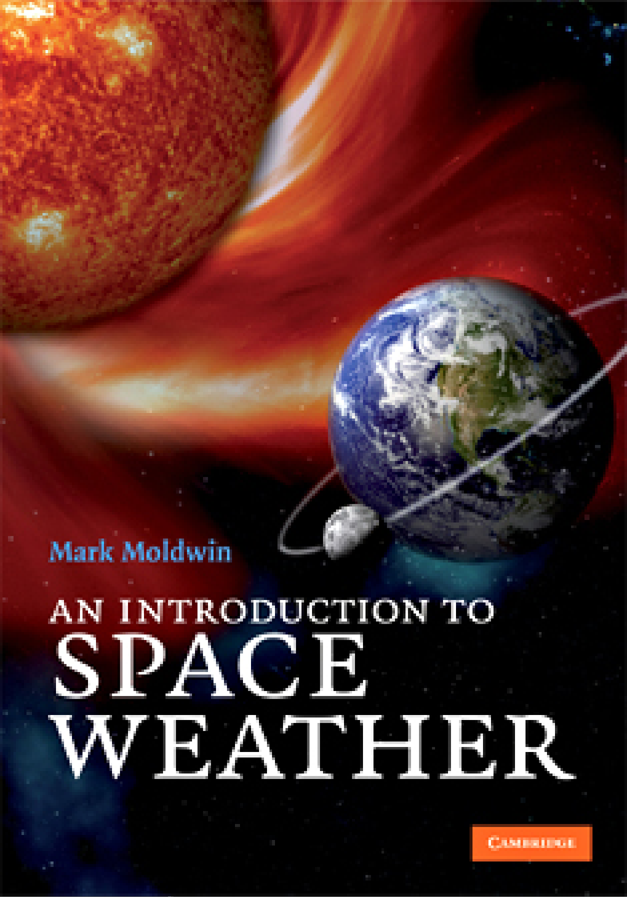


Mark Moldwin

# AN INTRODUCTION TO SPACE WEATHER

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521861496

This page intentionally left blank

## An Introduction to Space Weather

Space weather is an emerging field of space science focused on understanding societal and technological impacts of the solar–terrestrial relationship. The Sun, which has tremendous influence on Earth's space environment, releases vast amounts of energy in the form of electromagnetic and particle radiation that can damage or destroy satellite, navigation, communication, and power distribution systems, and injure or kill astronauts. This textbook introduces the relationship between the Sun and Earth, and shows how it impacts our technological society.

One of the first undergraduate textbooks on space weather aimed at non-science majors, it uses practical aspects of space weather to introduce space physics and give students an understanding of the Sun–Earth relationship. Definitions of important terms are given throughout the text. Each chapter contains key concepts, supplements, and review questions to help students understand the materials covered. This textbook is ideal for introductory space physics courses.

MARK B. MOLDWIN is Professor of Space Physics in the Department of Earth and Space Sciences and the Institute of Geophysics and Planetary Physics, University of California, Los Angeles. His primary research interests are magnetospheric and heliospheric plasma physics, and pre-college space science education and outreach activities.

# An Introduction to Space Weather

**Mark Moldwin**
Department of Earth and
Space Sciences
Institute of Geophysics
and Planetary Physics
University of California
Los Angeles

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York
www.cambridge.org
Information on this title: www.cambridge.org/9780521861496

First published in print format 2008

ISBN-13 978-0-511-39304-4 eBook (EBL)

ISBN-13 978-0-521-86149-6 hardback

# Contents

Preface ix
Acknowledgments xi

1 What is space weather? 1

1.1 Key concepts 1
1.2 Introduction 1
1.3 Brief history 4
1.4 Impacts of space weather on society 13
1.5 Supplements 13
1.6 Problems 16

2 The variable Sun 17

2.1 Key concepts 17
2.2 Introduction 17
2.3 Temperature and heat 19
2.4 Radiation and convection 19
2.5 Solar structure 20
2.6 Dynamics and processes 27
2.7 Supplements 31
2.8 Problems 36

3 The heliosphere 37

3.1 Key concepts 37
3.2 Introduction 37
3.3 The corona and the solar wind 37
3.4 The interplanetary magnetic field 38
3.5 Coronal mass ejections 41
3.6 The outer heliosphere 42
3.7 Cosmic rays 43
3.8 Supplements 44
3.9 Problems 49

4 Earth's space environment 50

4.1 Key concepts 50
4.2 Introduction 50

4.3 Dipole magnetic field 50
4.4 Structure of the inner magnetosphere 52
4.5 Interaction of the solar wind and magnetosphere 54
4.6 Magnetic reconnection 55
4.7 The magnetotail 56
4.8 Plasma sheet convection 57
4.9 Dynamics of the magnetosphere 58
4.10 Supplements 62
4.11 Problems 67

**5 Earth's upper atmosphere 68**

5.1 Key concepts 68
5.2 Introduction 68
5.3 The thermosphere 68
5.4 The ionosphere 71
5.5 Ionospheric structure 72
5.6 Ionospheric variations 73
5.7 The aurora 74
5.8 Impacts on communication 75
5.9 Supplements 76
5.10 Problems 77

**6 The technological impacts of space storms 79**

6.1 Key concepts 79
6.2 Introduction 79
6.3 Satellite orbits 80
6.4 Radiation impacts on satellites 82
6.5 Radio communication and navigation impacts 85
6.6 Ground system impacts 87
6.7 Supplements 89
6.8 Problems 94

**7 The perils of living in space 95**

7.1 Key concepts 95
7.2 Introduction 95
7.3 Radiation 96
7.4 Problems of long-duration space travel 102
7.5 Living on the Moon and Mars 105
7.6 Interstellar travel 105
7.7 Supplements 106
7.8 Problems 111

## 8 Other space weather phenomena 112

8.1 Key concepts 112
8.2 Introduction 112
8.3 Climate variability and space weather 114
8.4 Asteroid and comet impacts 115
8.5 Nearby supernova 117
8.6 Supplements 118
8.7 Problems 121

Appendix A: Web resources 123
Appendix B: SI units 125
Appendix C: SI prefixes 127
References 128
Historical bibliography 129
Index 131

Color plate section is located between pages 94 and 95

# Preface

In the last few decades our technological civilization has become dependent on satellites for global communication, navigation, and commerce. We have also begun the long journey to explore the Moon, Mars, and our Solar System.

This exploration has led to some amazing discoveries about our dynamic Sun and its interaction with Earth. We now know that the Sun is a variable star that expels high-energy particles and deadly radiation continuously out into space. This radiation can impact and destroy technological systems and is one of the major concerns for human space exploration.

In the 1990s, the commercial satellite industry boomed, with direct-satellite-to-home TV markets and satellite communication options expanding. In 2000, the satellite communications industry was doing nearly $100 billion per year of business with nearly a hundred new satellites launched each year. With the increased commercial businesses and the reliance of different markets on space, society began to notice when something went wrong in space.

Galaxy IV was an operating and profitable communications satellite until May 19, 1998, when, after experiencing weeks of intense radiation generated by the Sun and the Sun's interaction with the Earth's space environment, it failed. Galaxy IV carried the signals of over 90% of North America's pagers and several major broadcast networks, including the US National Public Radio (NPR) and CBS. Without the $200 million satellite, millions of pager messages, NPR radio, and CBS television programs never made it to their intended audience. Radio and TV producers were left scrambling to fill dead-air time and medical doctors and business people found themselves out of contact with their hospitals and clients. In all likelihood Galaxy IV was a victim of a space weather storm. Space storms can not only damage or destroy orbiting satellites, but can also injure or kill astronauts, degrade or blackout certain radio and navigation communications, and cause regional power failures by destroying critical components of electrical power grids. With the continued growth of the satellite communications industry and our growing dependence on wireless communication and instant access to global information, we are becoming more and more susceptible to problems caused by space weather.

This textbook introduces the reader to the emerging field of space weather using an approach that is both descriptive and quantitative. The mathematical sophistication of the reader is assumed to be at the level of high-school algebra. Since science is not just a collection of facts, but a process or way of understanding our natural world, the book attempts to answer the question “How do we know that?” by including discussions on the historical development of different concepts.

This book was derived from the notes for three undergraduate courses at UCLA – the first a freshman seminar, the second an Honors Collegium course, and the third a general education course for non-science majors entitled “The Perils of Space: an Introduction to Space Weather” first taught in Fall 2004.

Each chapter is divided into two parts: the main text describing space weather topics and supplements describing important physical concepts behind each topic. End-of-chapter problems allow students to delve deeper into aspects of the chapter. A list of key concepts is given at the beginning of each chapter, and the concepts are in bold in the main body of the text. Readers wishing to understand space weather should familiarize themselves with them. Definitions of important terms, which are given throughout the text, are indicated by bold page numbering in the index.

# Acknowledgments

Students in UCLA's Introduction to Space Weather course inspired the writing of this textbook. A number of friends and colleagues, including Hamid Rassoul, Chris Russell, Bob McPherron, Margy Kivelson, and Ray Walker, helped me with the book. I am especially indebted to Jeff Sanny for his very careful and thoughtful comments. The book would never have been written without the editorial assistance of Judy Hohl. Her professionalism and enthusiasm for the project made the publication possible. Also instrumental in the formulation of this textbook have been my graduate students (Matt Fillingim, Paul Martin, David Berube, Megan Cartwright, and David Galvan) and post-docs (James Weygand and Endawoke Yizengaw), who helped me combine my passion for research with my love of teaching. Mentors who supported my efforts to combine research and teaching include Jeff Hughes, James Patterson, and Michelle Thomsen. Finally, I would like to acknowledge the support and love of my wife (Patty) and children (Andi and Kyle) during the long process of putting this textbook together.

# Chapter 1
# What is space weather?

"Space weather" refers to conditions on the Sun and in the solar wind,
magnetosphere, ionosphere, and thermosphere that can influence the
performance and reliability of space-borne and ground-based technological
systems and can endanger human life or health. Adverse conditions in the
space environment can cause disruption of satellite operations,
communications, navigation, and electric power distribution grids, leading to a
variety of socioeconomic losses.

*National Space Weather Program Strategic Plan, 1995.*
*Office of the Federal Coordinator for Meteorological Services and*
*Supporting Research, FCM–P30–1995, Washington, DC.*

## 1.1 Key concepts

- space weather
- climate
- meteorology
- Earth's atmosphere

## 1.2 Introduction

In the last 50 years, we have become a spacefaring civilization. With robotic and manned spacecraft, we have started to survey our Solar System. We have learned that we live in the atmosphere of a dynamic, violent Sun that provides energy for life on Earth, but also can cause havoc among its fleet of satellite and communications systems. **Space weather** is the emerging field within the space sciences that studies how the Sun influences Earth's space environment and the technological and societal impacts of that interaction – damage to or destruction of Earth-orbiting satellites and threats to both astronaut safety during long-duration missions to the Moon and Mars and to the reliability and accuracy of global communications and navigation systems.

Modern society depends on accurate forecasts of weather (day-to-day variability of temperature, humidity, rain, etc.) and understanding of **climate** (long-term weather trends) for commerce, agriculture, transportation, energy policy, and natural disaster mitigation. The science of understanding weather, **meteorology**, is one of the oldest human endeavors to make sense of our natural environment. Like meteorology, the field of space weather seeks to understand and predict **climate** and weather, but of outer space. For millennia, space storms have raged above our heads unknown to us. But with the advent of the space age, we have begun to notice the destructive power of severe space weather.

Like weather, space weather has its roots in the Sun. The main distinctions between the two types of weather are where it takes place and the type of energy from the Sun that influences it. For weather, we are most concerned with the troposphere, which extends from Earth's surface to the top of the highest clouds at about 10 km. Space weather science is interested in the space environment around Earth all the way to the Sun. Space begins in a region of **Earth's atmosphere** called the thermosphere, which starts at roughly 100 km. The space shuttle and space station fly at an altitude of about 350 km. Plate 1 shows a picture of Earth's atmosphere from the space shuttle. The sharp contrast between the blue of Earth's atmosphere and the blackness of space is at approximately 100 km.

The second difference between weather and space weather is the type of solar energy that influences the two regions. The Sun continuously emits two main types of energy into space – electromagnetic (EM) radiation and corpuscular radiation. Visible light, radio waves, microwaves, infrared, ultraviolet, X-rays, and gamma rays are forms of EM radiation. The Sun's EM radiation bathes the top of Earth's atmosphere with about 1400 watts[1] of power per square meter and heats the lower atmosphere, surface and oceans unevenly. Winds are driven by these differences in atmospheric temperature.

The Sun also continuously emits corpuscular (minute particle) radiation, charged atoms and sub-atomic particles (mostly protons and electrons) in what is called the solar wind. Like winds on Earth, the solar wind is driven by temperature differences, but those differences are between the Sun's upper atmosphere and interplanetary space. The solar wind, which expands out into the Solar System carrying with it the Sun's magnetic field, carves out a region of interstellar space called the heliosphere ("helios", Greek for Sun).

The solar wind is not steady or uniform, but changes constantly. These changes affect Earth's space environment in a number of ways, including creation of new corpuscular radiation that bombards Earth's upper atmosphere, causing aurorae (northern and southern lights) and

---

[1] A watt is the SI unit of power (energy per time) named in honor of James Watt (1736–1819), a Scottish engineer and scientist credited with making the steam engine a practical device.

large electrical currents that can disrupt communication, power grids, and satellite navigation.

Occasionally the Sun's surface erupts and sends a large part of the solar atmosphere streaming away at high speeds. These events, called coronal mass ejections (CMEs), can contain $10^{12}$ (or 1 000 000 000 000) kg of material (equivalent to a quarter of a million aircraft carriers) and can move away from the Sun at over 1000 km $s^{-1}$ (over several million miles per hour) (Plate 2). If CMEs are directed towards Earth, a great space storm can develop far above our heads, crippling satellites, causing increased radiation exposure for airline crews and passengers, blacking out some forms of radio communication, and disrupting power systems on Earth.

