P. J. W. McAnally
Saturn and How to Observe It, J. L. Benton, Jr.
Uranus, Neptune and Pluto and How to Observe Them, R. Schmude, Jr.
Asteroids and Dwarf Planets and How to Observe Them, R. Dymock
Norton's Star Atlas and Reference Handbook, 20th Edition
Sky Atlas 2000.0, W. Tirion, R. Sinnott
Millennium Star Atlas, R. Sinnott, M. Perryman
Uranometria 2000.0, Volumes 1 & 2, Wil Tirion (Ed.)
Observing Handbook and Catalogue of Deep-Sky Objects, C. Luginbuhl, B. Skiff
The Night Sky Observer's Guide, Volumes I & II, G. Kepple, G. Sanner
Deep-Sky Companions: The Messier Objects, S. O'Meara
Deep-Sky Companions: The Caldwell Objects, S. O'Meara
Burnham's Celestial Handbook, R. Burnham
Practical Amateur Astronomy, S. F. Tonkin (Ed.)
Astronomical Spectroscopy for Amateurs, K. M. Harrison
Spectroscopy: The Key to the Stars, K. Robinson
Star Clusters and How to Observe Them, M. Allison
Double and Multiple Stars and How to Observe Them, J. Mullaney
Supernovae and How to Observe Them, M. Mobberley
Observing Variable Stars, G. A. Good
Nebulae and How to Observe Them, S. Coe
Planetary Nebulae and How to Observe Them, M. Griffiths
The Caldwell Objects and How to Observe Them, M. Mobberley
Faint Objects and How to Observe Them, B. Cudnik
Galaxies and How to Observe Them, Wolfgang Steinicke and Richard Jakiel

Astronomy and Astrophysics Books

Astrophysical Techniques, C. Kitchin
Discovering the Cosmos, R. Bless
The Cosmic Perspective, J. Bennett, M. Donahue, N. Schneider, M. Voit
Introductory Astronomy and Astrophysics, M. Zeilik, S. Gregory, E. Smith
Pathways to Astronomy, Schneider and Arny
Introduction to Modern Astrophysics, B. W. Carroll, D. A. Ostlie
An Introduction to the Solar System, N. McBride & I. Gilmour
Stars, J. B. Kaler
An Introduction to The Sun and Stars, S. F. Green and M. H. Jones
Extreme Stars, J. B. Kaler

Stars, Nebulae and the Interstellar Medium, C. Kitchin
Extreme Explosions, D. Stevenson
Galaxies in Turmoil, C. Kitchin
Galaxies and the Cosmic Frontier, W. H. Waller & P. W. Hodge
An Introduction to Galaxies and Cosmology, M. H. Jones & R. A. Lambourne
Exploring Black Holes, E. Taylor & J. A. Wheeler
Dark Side of the Universe, I. Nicolson

Magazines

Astronomy Now
Sky & Telescope
New Scientist
Scientific American
Science
Nature

The first three magazines are aimed at a general audience and so are applicable to everyone; the last three are aimed at the well-informed layperson. In addition there are many research-level journals that can be found in university libraries and observatories.

Organizations

The Federation of Astronomical Societies, UK (http://www.fedastro.org.uk/)
Society for Popular Astronomy, UK (http://www.popastro.com/)
The American Association of Amateur Astronomers, USA (http://www.astromax.com/)
The Astronomical League, USA (http://www.astroleague.org/)
The British Astronomical Association, UK (http://www.britastro.org/baa/)
The Royal Astronomical Society, UK (http://www.ras.org.uk/membership.htm)
The International Dark Sky Association (http://www.darksky.org/)

Spectroscopic Equipment

www.rspec-astro.com

Index

M. Inglis, *Astrophysics Is Easy!*, The Patrick Moore Practical Astronomy Series, https://doi.org/10.1007/978-3-031-16805-5