These space storms, like weather storms such as Hurricane Katrina in 2005, have caused severe damage to technological systems in the past. In March 1989, a large CME slammed into Earth causing massive power outages in eastern Canada. The emerging science of space weather is attempting to understand the causes of space storms and their impact on Earth's technological infrastructure with the hope that we can forecast space weather and mitigate damage.

### 1.2.1 It's Greek to me. The origin of technical names in science

The ancient Greeks envisioned the heavens as being on concentric spheres around Earth, with the planets (Greek for the wanderers), Sun and Moon moving on their own celestial spheres, while the stars moved in lock-step behind them on their own sphere. Science borrows from this worldview by giving concentric regions in and around the planets and the Sun a Greek prefix and the suffix "sphere". The rocky surface of Earth is often called the lithosphere ("litho", meaning stone), the water part the hydrosphere ("hydro", meaning water), the place where life is found the biosphere ("bio", meaning life). The region above Earth's surface is called the atmosphere ("atmos", meaning vapors). The atmosphere is further divided into sub-regions, which are listed in Table 1.1. The boundaries between the spheres are called "pauses" (e.g., the boundary between the troposphere and stratosphere is the tropopause). Several more "spheres" and "pauses" will be introduced in the following chapters. Figure 1.1 shows the layers of the atmosphere as a function of height. Note that each layer or sphere has a different temperature profile with height. (For example, the troposphere has temperature decreasing with altitude, while the stratosphere has temperature increasing with altitude.)

Table 1.1 *Greek prefixes for regions of Earth's atmosphere*

| Prefix | English translation | Height | Characteristic |
|---|---|---|---|
| tropo | mixing or changing | 0–10 km | where weather takes place |
| strato | layer | 10–50 km | where ozone layer is located |
| meso | middle | 50–80 km | coldest region |
| thermo | heat | 80 km– | where space begins |
| iono | to go | 80 km– | where aurorae occur |

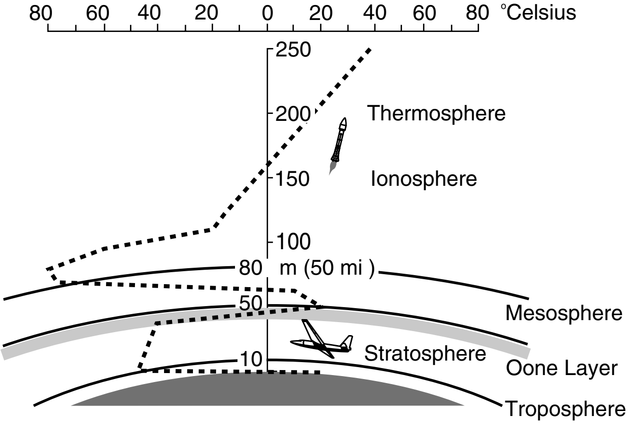


**Figure 1.1** The vertical temperature scale of Earth's atmosphere. The dashed line represents the temperature as a function of height. Each region is defined by how the temperature changes with height (courtesy of Cislunar Aerospace, Inc.).

## 1.3 Brief history

The study of space weather began with systematic observation of three natural phenomena: the aurorae (also called the northern or southern lights), Earth's magnetic field, and sunspots (dark regions observed on the surface of the Sun). Because aurorae can be seen with the unaided eye, they have been observed for thousands of years, though the systematic study of the aurorae didn't begin until the sixteenth century. Development of the sensitive compass and telescope in the early seventeenth century made possible the discovery of the nature of Earth's magnetic field and sunspots.

The understanding of space weather traces its roots to connections between these three phenomena. The first tentative connections were made in the middle of the nineteenth century. For the last 150 years, we have slowly expanded our knowledge of the Sun and Earth's space environments and, in so doing, have begun to develop a physical model of the Sun–Earth connection. This section gives a brief history of the discoveries and an introduction to some of the scientists who have led us to our current understanding of the solar–terrestrial relationship. As with all areas of science, the field of space weather developed in concert with our understanding of physics and chemistry and new technologies that allowed us to "see" the "invisible" – things too small or far away to be seen with the unaided eye or beyond our sensibilities and capabilities to see, hear, or feel, such as radio waves and magnetic fields. A website has been developed to provide a detailed timeline of our understanding of space weather. The web address is given in Appendix A.

### 1.3.1 The aurorae

Our earliest ancestors observed the aurorae. Until the eighteenth century most treatises on aurorae were based on speculation about their origin by men who may have never observed them. These speculations usually followed Aristotle's (384–322 BC) view of aurorae as burning flames, or René Descartes' (1596–1650) idea that aurorae were moon- or sunlight reflected off ice or snow crystals. Systematic observations of aurorae were first made in the sixteenth century. One of the greatest astronomers of all time, Tycho Brahe (1546–1601), recorded the occurrence of aurorae between 1582 and 1598 from his Uraniborg observatory in Denmark. He found that the number of aurorae varies from year to year, but did not note any systematic or regular variation. On September 12, 1621, the astronomer Pierre Gassendi (1592–1655) from the south of France and Galileo[2] in Venice observed the same aurorae. Gassendi called the lights aurorae borealis (Latin for northern dawn), a name that has been associated with polar lights ever since. He noted that the aurorae must occur high in Earth's atmosphere for observers at distant locations to be able to observe the same phenomena.

In the eighteenth century a number of observations began to illuminate the origins of aurorae. Frenchman Jean-Jacques d'Ortour de Mairan (1678–1771) made the first rough measurements of auroral height in 1726; these were consistent with Gassendi's observation that aurorae occur in the upper atmosphere. Using the triangulation method, English scientist Henry Cavendish (1731–1810) correctly estimated auroral height to be between 80 and 112 km in 1790. However, estimates of auroral height continued to have large uncertainties until around 1900, when Norwegian scientist Carl Størmer (1874–1957) measured the height accurately using photographic techniques.

[2] Galilei, Galileo (1564–1643), Italian physicist and astronomer and founder of the modern scientific method. The first to use a telescope for astronomical observations, he discovered the moons of Jupiter (named the Galilean moons in his honor); that Venus has phases, which offered direct support for the Copernican heliocentric theory; that the Milky Way is made up of individual stars; and that the Moon has mountains.

Captain James Cook was the first European to observe the southern lights (which he called aurora australis) while in the Indian Ocean near latitude 58°S on February 17, 1773. He wrote in his ship's log:

> lights were seen in the heavens, similar to those in the northern hemisphere, known by the name of Aurora Borealis.

In the nineteenth century, as reports from polar explorers were compiled, it became clear that aurorae appear in large ovals centered near the North and South Poles. Captain John Franklin, who later perished with his crew as they attempted to find the Northwest Passage, determined that the number of auroral sightings decreases nearer the Pole, suggesting an auroral zone. In 1833, the German geographer Georg Wilhelm Muncke (1772–1847) noted the existence of a zone of maximum auroral occurrence that is limited in latitude. In 1860, Professor Elias Loomis (1811–1888) of Yale University published the first map of the north polar region showing the zone where aurorae had been most commonly observed (see Figure 1.2).

So, by the middle of the nineteenth century a number of facts about aurorae were known: they occur in an oval around the north and south polar regions, and they are high in the upper atmosphere. The search was still on for the cause of the aurorae.

### 1.3.2 The geomagnetic field

In 1088, Chinese encyclopedist Shen Kua (1031–1095) wrote the first description of compasses, magnetized pins floated on a small cork in a bowl of water. Alexander Neckham of St. Albans (1157–1217) was the first European to describe a compass in his work *On the Nature of Things* published in 1187. Neckham had probably heard of the Chinese compass through the silk-road trade routes from China to Western Europe. In 1576, Robert Norman discovered that Earth's magnetic field has a vertical component called dip. Combining this discovery with his own work with a model magnetic field called a terrella, William Gilbert[3] (later the personal physician of Queen Elizabeth I) wrote a book called *De Magnete* in 1600. In this book he demonstrated that Earth's magnetic field behaves like a magnet, which led to the systematic study of magnetic field orientation as a function of position on Earth. These magnetic maps allowed the use of compasses for navigation. In 1722, George Graham (1674?–1751) built a compass sensitive enough to observe slight (usually

---

[3] Gilbert, William (1544–1603), English physicist and physician who pioneered the field of geomagnetism. The first English scientist to accept the Copernican view of the Solar System, he suggested that habitable worlds might be in orbit around other stars.

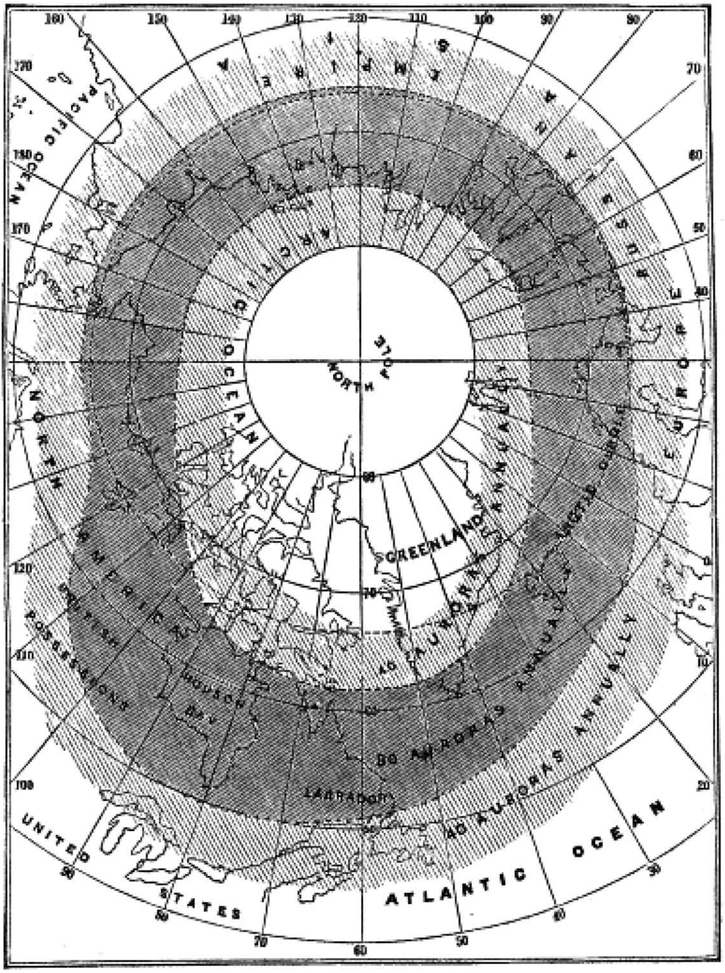


**Figure 1.2** The auroral oval from Professor Loomis' late nineteenth century study. Note that the aurora has a zone of occurrence centered around, but not at, the pole (from Loomis, 1869).

less than 1°) irregular variations of the geomagnetic field that caused the compass needle to "wriggle" slightly.

Thus the basic fact about geomagnetism known by the early eighteenth century was that Earth has a magnetic field like that of a regular magnet (called a dipole magnetic field) with both regular and irregular variations. The search was on for the cause of these geomagnetic fluctuations.

### 1.3.3 Sunspots

In 1610, Galileo turned his telescope to the Sun and observed sunspots by focusing the image onto a piece of paper (see Figure 1.3). Several other observers – Johannes Fabricius, Thomas Harriot and Christoph

Scheiner – essentially simultaneously observed sunspots with the newly developed telescope. But what were they? Scheiner argued that they were moons or planets (Mercury, Venus, or the mythical planet Vulcan) orbiting between the Sun and Earth, while Galileo argued that they were on the surface of the Sun. The first regular daily (when weather permitted) observations of sunspots began in 1749 at the Zurich Observatory in Switzerland. Using data from Zurich, Samuel Heinrich Schwabe (1789–1875) recognized the occurrence of an 11-year solar cycle in about 1844. His original goal had been to find intra-mercurial planets such as those conjectured in Galileo's time. He began to systematically look for "transits" of these hypothesized planets across the Sun. In so doing, he meticulously recorded the position of every sunspot for 18 years. With this data set he discovered the 11-year sunspot cycle. He never did discover a planet crossing the Sun.

So by the mid-nineteenth century, it had been clearly demonstrated that sunspots exist on the Sun and they have an 11-year cycle during which their number waxes and wanes. The question of what sunspots are and what if any effect they have on Earth still remained.

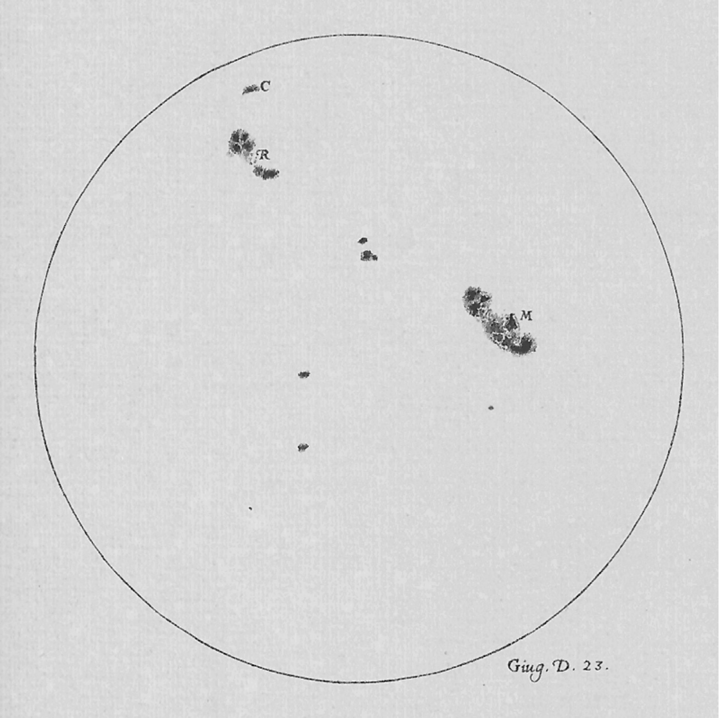


**Figure 1.3** Sunspot drawings from Galileo made in about 1610 (from Galileo Galilei, 1613, courtesy of Owen Gingerich).

### 1.3.4 Making the connection between aurorae, the geomagnetic field, and sunspots

English astronomer Edmond Halley (1656–1742), of Halley's Comet fame, noted that the aurorae that occurred on March 16, 1716 over London appeared to have rays converging toward Earth similar to Earth's magnetic field lines. He drew the magnetic field lines outside Earth by extrapolating from the shape that iron filings make around a magnet. This hint that auroral rays are aligned with Earth's magnetic field was not confirmed until 1770 when the Swedish scientist Johann Wilke observed that auroral rays actually lie along the geomagnetic field lines. In 1722, the British instrument maker George Graham observed slight magnetic fluctuations with his compass, which later were observed to be correlated with observations of aurorae by Anders Celsius (the scientist whose name was given to a temperature scale) and his student (and brother-in-law) Olaf Hiorter in Uppsala, Sweden in 1747. Professor Celsius' instrument was obtained from Graham, and through regular correspondence Celsius and Graham found that days with geomagnetic activity (the name given to magnetic fluctuations) in London were also days with geomagnetic activity in Uppsala. This established that geomagnetic activity and hence aurorae occur over large distances. Hiorter wrote "That aurorae must be the highest phenomena of our atmosphere, so high and extensive, that they can simultaneously, here and in England, at Uppsala and London, . . . disturb the magnetic needle." They and others then began to observe periods of very large geomagnetic fluctuations – many degrees of magnetic needle fluctuation in several minutes. These large geomagnetic disturbances are now called geomagnetic storms.

In the mid-nineteenth century, Col. Edward Sabine and Prof. Rudolf Wolf independently were the first to publish results showing the correlation between sunspots and geomagnetic activity. Sabine was a British military officer (later knighted) in charge of British observatories around the world at which surface weather and changes in the geomagnetic field were monitored. Understanding the geomagnetic field was important for navigation because of the use of magnetic compasses. Rudolf Wolf was director of the Zurich Observatory and therefore had access to the longest record of sunspot occurrence in the world. The studies of Graham and Celsius now linked aurorae with geomagnetic activity, and the studies of Sabine and Wolf linked geomagnetic activity with solar activity. Because of the unreliable reporting of the occurrence of aurorae (summer time, clouds, limited populations at high latitudes), it was several years before the occurrence of aurorae was also clearly shown to be linked with solar activity.

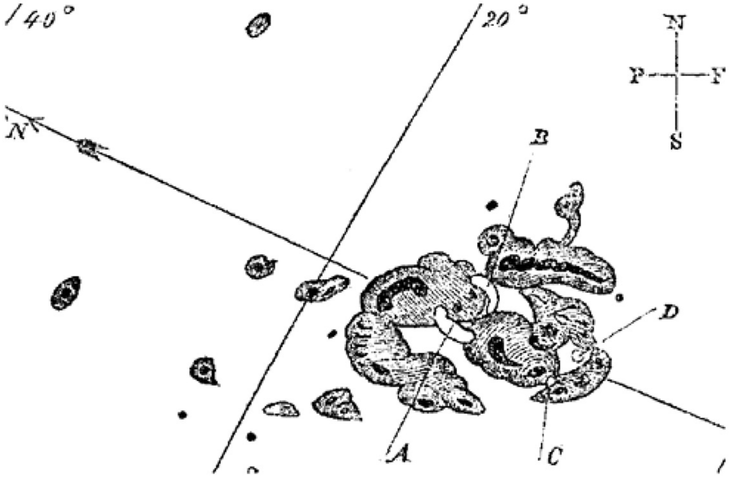


**Figure 1.4** Carrington's 1859 drawing of a white light solar flare associated with a sunspot group (from Carrington, 1860).

Immediately the question was raised: what connected sunspots with Earth's magnetic field? In 1859 when Richard Carrington, who worked as an astronomer at Greenwich Observatory, observed a white light flare above a sunspot, he remarked in a letter to the British Royal Society that the flare was followed within a day by a geomagnetic disturbance. (Figure 1.4 shows Carrington's drawing of the flare.) In this letter he suggested that perhaps there was a causal relationship. However, in 1859 it was assumed that space was a complete vacuum and the discovery of electrons and the other sub-atomic particles was 40 years away. In fact, in 1859 there was still not a clear understanding of electromagnetic radiation (such as light). In 1863 Lord Kelvin,[4] one of the dominant physicists of the nineteenth century, cast strong doubt on the connection between sunspots and geomagnetic activity after calculating that the solar magnetic field could not possibly impact Earth's magnetic field over the tremendous distance between the Sun and Earth. He later expressed these thoughts in his Presidential Address to the Royal Society in 1892 as: "It seems as if we may also be forced to conclude that the supposed connexion [sic – old British spelling] between magnetic storms and Sun-spots is unreal, and that the seeming agreement between periods has been mere coincidence."

---

[4] Kelvin, William Thomson (1824–1907), Irish-born Scottish physicist and one of the greatest scientists of the nineteenth century. His work included studies in thermodynamics and electricity and magnetism. He was also an entrepreneur and became wealthy with inventions that made the first transatlantic telegraph successful. Proposed the absolute temperature scale, whose degree is now called the kelvin (K) in his honor.

Despite this opinion, several discoveries suggested that perhaps matter from the Sun could travel to Earth. In 1869, J. Norman Lockyer developed the solar spectrograph and observed prominences (large loops of plasma) reaching high above the solar atmosphere. The following year Charles Young took the first photographs of a solar prominence. Observations of both solar prominences and total solar eclipses, which showed that the upper atmosphere of the Sun extends away from the solar surface, led scientists to speculate that solar material could be ejected into interplanetary space. Over the next several decades, as more observations of activity on the Sun were correlated with geomagnetic activity, several scientists including Henri Becquerel in 1878 (who later won a Nobel prize for his discovery of radioactivity) suggested that something (either magnetic fields or matter) occasionally expelled from the Sun impinged upon Earth's upper atmosphere, causing currents to flow and giving rise to geomagnetic activity. As the nineteenth century was drawing to a close, a number of important concepts in fundamental physics that helped in solving the solar–terrestrial puzzle were being developed. These included the realizations that electricity and magnetism are connected, that light is electromagnetic radiation, and the discovery of corpuscular radiation. The discovery of corpuscular radiation in the 1880s by the English physicist William Crookes spurred a number of scientists, among them the Irish mathematician FitzGerald in 1892 and the English scientist Sir Oliver Lodge in 1900, to suggest that this type of radiation was responsible for geomagnetic activity. At Cambridge University in 1897, J. J. Thomson (1856–1940) showed that corpuscular radiation is made up of sub-atomic particles called electrons, the discovery of which won him a Nobel Prize. In addition, wireless radio was coming on the scene, and in 1901 the Italian inventor Guglielmo Marconi[5] launched the first transatlantic radio message (which earned him a Nobel Prize). The puzzle of how electromagnetic radiation could propagate around the curved Earth inspired scientists such as the English physicist Oliver Heaviside[6] and the Irish physicist Arthur Kennelly (1861–1939) to theorize that a layer of conducting gas in the upper atmosphere acted to reflect Marconi's radio waves. This theory fitted the idea that currents flowing in the upper atmosphere are

---

[5] Marconi, Guglielmo (1874–1937), Italian electrical engineer and inventor who was one of the first to appreciate the potential of radio waves for communication and was the first to send a transoceanic radio message. He developed the first microwave receiver in 1932, which ushered in the development of radar.

[6] Heaviside, Oliver (1850–1925), English physicist and electrical engineer who made seminal contributions to the understanding of electric circuits by developing the concepts of inductance, capacitance, and impedance. Re-wrote Maxwell's equations of electricity and magnetism in the vector calculus notation used today.

associated with aurorae. The existence of a conducting layer in the upper atmosphere was soon experimentally verified by the English physicist Edward Appleton[7] (who won a Nobel Prize for his work). The region of the upper atmosphere associated with current flow and the aurorae was named the ionosphere.

Shortly after the discovery of the ionosphere, British geoscientist Sydney Chapman and his student Vincenzo Ferraro published a paper suggesting that streams of particles from the Sun that hit Earth cause geomagnetic storms. Twenty years later, German astronomer Ludwig Biermann[8] suggested that the Sun continuously emits a gas (now called the solar wind) in order to explain the observation that comet tails always point away from the Sun. Around this time, Swedish physicist Hannes Alfvèn[9] proposed the existence of electric and magnetic waves defined by the motion of a magnetized plasma. These waves, now called Alfvèn waves, explained how energy and momentum propagate in plasma such as the solar wind (Alfvèn won a Nobel Prize in 1972 for this work). Eugene Parker (1927–), from the University of Chicago, then developed a theory of the solar wind that explained how the solar atmosphere could continuously expand out into interplanetary space taking with it the solar magnetic field. At Imperial College in London in 1961, James Dungey (1923–) combined these ideas by proposing that the solar wind's magnetic field, called the interplanetary magnetic field (IMF), could connect or merge with Earth's magnetosphere and couple solar wind energy and momentum directly into Earth's magnetosphere. The Australian solar physicist Ronald Giovanelli (1915–1984) had suggested the concept of magnetic field merging, or reconnection, ten years earlier in an attempt to explain where the energy came from to power solar flares (like the one observed by Carrington in 1859). These ideas were to be tested over the next 40 years during the Space Age, when satellites could be sent directly into space to measure what was there. The launch of Sputnik in 1957 ushered in the dawn of the Space Age and not only brought us understanding of what was in space, but also led to the technological revolution of global communication and Earth observing that we take for granted today. With the advent of satellite and

---

[7] Appleton, Edward Victor (1892–1965), atmospheric physicist who played important roles in the development of radar during World War II. He was knighted in 1941 for his contributions in understanding Earth's ionosphere.

[8] Biermann, Ludwig (1907–1986), German astrophysicist who made important contributions to the understanding of stellar interiors and the understanding of the role of magnetized plasmas in the Solar System and Galaxy.

[9] Alfvèn, Hannes (1908–1995), Swedish space physicist who founded the sub-field of plasma physics called magnetohydrodynamics with his discovery of hydromagnetic waves. These are now called Alfvèn waves in his honor.

cable TV, Global Positioning System (GPS) navigation, and continental power grid systems we entered an age in which the turmoil of space began to affect our everyday lives.

## 1.4 Impacts of space weather on society

Over 500 operational satellites currently orbit Earth. Many of these are commercial communications satellites that provide global news TV coverage, telephone connections, and credit card transactions (next time you visit a gas station that has "pay-at-the-pump" credit card readers, look at the roof of the gas station. You will probably see a satellite dish that beams your credit card information nearly instantly to your bank to verify your credit.) Governments operate many other satellites to provide weather images, navigational signals, land use information, and military surveillance. All are susceptible to damage and degradation due to the harsh space environment.

Many other systems, including airline crews and passengers, pipelines, and electric power grids, are susceptible to space weather effects, as well. Though space weather effects have been observed since the first telegraph lines in the mid-nineteenth century, it was not until the last decade or so that scientists began studying the problem in earnest. The new interest in space weather is primarily due to the rapid growth of the commercial satellite communications industry and the development of continental-sized power and communication grids. With these developments, we have become more susceptible to space weather storms through our reliance on high-tech information systems and our growing global interconnections. This book describes space weather causes, effects, and impacts on society. With time, space weather will play a larger role in our everyday lives. Perhaps in the not too distant future, the newspaper will carry a weather forecast and a space weather forecast in order to help you plan your next visit to the orbiting Hilton Hotel or the new resort on the Sea of Tranquility to visit the Apollo 11 landing site on the Moon.

## 1.5 Supplements

> Measure what is measurable, and make measurable what is not so.
>
> Galileo Galilei

To have a physical understanding of nature, including the environments of the Sun and upper atmosphere of Earth, you must comprehend a number of over-arching concepts of physics. Among these are energy and force, which are at the heart of much of physics and help us understand the interconnections and cause-and-effect relationships of the world around us.

In addition to understanding these fundamental concepts of physics, it is also important to learn the language of physics in order to communicate the value of an observed quantity. One could use qualitative terms, such as “fast”, “slow”, “heavy”, “light”, etc., as descriptors, but to truly understand an object, you need to know its speed or mass quantitatively. Scientists all over the world use a special set of units called the Système International d’Unites (SI), which allows them to communicate easily. You may know SI as MKS (meters–kilograms–seconds), though the American public still uses “English” units, such as yards, pounds and seconds. An advantage of the SI (or metric) system is that it is a base ten system (all the units are evenly divisible by 10). In addition, there is a physical relationship between the fundamental units[10] of length, mass, and time. In SI these units are measured in meters, kilograms, and seconds. The relationship between them involves a volume of water and a swinging pendulum (though the latter is only an approximation, and the history of the definition of the meter and second is long and full of twists and turns). The mass of a cubic centimeter of water at standard temperature and pressure (essentially room temperature conditions) is equal to exactly 1 gram. One half the period of a swinging pendulum with a length of one meter is almost exactly one second. Today a second is defined by counting the oscillations of a cesium atom and not a swinging pendulum. The development of atomic clocks has made possible a wide range of technologies including satellite navigation. The developers of this technique (Roy J. Galuber, John Hall, and Theodor W. Hansch) were awarded the 2005 Nobel Prize in Physics.

### 1.5.1 SI units

The base unit of distance is the meter (slightly longer than a yard). Fractions of meters divisible by ten are given prefixes such as deci (1/10th), centi (1/100th), or milli (1/1000th). For distances that are relevant to those across the surface of Earth, kilometers (km) are often used. This nomenclature is used for all units so you can have a gram or kilogram or milligram. We often measure time in hours, minutes, and seconds. Therefore, when using time, care must be used in making sure that you use one type of time unit (years, days, hours, minutes, or seconds) and not intermix them.

---

[10] Four units (length, mass, time, and electric charge) are generally considered as fundamental. Nearly all other units can be written in terms of these fundamental units and are therefore called “derived units”.

All of the other fundamental parameters of physics (such as velocity, acceleration, force, and energy) use combinations of the base SI units. For example, velocity (or more properly speed when referring to the magnitude of velocity) is calculated by dividing distance by time ($v = d/t$). Therefore the units of velocity are the units of distance (m) divided by time (s) or meters per second (m/s). A list of useful SI units and the equivalent base units is given in Appendix B. Two other fundamental units that are used in this book are for electric charge (coulomb in SI) and temperature (kelvin in SI).

Example: how fast would a CME need to be traveling if it took three days to reach Earth from the Sun?

$$\begin{aligned}
\text{Velocity} &= \frac{\text{distance}}{\text{time}} \\
&= \frac{\text{distance between Sun and Earth}}{3\text{ days}} \\
&= \frac{150\,000\,000\text{ km}}{3\text{ days}} \\
&= \frac{150\,000\,000\text{ km}}{3\text{ days} \times 24\text{ hrs/day}} \\
&= 2\,000\,000\text{ km hr}^{-1}.
\end{aligned}$$

Note that days were converted to hours by multiplying the number of days by the number of hours per day.

## 1.5.2 Scientific notation

We often measure very large or very small things. For example, the average distance between the Sun and Earth is about 150 000 000 km. Scientists often do two things to make large or small numbers more manageable. The first is to define a new unit. In the case of the distance between Earth and the Sun, we define this distance as 1 astronomical unit (AU) (1 AU = 150 000 000 km). This unit is useful for describing the distance between the planets (i.e., Mars is 1.5 AU from the Sun, Jupiter is 5 AU, etc.).

Another way of easily handling large or small numbers is to use scientific notation. This is a way to write multiples of 10 using exponents. It is simply a way of keeping track of all the places to the right or left of the decimal point. For example, 1000 can be written as $10^3$, or 150 000 000 km can be written as $1.5 \times 10^8$ km. For numbers less than 1, the exponent is negative (i.e., 1/1000th = 0.001 = $10^{-3}$). What makes scientific notation useful (besides providing a more compact way to write large or small numbers) is that multiplying and dividing are as simple as adding and subtracting. To multiply numbers together that are written in scientific notation, add all the exponents together. For example, $10^4 \times 10^4 = 10^{4+4} = 10^8$. For division, simply subtract the exponents ($10^4 \div 10^6 = 10^{4-6} = 10^{-2}$).

Example: the radius of the Sun is $7 \times 10^5$ km. The radius of Earth is approximately $7 \times 10^3$ km. How many Earths could fit across the face of the Sun?

Answer: the answer can be found by calculating how many Earth radii it takes to equal one solar radius.

$x \times 1r_{\mathrm{E}} = 1R_{\mathrm{solar}}$

$$x = \frac{1R_{\mathrm{solar}}}{1r_{\mathrm{E}}} = \frac{7 \times 10^5\ \mathrm{km}}{7 \times 10^3\ \mathrm{km}} = \frac{7}{7} \times 10^{5-3} = 1 \times 10^2.$$

Therefore, $10^2$ or 100 Earths can fit across the Sun.

## 1.6 Problems

1.1 Have you used a satellite recently? Think about how you have used information from a satellite or gained access to information using a satellite and write a short paragraph describing this use.

1.2 What is the name of the boundary that separates Earth's magnetosphere from the solar wind? How would you tell when you crossed this boundary?

1.3 A satellite TV company spends US$400 million (a typical cost in 2004) to build and launch a communications satellite. If the satellite has an expected lifetime of six years, how many satellite TV subscribers must the company attract if each subscriber pays $30 per month, in order to break even? (In 2004, DirecTV had over 13 million customers. Based on your calculation, is satellite TV profitable?)

1.4 What is the distance from the Sun to Earth in terms of solar radii? Earth radii?

1.5 How long (in days) does it take a parcel of solar wind traveling at 400 km $\mathrm{s}^{-1}$ to reach each planet? (Mercury = 0.4 AU, Venus = 0.7 AU, Earth = 1.0 AU, Mars = 1.5 AU, Jupiter = 5 AU, Saturn = 10 AU, Uranus = 20 AU, Neptune = 30 AU, Pluto = 40 AU).

1.6 How does the development of new observational instruments contribute to our understanding of our natural world?

# Chapter 2
# The variable Sun

The spots do not remain stationary upon the body of the sun, but appear to move in relation to it with regular motions.

> *Galileo Galilei in* Letter to Mark Welser*, 1613. These "Sunspot Letters" were one of the first written scientific discussions of sunspots.*
> Discoveries and Opinions of Galileo *translated with an Introduction and Notes by Stillman Drake, Anchor Books, 1957.*

## 2.1 Key concepts

- electromagnetic radiation
- heat transfer
- Standard Solar Model
- solar atmosphere
- solar cycle

## 2.2 Introduction

Since the dawn of man, the Sun has elicited worship, inspiration, and study, and tremendous mysteries about its dynamics still occupy the attention of solar astronomers and space physicists. This ignorance has profound implications. We now rely on space for global communication, navigation, and Earth observing, and solar dynamics cause degradation and failure of satellites and space instruments. Understanding solar dynamics is a key part of understanding space weather. This chapter describes what we know about the Sun and how we know it. Much of our knowledge comes from observations of the Sun, and much of it comes from applying the laws of physics (such as thermodynamics and nuclear and atomic physics) within quantitative models to make predictions of observable quantities. This is how we know what goes on inside the core of the Sun without going there or observing it directly. This combination of observation and physics has allowed us to know more about our natural surroundings than at any other time in human history. What we are finding out is that we know very little of how things work. But what we do know (and how we have come to know it) is truly amazing.

The Sun is one of an estimated 100 billion stars in the Milky Way galaxy. We know this by extrapolating surveys of stars in this galaxy and others. From studying the properties of nearby stars and the age of meteorites, we have learned that the Sun is a typical star about 4.5 billion years old. By observing star formation regions within our galaxy, we know that the Sun was formed out of a giant cloud of gas and dust called the solar nebula. An area within this nebula contained slightly more material than its surroundings and therefore gathered material to itself because of its gravitational pull (a self-sustaining process called accretion). As the proto-Sun (proto, Greek for "first") got bigger and bigger, its increasing mass strengthened its gravitational attraction, pulling more material to it. As the gas and dust were gathered into the proto-Sun, gravity forced it closer and closer together, increasing its density (amount of material in a given volume). The pressure inside the proto-Sun therefore increased since the pressure of a gas is dependent on its density from the ideal gas law (this law can be expressed as $P = nkT$ where $P$ is pressure, $n$ is number density, $k$ is Boltzmann's constant,[1] and $T$ is temperature). As more material was accreted onto the proto-Sun, its density and temperature continued to increase until they reached a critical temperature at which thermonuclear fusion can occur (about 15 million K). At this point the Sun was born. The star becomes stable when there is a balance between the force of gravity pulling material into it and the force due to the pressure of the gas pushing out. This is called hydrostatic equilibrium (see Chapter 5 for further discussion of this concept). The Sun's energy comes from thermonuclear reactions in the core that fuse protons together to form helium nuclei. In the process, energy is liberated; some of this energy eventually makes its way to the surface and propagates out into space as **electromagnetic radiation**.

Because of the high temperatures on and within the Sun, the gas is ionized. An atom or molecule becomes ionized when an electron is knocked out, giving the atom a net positive electric charge. The charged particles feel the force of electric and magnetic fields. In addition, moving charged particles can create an electrical current, which in turn gives rise to a magnetic field. Therefore, due to the motion of ionized gas in the Sun, a strong magnetic field is generated. Changes in this highly variable solar magnetic field cause changes in the amount of energy released from the Sun's surface. The Sun is a dynamic and variable star because

---

[1] The Boltzmann constant was named in honor of Ludwig Boltzmann (1844–1906), an Austrian theoretical physicist who made significant contributions to thermodynamics, electricity, and magnetism and is credited with developing the foundation of statistical physics through his insights into the kinetic theory of gases. His relationship describing the mean total energy of a molecule to its temperature includes a constant $k$, now known as the Boltzmann constant (equal to $1.380\,66 \times 10^{-23}$ joules per kelvin.)

of these moving charged particles and the magnetic field they create. The structure and processes of the Sun that give rise to solar variability are examined in this chapter.

Before discussing the atmosphere of the Sun, a few important physical concepts to help explain the Sun's structure are introduced.

## 2.3 Temperature and heat

The temperature of an object, gas, or liquid describes the thermal motion of the atoms and molecules that make up the object. Atoms or molecules in hot things have large thermal velocities, and atoms or molecules in cold things have small thermal velocities. A low-temperature gas or liquid has slowly moving molecules; a high-temperature gas or liquid has quickly moving molecules. Put a tea bag (or a drop of food coloring) in hot water and then another in cold water. Does the tea (or food coloring) spread more quickly in hot water or cold water? What does this say about the motion of the individual water molecules in the cup?

When objects with different temperatures are placed in contact with each other, the temperature of both objects changes. The hotter object cools while the colder object warms. The temperatures change until they are in equilibrium or equal. We call the transfer of energy from a hot object to a cold object heat. Heat has the units of energy (joules in SI) and the transfer of heat from one place to another drives weather on Earth and is a fundamental aspect of space weather.

## 2.4 Radiation and convection

There are three forms of **heat transfer**: conduction, convection, and radiation. Conduction is the transfer of heat in the absence of fluid motions. This occurs when two solid objects are brought into contact (an electric stove top and a frying pan, for example). The hot stove top's heat flows from the electric coils into the metal frying pan, heating the frying pan.

Convection is the transfer of heat by fluid motions. Fluid is a generic name given to everything that flows easily, and fluids can include gases. An example of convection is how heat is transferred through a pot of boiling water. The water near the bottom in contact with the hot pot rises because it is hotter and hence less dense; the cooler water sinks because it is more dense. If you watch a pot of water boil, you will see convection cells set up with the hot water rising at the center of the pot and sinking at the sides.

Radiation is the transfer of heat through electromagnetic radiation. The Sun warms Earth's surface when sunlight is absorbed by Earth. The temperature of Earth depends on the amount of sunlight absorbed by Earth and its atmosphere.

These last two processes (convection and radiation) are responsible for the transfer of heat from the core of the Sun into outer space. The heat from the Sun's core is transferred towards its surface in the form of electromagnetic radiation. At a certain point inside the Sun, fluid flow can effectively start transferring that heat via convection. The gas is heated from below and rises, reaches the surface, emits radiation (and hence heat) out into space, and cools and sinks.

## 2.5 Solar structure

We can make direct observations of the solar surface and atmosphere. The primary means of study of these regions is analysis of absorption lines in the solar spectrum. By studying these lines, we know the composition of the Sun to very high accuracy. Table 2.1 lists the five most common elements in the Sun and their relative abundance. Hydrogen is by far the most common, followed by helium. The **Standard Solar Model** suggests that these abundances are representative throughout the Sun except in the core where thermonuclear reactions continuously change the composition. The Sun contains all of the natural elements found on Earth and in the periodic table. In fact, essentially all of the elements beyond hydrogen and helium on Earth (and in our bodies) were made inside a now-dead star, whose remains made up the original solar nebula. Literally, we are made of star-stuff.

The Sun contains $1.9 \times 10^{30}$ kg of material – over 99% of the total mass in the Solar System – or about 300 000 Earth masses. The regions of the Sun are illustrated in Figure 2.1. Table 2.2 shows some of the Sun's physical characteristics.

Table 2.1 *The five most common elements in the Sun*

| Element | Symbol | Relative abundance |
|---|---|---|
| hydrogen | H | 92.1% |
| helium | He | 7.8% |
| oxygen | O | 0.061% |
| carbon | C | 0.030% |
| nitrogen | N | 0.0084% |

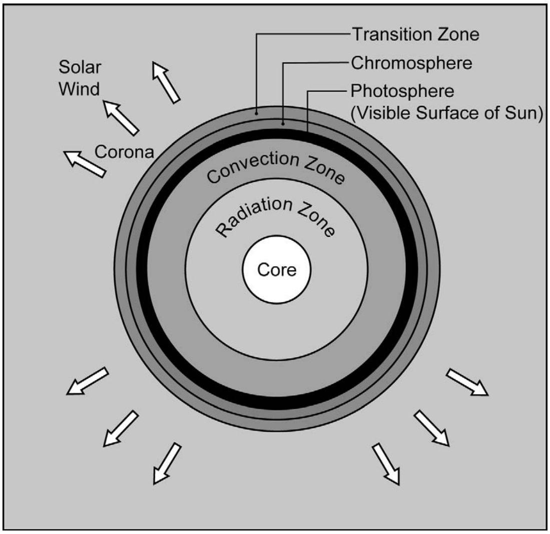


**Figure 2.1** The main regions of the Sun. The regions inside the Sun are defined by how energy is transferred from the core to the surface. The regions of the Sun's atmosphere are defined by their density and temperature.

Table 2.2 *Solar properties*

| | |
|---|---|
| radius | 696 000 km |
| mass | $1.9 \times 10^{30}$ kg |
| average density | 1410 kg $m^{-3}$ |
| distance from Earth | 150 000 000 km (or 1 astronomical unit) |
| surface temperature | 5800 K |
| luminosity | $3.86 \times 10^{26}$ W |

## 2.5.1 Interior

The Sun's interior is divided into three main regions: the convection zone, the radiation zone, and the core. The next sections describe these regions.

### *Convection and radiation zones*

Below the photosphere, the visible surface of the Sun, extending down about 200 000 km, is the convection zone. This region undergoes convective motion (hot gas rising and cooler gas sinking) similar to a pot of boiling water. Convection is the process that transports heat through bulk fluid motion. Energy from the outer layer of the Sun is transported by convection to the solar surface where it can radiate out into space. Below the convection region is the radiation zone, where energy is transported primarily by electromagnetic photons. The convection zone begins where the flux of radiation energy is so high that the energy cannot easily make its way through the gas, hence convection begins in order to transport the energy through bulk motions of the solar material.

How do we know what goes on inside the Sun? As previously mentioned, we have developed a solar model (the Standard Solar Model) based on our best understanding of the physics that takes place there. The model is compared to observations that we can make – including direct observations of the solar surface – which allows measurement of in-and-out surface motions. These radial motions or oscillations are observed through periodic changes in the Doppler[2] shifts of spectral lines of the gases in the photosphere and chromosphere. The observed spectral frequency (or color for visible light) is shifted toward the blue for gas that is moving to the observer and toward the red for gas moving away (see supplement for details). Solar surface motions are due to sound waves from the interior of the Sun. Different regions of the Sun refract (bend) or reflect these sound waves and confine the motion of the waves to specific regions. Observers can then deduce the temperature and density structure of the interior from the behavior (frequency) of these oscillations. This is similar to how we can use seismic waves to deduce the structure of Earth's interior. Because of these similarities, the field of using sound waves to study the interior of the Sun is called helioseismology.

### *The core*

Thermonuclear reactions occur in the Sun's core, which is about 200 000 km across. These reactions power the Sun and release tremendous amounts of energy every second. We can measure the Sun's total energy output by placing a device – such as a solar cell – above Earth's atmosphere perpendicular to the Sun's rays and determining how much energy is intercepted per square meter every second. This quantity of energy, called the solar constant, is about 1400 W $m^{-2}$. An average California residential home uses 1600 W,[3] so the energy intercepted by a little more than one square meter could provide the electrical energy of an average household.

---

[2] Doppler, Christian Johann (1803–1853), Austrian physicist who discovered that the observed frequency of a sound wave depends on the relative velocity of the source and observer. This is called the Doppler effect, and it applies to electromagnetic radiation (like light) waves, as well, which allows the remote sensing of the solar surface's radial velocity.

[3] Residential energy use is metered and billed in kilowatt-hours (kWh) or the number of kilowatts used each hour. A typical California home uses 5914 kWh or an average of 1600 W. Of course not all the energy from the Sun makes it through Earth's atmosphere. Some of it is absorbed or reflected by the atmosphere and clouds, so on average only 30–50% of the energy makes it to Earth's surface.

Now that we know how much solar energy passes through a square meter above Earth's atmosphere per second, we can calculate the total amount of energy emitted by the Sun each second. This quantity, called luminosity, has units of power. Assuming that the Sun radiates energy uniformly in all directions, we can imagine that the amount of energy intercepted at one square meter above Earth's atmosphere is the same amount of energy intercepted at any square meter of an imaginary sphere that has at its center the Sun and its radius is the distance from the Sun to Earth. Now all we have to do is add up all the square meters that cover that surface of that sphere. The surface area of a sphere is $4\pi r^2$. Therefore the total amount of energy emitted by the Sun per second is the solar constant times the area of a sphere whose radius is the distance between the Sun and Earth ($1400\ \mathrm{W\ m^{-2}} \times 4\pi\ (1\ \mathrm{AU})^2$), approximately $4 \times 10^{26}$ W. Although the luminosity of the Sun is typical of the other stars you see in the sky, the Sun appears much brighter (and bigger) because it is much closer to us than even our nearest stellar neighbor. To put its luminosity in perspective, the Sun emits the same amount of energy in one second that Earth would produce at the current rate in over 900 000 years. Put another way, the Sun emits the equivalent of 100 billion one-megaton nuclear bombs every second. Needless to say, nothing on Earth compares to the vast amounts of energy produced each second by the Sun.

A natural question that had puzzled scientists until the twentieth century is, where does the Sun get this energy? One of the biggest contributions to our understanding of the Sun's energy source was Einstein's[4] famous equation, $E = mc^2$, which tells us that energy equals mass times the speed of light squared or, physically, that mass and energy are intimately related. The speed of light is itself a very large number – $3 \times 10^8\ \mathrm{m\ s^{-1}}$ –, and this number squared is huge. Therefore, even a small amount of mass is equivalent to a large amount of energy. Einstein's simple formula led to the development of the concept of nuclear fusion, now known to be the only energy-generating mechanism capable of producing the enormous amounts of energy emitted by the Sun. In this process light nuclei are fused into heavier nuclei. High temperatures in the Sun's core strip electrons from nuclei of atoms (mostly hydrogen) so that there are essentially protons (or the nuclei of hydrogen) whipping around the core. Occasionally two protons collide, which starts a process called a proton–proton chain that eventually leads to one helium nucleus. Helium contains two protons and two neutrons and therefore has an atomic mass of four. It is made by fusing together four protons. If

---

[4] Einstein, Albert (1879–1955), German-born American Nobel-Prize-winning physicist whose work revolutionized our understanding of energy and matter.

you compare the mass of four protons with the mass of a helium nucleus, you find that there is a discrepancy. The helium nucleus is slightly lighter. Where did the missing matter go? It went into energy, according to Einstein's formula. The difference in mass is about $0.0477 \times 10^{-27}$ kg – not a lot – but when converted to energy it is equivalent to $4.3 \times 10^{-12}$ J. Therefore, fusing together 1 kg of hydrogen (and converting the tiny fraction of its mass into energy) can generate $6.4 \times 10^{14}$ J (or the amount of energy Earth would generate at current levels for the next 1600 years – now you can appreciate the promise of nuclear fusion for solving our energy needs). To generate enough energy to account for the Sun's current luminosity, you must convert 600 million tons of hydrogen into helium each second. This is a lot of mass, but very little compared with the total mass of the Sun. The energy from nuclear fusion is emitted in the form of gamma rays – the highest-energy form of electromagnetic radiation. As gamma-ray photons make their way through the Sun – colliding and being absorbed and re-emitted by the matter in the Sun – they lose energy. Eventually the energy leaves the photosphere, mostly in the form of visible light. A small amount of the energy is carried off by neutrinos as a by-product of the fusion process. These sub-atomic particles (the name derives from the Italian for "little neutral one") move off at essentially the speed of light and effectively escape the Sun without any interactions. They are so weakly interacting that they pass through you and Earth continuously (in fact, they can pass through several light-years of lead). However, we have developed sensitive instruments that can detect a very small fraction of the neutrinos intercepting Earth. These measurements have helped confirm our understanding of the processes in the core of the Sun.

### 2.5.2 Solar atmosphere

***The photosphere***

The visible surface of the Sun – the photosphere (photo from the Greek for "light") – is opaque to visible light, and hence we see a sharp edge. However, since the Sun is made of gas, it does not have a solid surface.

Galileo first systematically observed the visible surface of the Sun in about 1609. Figure 1.3 is a drawing of Sun's surface made by Galileo showing that the Sun contains spots (whose number waxes and wanes over an 11-year solar cycle). These sunspots are now known to be regions of strong magnetic field. Sunspots appear darker than the surrounding solar surface because they are slightly cooler (typically 4500 K compared to the 5800 K of the surrounding photosphere). They are usually about 10 000 km (one to three Earth radii) across and can last for weeks.

As mentioned in the Introduction, sunspots were used by Galileo to estimate the Sun's rotation rate. The Sun takes about a month to make a complete rotation. Sunspots were used by Richard Carrington 250 years after Galileo first discovered solar rotation to determine that the Sun rotates differentially (that is, moves at different speeds at different solar latitudes). The equator rotates faster than the poles, with an equatorial rotation rate of about 25 days, while the poles rotate every 36 days. The average rotation rate of most importance to Earth is about 27 days.

High-resolution images of the photosphere (Figure 2.2) show that the Sun's visible surface is highly mottled, with dark and light regions called granules. Granulation is direct evidence of solar convection. Bright areas show a Doppler blue shift indicating upward motion. Dark areas show a red shift indicating motion down into the solar interior. The bright regions are about 1000 km across (or about the size of the state of Texas in the US). The bright spots come and go with a lifetime of about 5 to 10 minutes. The bright spots are typically about 500 K hotter than the dark regions.

In addition to this fine-scale granulation, there are observations of super granulation with scale-sizes of tens of thousands of kilometers

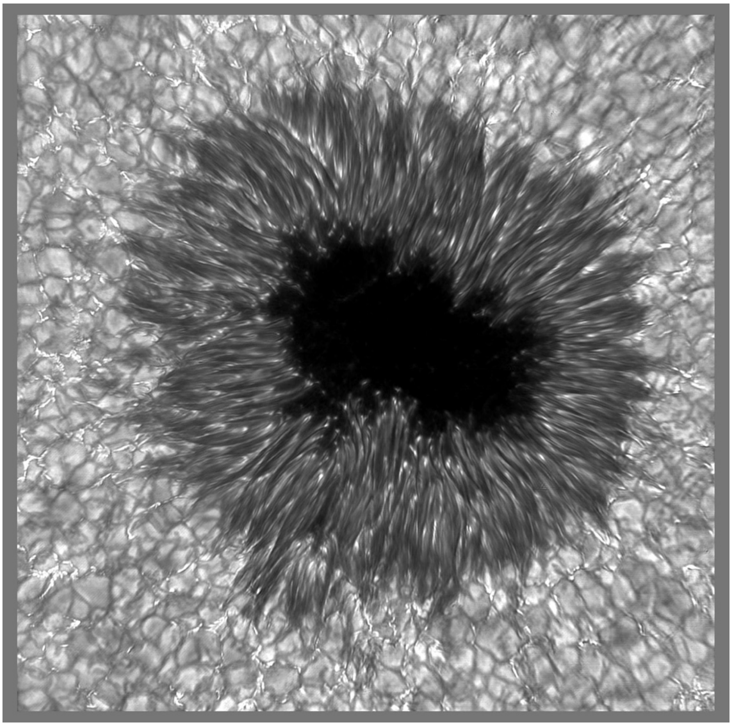


**Figure 2.2** High resolution image of the photosphere showing a sunspot. The smaller features, called solar granulation, provide evidence for convection (from Friedrich Woeger, KIS, and Chris Berst and Mark Komsa, NSO/AURA/NSF).

(or several Earth radii). These super granules are thought to be the signature of larger-scale convection in the outer convective zone.

### *The chromosphere*

The region just above the photosphere is called the chromosphere. The chromosphere, which is about 1500 km thick, is characterized by a temperature higher than that of the photosphere (about 10 000 K compared to the 5800 K temperature of the photosphere). The element helium (helios is Greek for Sun) was discovered in the chromosphere before it was discovered on Earth. The density of the plasma (and hence the amount of light emitted) drops rapidly with height in the chromosphere. Therefore, the chromosphere is not visible against the bright background of the photosphere. However, scientists have known about the chromosphere for a long time by observing total solar eclipses. When the Moon blocks the bright photosphere, the chromosphere becomes visible with its distinctive reddish hue due to the H-alpha (H$\alpha$) emission line of hydrogen. H$\alpha$, which refers to a specific electronic transition within hydrogen, is associated with a unique wavelength of visible light. The chromosphere can be very dynamic with hot jets of gas (spicules) extending high above the surface. These can extend thousands of kilometers above the solar surface and are made of material that has been expelled from the surface at velocities of about 20 to 100 km $s^{-1}$. The chromospheric temperature is inversely proportional to density, so it increases rapidly with height.

When the Sun is viewed through a special filter that only allows H$\alpha$ light through, a wealth of features become apparent (Figure 2.3). Chromospheric networks are web-like patterns most easily seen as bright emissions in H$\alpha$. These networks outline the supergranulation cells mentioned above. Spicules and solar prominences tend to be found at the edges of these cells. Very narrow dark filaments and bright filaments (called plages – French for beach) also are visible in H$\alpha$. Filaments observed above the limb of the Sun, called prominences, are often characterized by their loop-like appearance. The loops map out the solar magnetic field, and the magnetic field is what constrains the motion of the bright gas.

Above the chromosphere the temperature increases dramatically with height in a relatively thin transition region. Above this is the outer atmosphere of the Sun, called the solar corona, which is visible during solar eclipses and from special telescopes called coronagraphs. The corona (Latin for crown) expands out into space supersonically. The solar gas that escapes into interplanetary space is called the solar wind.

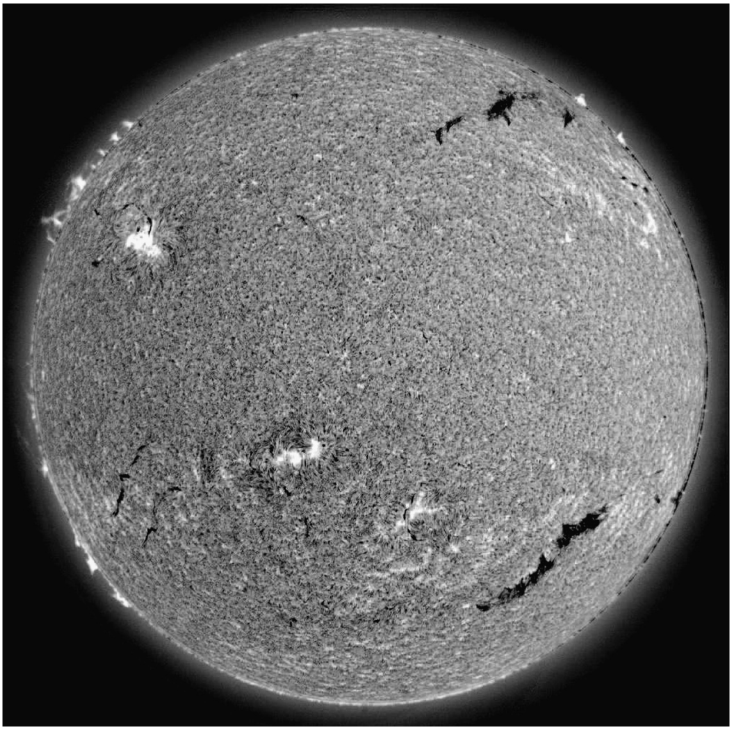


**Figure 2.3** The Sun observed in the light of hydrogen-alpha (courtesy of Big Bear Solar Observatory).

## 2.6 Dynamics and processes

### 2.6.1 Solar magnetism

The magnetic field strength of various regions of the Sun can be determined spectroscopically using the Zeeman effect discovered by Pieter Zeeman[5] in 1896. In the presence of a magnetic field, a gas' spectral lines will be split into two or more components. The actual frequency of the spectral lines depends on the strength of the magnetic field. This discovery won Zeeman and his former teacher (Lorentz)[6] the Nobel Prize in Physics in 1902 and allowed direct measurement of the solar surface magnetic field for the first time. Using this technique, we know that the magnetic field inside a sunspot is about 1000 times stronger than in the surrounding normal solar surface. In addition, the fields are directed either into or out of the solar surface and not randomly oriented. Magnetic fields have a direction and magnitude. For example, a magnet has

---

[5] Zeeman, Pieter (1865–1943), Dutch Nobel-Prize-winning physicist who discovered that spectral lines split in the presence of a strong magnetic field. This is called the Zeeman effect and allows the remote sensing of the strength of the Sun's magnetic field.

[6] Lorentz, Hendrik Antoon (1853–1928), Dutch Nobel-Prize-winning physicist who helped develop the theory of electromagnetism and predicted the Zeeman effect.

a north and south pole with the north pole defined as the side that has magnetic field moving away from the magnet and the south pole where the field is directed towards the magnet. It is thought that the strong fields associated with sunspots interfere with the normal convection of the Sun and hence the gas can cool and appear darker than its surroundings. Each sunspot generally has a single polarity (or magnetic field direction), and they usually appear in pairs of opposite polarity. The north polarity (or outwardly directed magnetic field) makes a loop about the solar surface and then dives into the south polarity (inwardly directed) sunspot. Though sunspots often have an irregular appearance, the magnetic field orientations exhibit a great deal of order. All the sunspot pairs in a given hemisphere of the Sun have the same magnetic configuration. Specifically, when the leading spot (measured in the direction of solar rotation) has a north polarity, then all leading spots in the hemisphere will have a north polarity. What is more, all the sunspot pairs in the opposite hemisphere will have the opposite orientation (south leading). This ordering of the solar field is due to the differential rotation of the Sun. Recall that the Sun rotates faster at the equator than at the poles. As seen in Figure 2.4, this differential rotation causes a twisting of the overall solar magnetic field.

## 2.6.2 Solar active regions

Many pairs of sunspots are associated with explosive releases of energy from the photosphere. These areas of activity are simply called active regions. Though the exact mechanisms that cause the explosive release of energy from the Sun's surface are not known, we do know that they are related to the rapid conversion of magnetic energy into particle kinetic energy. This conversion takes place in regions of strong magnetic fields, and the twisting of the surface magnetic field often leads to rapid energy release. Prominences are an example of this energy release. Another example, solar flares, are much more energetic than prominences. Flares

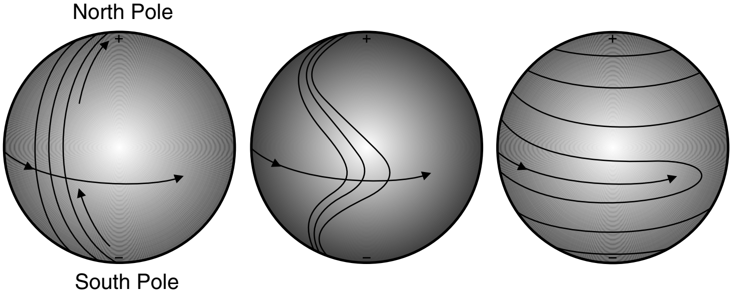


**Figure 2.4** The Sun rotates faster at the equator than at the poles. This is possible because the Sun is a ball of gas. The three panels show the effect of differential rotation on the Sun's magnetic field (from NASA TRACE Mission).

release tremendous amounts of energy in a few minutes and can reach temperatures of 100 million K (much hotter than even the core of the Sun). This energy is equivalent to hundreds of millions of megaton hydrogen bombs exploding at the same time. The energy of these flares is so intense that the charged particles that make up the solar atmosphere are blasted out into space – some at nearly the speed of light. In addition, the heated gas glows at essentially all wavelengths including X-rays. These energetic particles and electromagnetic radiation are ejected into interplanetary space and often can impact Earth's space environment – one of the causes of space weather.

### 2.6.3 Solar cycle

The number of sunspots on the Sun waxes and wanes over an 11-year **solar cycle**. Since sunspots are associated with solar activity (i.e., flares and other rapid releases of energy that can heat localized regions of the atmosphere of the Sun to many millions of degrees kelvin), the solar cycle also describes the level of activity and variability of the Sun. Plate 3 shows an X-ray photograph of the Sun taken by the Japanese Yohkoh Satellite over an 11-year solar cycle. Each snapshot of the Sun shows the structure of the Sun's atmosphere in X-rays at a given day, one snapshot per year. Note that the amount of X-rays increases and decreases over this solar cycle – with the amount of time from the most active or energetic interval (called solar maximum) to the time of least activity (solar minimum) being approximately five to six years. X-rays are very energetic electromagnetic radiation and are created when a gas is heated to about one million degrees kelvin. Notice that during solar maximum, the Sun's atmosphere has many places of bright X-ray emission, whereas at solar minimum the atmosphere of the Sun is essentially black in X-rays, indicating that very little heating of the upper atmosphere of the Sun is taking place. Also note that the X-ray emission is not coming from the surface of the Sun, but from the atmosphere above the photosphere.

This cycle has persisted for centuries. Figure 2.5 shows the number of sunspots observed on the Sun since the 1800s, when sunspots were observed on a daily basis from a number of different solar observatories. The number of sunspots on the Sun continuously changes – sunspots typically have lifetimes from 1 to 100 days – but the total number of sunspots changes with this quasi-regular 11-year cycle.

Recall that sunspots are regions of intense magnetic fields and the variation in their number indicates that the Sun's magnetic field is also changing. In fact the structure and orientation of the Sun's magnetic field changes over a 22-year cycle with the polarity of the Sun's field

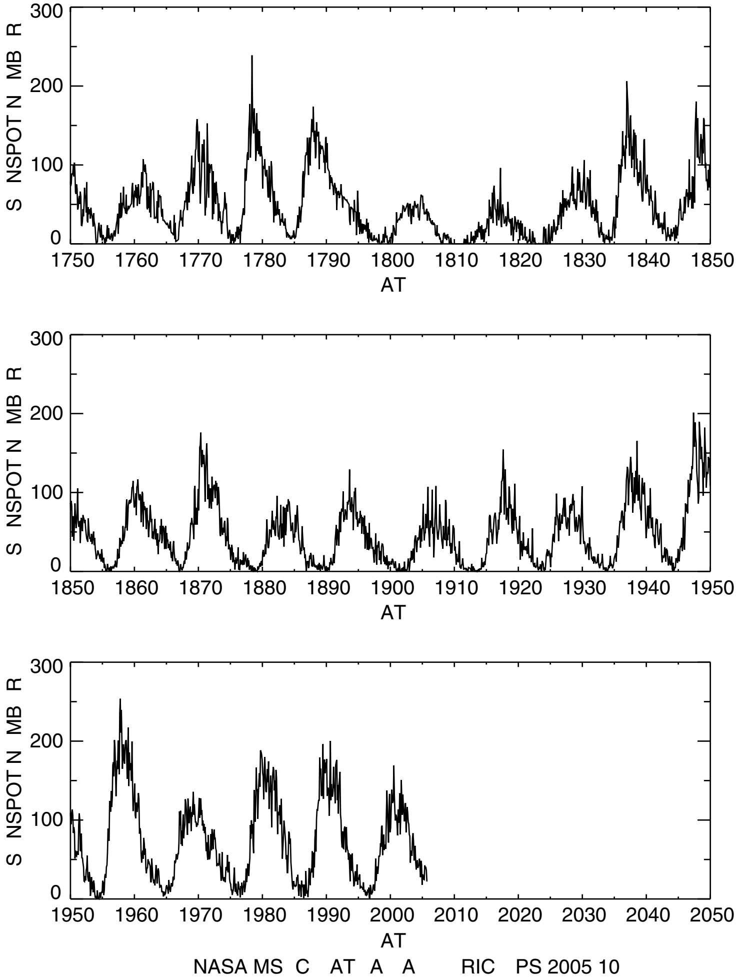


**Figure 2.5** The number of sunspots observed on the Sun since the 1800s. Note the approximately 11-year periodicity of the occurrence of sunspots, which is called the sunspot cycle (from NASA).

reversing over this time interval. The Sun's magnetic field is generated and modulated by the Sun's dynamo, which is powered by the differential rotation of the Sun and the solar convection processes. During solar minimum the Sun's field is relatively simple and well ordered and resembles a dipole magnetic field, with the magnetic field coming out of one hemisphere and going in the other. Over the next five to six years, as the Sun approaches solar maximum, the nice dipole configuration slowly disappears and the Sun becomes magnetically disorganized and highly structured. Past solar maximum, over the next five to six years the magnetic field again becomes more organized and more dipolar. Early in this transition the tilt of the weak dipole field can be very large with respect to the spin axis of the Sun, but as solar minimum approaches, the dipole axis orientation becomes more and more aligned with the spin axis. When the dipole field is reformed, it has the opposite polarity to the previous one. This change in polarity is what defines the 22-year magnetic cycle of the Sun, sometimes called the double solar cycle or the Hale[7] cycle. The polarity of sunspot pairs also follows this cycle. For the first 11 years of the magnetic cycle, the leading spots of a particular hemisphere always have the same polarity, which is opposite to the polarity of the leading spot in the other hemisphere. The sunspot polarities then reverse for the next 11 years.

Since the amount of solar activity follows this sunspot or solar cycle, one would expect that the number of solar disturbances that impact Earth would also follow this cycle. This is exactly what happens. Therefore space weather has "seasons", with solar maximum indicating a strong likelihood of severe space weather and solar minimum predominantly quiet space weather. The next chapter describes the outer atmosphere of the Sun – the solar corona and solar wind. Earth resides in the outer atmosphere of the Sun and so our space environment is intimately connected to the structure and dynamics of the Sun. We live in the outer atmosphere of the Sun.

## 2.7 Supplements

### 2.7.1 Electromagnetic spectrum and radiation

***Frequency and wavelength***

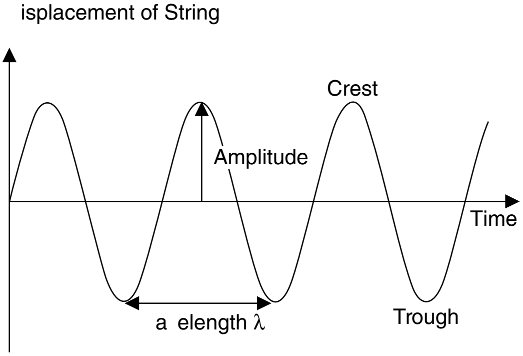
Many things in nature, such as a pendulum, an ocean buoy, the walls of our hearts, and a piano string, move back and forth or up and down in a regular manner (are said to oscillate). Some oscillations perturb their surroundings and create waves, and the perturbations or waves carry energy from their source to their surroundings. A piano string, for example, creates sound waves that perturb the surrounding air's density and push it back and forth as the string oscillates. Our ears are sensitive to these air density fluctuations, and we can hear the sound waves.

A characteristic of these waves is the time between oscillations (a cycle), or in the case of a sound wave generated by a piano string, the time it takes the string to move back and forth or the time it takes one density perturbation cycle to pass us. We call this time the period ($T$). A closely related parameter called frequency ($f$, sometimes written as $\nu$) is the number of events (or in the case of the piano wire, the number of back and forth oscillations or cycles) in a given time interval. Period and

---

[7] Hale, George Ellery (1868–1938), American solar astronomer instrumental in founding a number of astronomical observatories.

**Figure 2.6** The displacement of a string away from its resting position as a function of time.



frequency are related by $f = (1/T)$. The SI unit of frequency, equivalent to one cycle per second, is the hertz[8] (Hz).

One way to graphically visualize a wave is to record the displacement (or amount of motion) of the string as a function of time, or, in other words, plot the string's position (along the $y$-axis) as a function of time (along the $x$-axis). Figure 2.6 shows the position of a point at the middle of the piano string after it has been struck as a function of time. Note that the string moves back and forth about its resting position. The amount of displacement (a measure of the size of the wave) is called the amplitude. Another distance shown in the figure is the distance between troughs (or crests) of the waves, called the wavelength (often written as the Greek letter $\lambda$).

Now let's examine a sound wave created by the vibration or oscillation of a piano string. Instead of plotting the position of the string as a function of time, we could record the density of the air at a given point next to the string as a function of time. This plot would look just like the position of the string as a function of time because when the string is moving towards our microphone or ear, the air is compressed and its density goes up. When the string moves away from our microphone, the air is rarefied and its density goes down. The timing between the air density changes is exactly the same as the string's back and forth motion that created them. Therefore the sound wave also can be described by its period or frequency, its wavelength and amplitude. In fact, all periodic waves can be described in exactly the same way.

Recall that waves are disturbances that move away from their source. A water wave moves away from your hand as you splash in a pool; a sound wave moves away from a piano string. How fast do these waves

---

[8] Hertz, Heinrich Rudolf (1857–1894), German physicist whose study of electromagnetic radiation led to the discovery of radio waves. The SI unit of frequency is named in his honor.

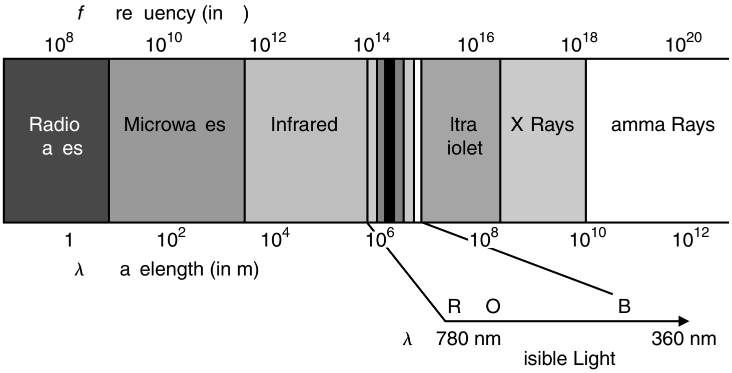


**Figure 2.7** The electromagnetic spectrum ordered left-to-right from low energy to high energy.

move? The answer depends on the density and temperature of the material the wave propagates through. A wave will move at different speeds through different materials (sound will travel at different speeds through air and water, for example) and its speed also depends on the temperature of the particular material (i.e., sound waves move faster through warm water than through cold water).

There is a relationship between a wave's frequency, wavelength and velocity. The velocity is equal to the wavelength times the frequency

$$V = \lambda f.$$

For light, the velocity is written $c$ (as in Einstein's famous equation) and the energy of the electromagnetic radiation depends on the frequency (or wavelength). High-frequency EM radiation has more energy than low-frequency radiation. This linear relationship can be written

$$E = hf,$$

where $h$ is Planck's constant[9] and $E$ is energy. So for visible light, red is lower frequency than blue and therefore red light carries less energy than blue light. Figure 2.7 shows the frequency and wavelength of the entire EM spectrum.

### *The Doppler effect*

The Doppler effect is the shift in frequency of a wave due to the relative motion of the sound emitter and observer. For example, as a fire truck with its sirens blaring approaches, we hear a higher-pitched tone than when it is receding.

---

[9] Planck, Max Karl Ernst Ludwig (1858–1947), German Nobel-Prize-winning physicist whose discovery that energy exists in fundamental units called quanta marked the beginning of the development of quantum theory and the founding of modern physics. The Planck constant is equal to $6.6261 \times 10^{-34}$ J s$^{-1}$.

The effect only occurs for relative motion towards or away from the observer. The relationship between the component of velocity along a line connecting the emitter and the observer is

$$f' = f_0 \left( \frac{v \pm v_\mathrm{o}}{v \pm v_\mathrm{s}} \right), \tag{2.1}$$

where $f'$ is the perceived frequency; $f_0$ is the actual frequency; $v_\mathrm{s}$ is the speed of the source; $v_\mathrm{o}$ is the speed of the observer; and $v$ is the speed of the waves. The $(+)$ is used when the source and observer are moving toward each other, and $(-)$ when they are moving apart.

### *Photons and energy*

Light is one form of electromagnetic radiation. What is interesting about electromagnetic radiation is that it behaves as both a wave and as a discrete particle. Light will act as a wave if we send it through a prism or diffraction grating and will separate out into a rainbow because of refraction. If we shine light at a metal, electrons can be emitted if the light's energy (or frequency) is high enough (the photoelectric effect). Produced when the light energy is above a threshold level, the photoelectric effect demonstrates the particle nature of light (its frequency rather than its intensity determines whether electrons are emitted). Because EM radiation can act like a particle (i.e., it can be thought of as a discrete object as opposed to a continuous wave), it is given a special name – photon. Therefore you can think of light as either a wave or a photon. How you think about it depends on the effect that you are observing. The bottom line, though, is that you should think of electromagnetic energy as waves or photons of energy. The amount of energy depends on the frequency of the radiation. This wave–particle duality of light caused lots of controversy early in the formation of modern physics and led to the development of quantum mechanics.

### *Blackbody radiation*

If you look closely at a candle flame, you will notice several colors in the vicinity of the wick (blue close to the wick and yellow and orange farther away). What we just learned about frequency (and hence color) tells us that the energy of the light we see must indicate different temperatures in various parts of the flame. All matter in the universe (including you) emits electromagnetic radiation. The amount and frequency of the radiation depends on the object's temperature. Hopefully your body temperature is near 98.6 °F (or 37 °C), and therefore you are emitting electromagnetic radiation mostly in the infrared (IR) portion of the electromagnetic spectrum. Some animals such as nocturnal snakes have photoreceptors (fancy name for eyes) sensitive to IR radiation. A candle

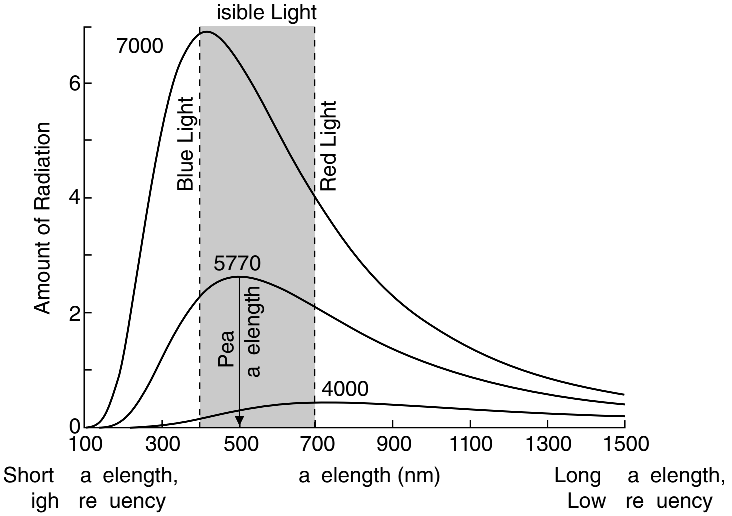


**Figure 2.8** All objects emit EM radiation at a spectrum of wavelengths depending on their temperature. The peak of the blackbody curve moves to shorter wavelengths (higher energy) as the temperature increases. The Sun emits a spectrum corresponding to a photospheric temperature of about 5770 K.

flame burns hotter than our bodies' temperature, and therefore the radiation that it emits is higher energy (and frequency). Hence we see light shining from a candle flame. The amount of EM radiation and its peak frequency depend on the temperature of the object. This type of radiation is called blackbody radiation. Blackbody radiation has a property that makes it easy to identify – that is, the amount of radiation at any specific frequency depends only on the temperature. Figure 2.8 shows what is called a blackbody curve. The plot shows the amount of radiation emitted as a function of wavelength. Each curve represents an object at a certain temperature. Note that as the temperature of an object increases, the amount of radiation emitted increases and the wavelength of the peak (or maximum) energy gets shorter. Since wavelength is inversely proportional to frequency, the peak frequency (and hence energy) emitted increases. A simple relationship called Wein's law[10] relates the peak wavelength to the temperature:

$$\lambda_{\text{peak}} T = 2.898 \times 10^{-3}\ \text{m} \cdot \text{K}.$$

Space scientists know the surface temperature of the Sun because of this. We can measure the amount of electromagnetic radiation emitted from the Sun as a function of wavelength and then fit the data to a blackbody curve that has a nice analytical mathematical expression, and voilà, we know the Sun's surface temperature. Much of astronomy and

---

[10] Wein, Wilhelm Carl Werner Otto Fritz Franz (1864–1928), German Nobel-Prize-winning physicist whose work in thermodynamics and blackbody radiation led to the development of the relationship between wavelength and temperature for a blackbody radiator now named in his honor.

space science uses this relationship between the amount of radiation emitted as a function of frequency and temperature to understand the evolution and dynamics of stars and the temperature of planets.

## 2.8 Problems

2.1 How many Earths could fit inside the Sun?

2.2 What has been the average length of time between solar maxima for the last five solar cycles (Figure 2.5).

2.3 During fusion of hydrogen to helium, 4 billion kg of matter are converted to energy each second. What fraction of the Sun's total mass is lost each year to this process?

2.4 The Sun expels 1 billion kg of matter each second as the solar wind. What fraction of the Sun's total mass is lost each year to this process?

2.5 Order the following electromagnetic bands in terms of energy from lowest to highest: (X-ray, visible, gamma ray, radio, microwave, ultraviolet, infrared). Which band has the highest frequency? Which band has the longest wavelength? Which frequency of EM radiation corresponds to a wavelength of 1 m?

2.6 Using dimensional analysis (a technique of making sure both sides of an equation have the same physical dimensions), write a proportionality equation showing the relationship of wavelength to frequency (note that the product is equal to a constant – the velocity of the wave).

2.7 What is the thermal pressure of the gas at the surface of the Sun?

2.8 There is a direct relationship between an object's angular size and its distance from the observer. The Sun is 400 times the size of the Moon. How much farther away from Earth is the Sun than the Moon for them to have the same ($0.5^\circ$) angular size as seen from Earth?

2.9 Using Wein's law, at what wavelength does the Sun emit the most electromagnetic radiation? Is this in the visible part of the electromagnetic spectrum? If so, to what color does this correspond?

2.10 What is the Sun's total luminosity (power or energy per second)? What is the amount of power that intersects one square meter at the Earth (this is called the solar constant)? At Mars (1.5 AU)?

2.11 What is the wavelength shift of a chromospheric spicule emitting in the H$\alpha$ that has a motion of 100 km $s^{-1}$ towards the observer?

# Chapter 3
# The heliosphere

The heliosphere is defined as the region of interplanetary space where the solar wind is flowing supersonically.

> *Dessler, A.J.* Reviews of Geophysics*, **5**, 33, 1967. The first use of the term heliosphere in the scientific literature.*

## 3.1 Key concepts

- plasma
- solar wind
- heliosphere
- interplanetary magnetic field
- coronal mass ejection
- cosmic rays

## 3.2 Introduction

Sunlight, which bathes Earth with heat and light, is only part of the energy that flows constantly from the Sun. Ionized gas (**plasma**) and magnetic field are continuously expelled as **solar wind**, as well. Solar wind was discovered in the 1950s when it was noticed that the plasma tail of a comet always points away from the Sun, even when the comet is moving back towards deep space. Figure 3.1 shows a schematic of a typical comet and its tail at different points in its orbit around the Sun. A comet tail is made up of cometary material that has been heated by sunlight as it gets close to the Sun and has escaped from the nucleus. The glowing cloud of neutral and ionized gas and dust around the nucleus is called the coma. The material in the coma is then "blown" back away from the Sun. Creation of a comet tail requires not only sunlight, but also the energy and momentum of a gas flowing supersonically away from the Sun, the solar wind. The solar wind expands out into interplanetary space and carves out a region surrounding the Sun called the **heliosphere**.

## 3.3 The corona and the solar wind

During a total solar eclipse (when the Moon is lined up exactly between the Sun and Earth), light can be seen around the Sun. This light, referred to as the corona, is sunlight scattered by electrons in the Sun's outer

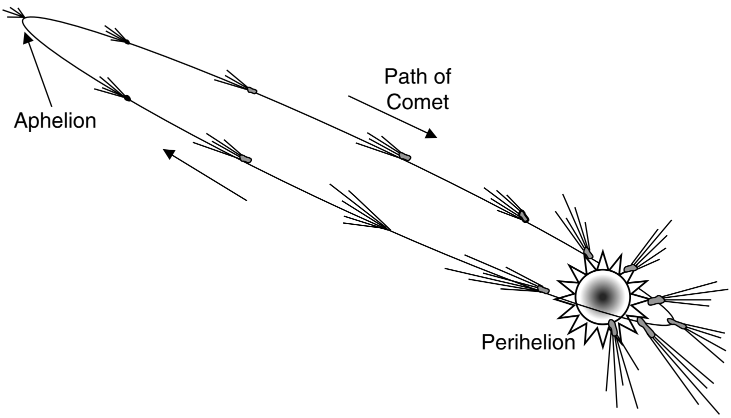


**Figure 3.1** As a comet orbits the Sun, its tail always points away from the Sun whether it is moving towards or away from the inner Solar System. Bierman used this observation to predict the existence of the solar wind that continuously blows away from the Sun.

atmosphere. The corona is not spherically symmetric or equally bright in all directions. Its spatial structure, with more emission (and hence plasma) near the equator than the poles, is due to the structure of the Sun's magnetic field. Plate 4 shows the corona during a total solar eclipse. The visible photosphere is a million times brighter than the corona, and therefore the corona can only be seen when photospheric light is blocked. Telescopes called coronagraphs now routinely observe the corona even without a total solar eclipse by using an occulting disk to block out photospheric light.

For a particle to escape the Sun's gravitational pull in the lower corona, it must be moving faster than the Sun's escape velocity, 618 km $s^{-1}$, and be in a region tenuous enough that collisions with other particles are infrequent. The Sun's escape velocity is the speed at which a particle needs to be moving away from the Sun in order for it to never slow down and turn around and fall back onto the Sun. Coronal particles have a temperature of over 1 million kelvin, and therefore many particles have velocities great enough to escape the Sun.

These particles make up the solar wind, a plasma composed of mostly protons, helium nuclei, and electrons that moves supersonically away from the Sun and carries with it the Sun's magnetic field. Therefore, the solar wind is a magnetized plasma.

## 3.4 The interplanetary magnetic field

The part of the Sun's magnetic field that is pulled out into the heliosphere by the solar wind is called the **interplanetary magnetic field** (IMF). Its characteristic spiral configuration when viewed from above or below the equatorial plane is due to the Sun's rotation. The magnetized solar wind expands radially (directly away from the Sun), pulling

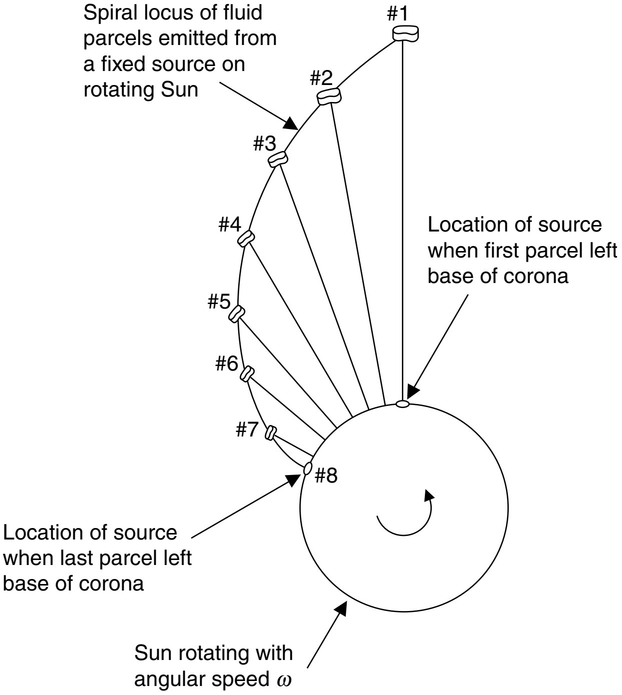


**Figure 3.2** Because of the Sun's rotation, its magnetic field is wrapped into an Archimedian spiral. The figure shows the location of eight parcels emitted at constant speed from a source fixed on the rotating Sun (from Kivelson and Russell, 1995).

the solar magnetic field along with it. As the Sun rotates, the position or footpoint of where the solar wind stream leaves its surface moves counter-clockwise when viewed from above the Sun. This causes the magnetic field to start to spiral as it moves farther from the Sun with respect to the original footpoint position (see Figure 3.2). This is called an Archimedian spiral after the Greek scientist Archimedes,[1] who first described it mathematically. It is analogous to a stream of water being shot out of a rotating sprinkler head. Even though each individual drop or parcel of water moves radially away, since the sprinkler is rotating, the stream appears to curve or spiral out.

Because the Sun's rotation rate is essentially constant in time, the angle the spiral makes with respect to the Sun–Earth line (an imaginary line connecting the Sun and Earth) is due to the speed of the solar wind alone. Faster solar wind will create a smaller angle because it will

---

[1] Archimedes (c. 287–212 BC), Greek mathematician and scientist who is credited with discoveries in hydrostatics and mechanics. Most famous for his Archimedes' principle which states that a body immersed in water will displace a volume of fluid that weighs as much as the body would weigh in air.

move farther away from the Sun in a given time than a slower parcel. Because of the structure of the Sun's magnetic field (as revealed in the eclipse picture, Plate 3), the solar wind streams away from the Sun at different velocities. Solar wind flowing from closed-field regions near the Sun's equatorial plane is slower than that moving away from the Sun's open magnetic field regions. In the coronagraph photo, regions of dipole-like field with a loop of plasma coming out of the rim are called closed-field regions. Near the poles are open-field regions, where a "ray" of plasma flows along a field line not looping back to the surface. In X-ray photographs of the Sun (e.g., Plate 3), these open-field regions, referred to as coronal holes, appear dark. The location and existence of these two types of field regions change, which means that the speed of the solar wind observed at Earth changes constantly, but is most simply characterized by intervals of high-speed solar wind and slow-speed solar wind.

The IMF has not only an Archimedian spiral pattern, but also structure in the north–south direction because the magnetic equator of the Sun is not perfectly aligned with the Sun's spin axis. Therefore, from the vantage point of Earth, we see solar wind coming from alternating sides of the Sun's magnetic equator as the Sun rotates. This gives rise to a flapping or wavy structure of the IMF. Figure 3.3 shows the structure of the IMF in three dimensions. The surface shown is the location of the plasma that came from the Sun's magnetic equator. Called the heliospheric current sheet, this divides IMF that is pointing away from the Sun from IMF pointing towards the Sun. Table 3.1 gives some properties of the solar wind and the IMF.

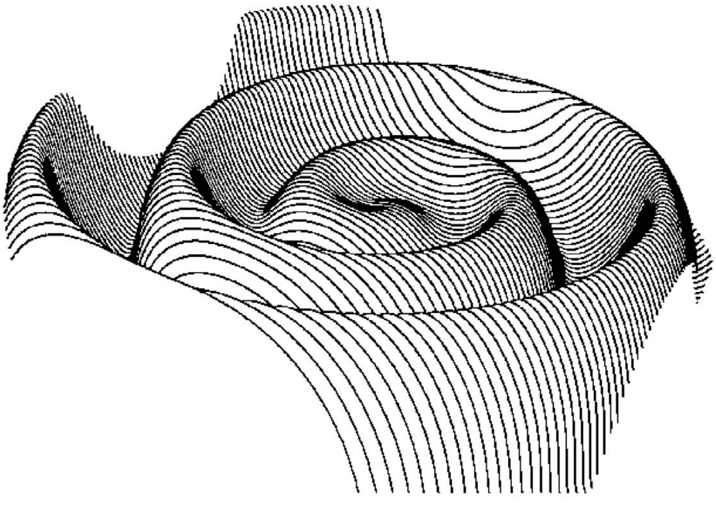


**Figure 3.3** A perspective plot of the heliospheric current sheet showing the wavy nature of the interplanetary magnetic field due to the tilt of the Sun's magnetic axis with respect to its spin axis (from Jokipii and Thomas, 1981).

Table 3.1 *Average properties of the solar wind and IMF at 1 AU*

| | |
|---|---|
| Number density | 5 particles $\mathrm{cm}^{-3}$ |
| Temperature | 1 000 000 K |
| Velocity | 400 km $\mathrm{s}^{-1}$ |
| Composition | 90% H, 8% He, and trace amounts of other heavy ions |
| IMF strength | 10 nT |

## 3.5 Coronal mass ejections

The motion and structure of the Sun's magnetic field make the corona ever changing. Occasionally, reconfigurations of the solar magnetic field cause a large portion of the corona to blast away from the Sun and out into the heliosphere. These **coronal mass ejections** (CMEs) were discovered with the first space-based coronagraph images in the 1980s. CMEs are large-scale magnetic structures that can contain over a trillion ($10^{12}$) kilograms of hot coronal material. A trillion kilograms is equivalent in mass to over a quarter of a million (250 000) aircraft carriers. CMEs can move away faster than the background solar wind and have velocities of over 1000 km $\mathrm{s}^{-1}$ (several million miles per hour). Many of these fast CMEs set up shock waves in front of them as they stream away from the Sun. A shock wave is formed when the speed of an object exceeds the sound speed of the background material. Sound speed is dependent on a material's density and temperature. An example would be a bullet moving through air. The sound speed of air is approximately 300 m $\mathrm{s}^{-1}$, so if the bullet moves faster than 300 m $\mathrm{s}^{-1}$, a shock wave forms in front of it. With respect to a CME, the background solar wind is moving on average 400 km $\mathrm{s}^{-1}$. The sound speed of the solar wind is approximately 40 km $\mathrm{s}^{-1}$. Therefore, if a CME is launched with a speed greater than about 40 km $\mathrm{s}^{-1}$ (faster than the background solar wind), a shock wave will form. One important aspect of a shock wave in interplanetary space is that it is a very good particle accelerator. Therefore fast CMEs drive a shock that can generate large amounts of energetic particles as it plows through the slower solar wind. These energetic particles can reach Earth and cause damage to satellites.

CMEs often have distinctive, loop-like magnetic field structures called flux ropes (see image from the NASA SOHO satellite in Plate 2). When referring to a CME specifically, these structures are called magnetic clouds. When observed in interplanetary space, CMEs have distinct plasma and magnetic field properties that are used to identify them.

They often take over a day to pass by Earth, implying a length scale of a quarter of an astronomical unit (0.25 AU). CMEs often have a shock and a high-density "plug" of plasma in front due to slower solar wind plasma being "swept up" like snow by a snowplow. Fast CMEs are the leading cause of major geomagnetic storms and therefore are one of the most important solar phenomena when discussing space weather.

## 3.6 The outer heliosphere

The outer heliosphere is defined as the region well beyond the orbit of Pluto where the solar wind interacts directly with interstellar space. Pluto is on average 40 AU away from the Sun. The heliopause – the boundary between the heliosphere and the interstellar medium (ISM) – is thought to be on the order of 100 AU away. In 2003 Voyager I passed the termination shock (formed between the supersonic flow of the solar wind and the interstellar medium) when it was 90 AU from the Sun. The ISM is thought to be a mixture of electrically neutral atoms and magnetized plasma. Since the Sun orbits the center of our galaxy at approximately 220 km $s^{-1}$, there is a relative velocity difference between the heliosphere and the surrounding ISM. This gives rise to a teardrop shaped heliosphere with a termination shock inside the heliopause (see Plate 5).

The heliopause proper separates the Sun's magnetized plasma from the ISM. This is analogous to the magnetopause that separates Earth's magnetic field from the magnetized plasma of the shocked solar wind. In the direction of motion of the Sun, another shock is formed between the moving heliosphere and the ISM. The heliosphere is estimated to be moving roughly perpendicular to the line between the Sun and the center of the Milky Way galaxy – hence we are approximately orbiting the galaxy in a circular path. At 220 km $s^{-1}$ it takes approximately two hundred and fifty million (250 000 000) years to orbit the galaxy once. In its 4.5 billion year history, Earth has orbited the galaxy 18 times. So Earth is about 18 "galactic" years old. Beyond this bow shock is unperturbed ISM that extends out to the next stellar atmospheres. The closest star, Alpha Centauri, is about four light-years away (a light-year is the distance light travels in one year or about 10 trillion km). Of course, there is space weather around that star as well, but currently we do not know if Alpha Centauri has a planetary system, though most models of star formation suggest that a planetary disk can form out of the stellar nebula. Hence we anticipate that there may be a planetary system about Alpha Centauri, as well.

## 3.7 Cosmic rays

Earth is constantly bombarded from every direction by highly ionized atoms and other subatomic particles known as **cosmic rays**. "Ray" is a misnomer, since cosmic rays consist of energetic particles. Cosmic rays are further divided into two components: particles that originate outside our heliosphere (called galactic cosmic rays) and those that originate from our Sun (called solar energetic particles). Cosmic rays travel at nearly the speed of light, and most are nuclei of atoms. The composition of cosmic rays spans the periodic table, from the lightest particles, such as hydrogen and helium, to the heaviest, such as iron. Cosmic rays also include electrons, positrons (the mirror particle of an electron – essentially a positively-charged electron), and other subatomic particles. The energy of cosmic rays is usually measured in MeV (mega-electron-volt or a million eV) or GeV (giga-electron-volt or a billion eV). An electron-volt is an energy unit equivalent to the energy gained by an electron accelerated through a one-volt potential electric field. One eV is equal to $1.6 \times 10^{-19}$ joules (J). Typical energies of galactic cosmic rays are between 100 MeV and 10 GeV. To put this in perspective, if the galactic cosmic ray is a proton, it has a velocity between 43% and 99.6% the speed of light to have this much energy. The highest energy cosmic rays ever measured have kinetic energies equivalent to that of a tennis ball hit by a top men's professional tennis player – all contained in a single atomic nucleus.

The current theory is that galactic cosmic rays originate in supernovae (stellar explosions). It is estimated that one supernova happens every 50 years in a galaxy like the Milky Way. One type of supernova is the death-throes of a massive star. After the star has consumed all of its fuel by thermonuclear fusion, its outer layers collapse and cause a huge explosion that expels stellar material into space and causes shock waves to form. The explosion and shock waves then produce particles of very high energy. The shock waves continue to propagate away from the progenitor star (or the star that went supernova), continuously accelerating particles for many years after the explosion.

Since cosmic rays are charged particles, their motion is deflected by galactic magnetic fields as they propagate through interstellar space. Therefore it is not directly possible to identify their source since their path is a "random walk" from their source to us (i.e., since the galactic magnetic field is highly structured, cosmic rays are essentially scattered in every direction as they propagate through space). The Sun's magnetic field (the IMF) and Earth's magnetic field also influence the motion of cosmic rays through our Solar System and to Earth's surface.

When high-energy cosmic rays hit Earth's atmosphere, they collide with atmospheric particles and cause showers of secondary particles to hit the surface. Each collision takes energy from the original cosmic ray, creates new particles, and energizes the atmospheric gas particles that are hit. These in turn can hit other particles, energizing them. They then can hit new particles, etc. Depending on the energy of the incoming cosmic ray, large fluxes of secondary particles can reach Earth. A by-product of these collisions is the creation of pions, unusual sub-atomic particles, which usually decay quickly to produce muons, neutrinos, and gamma rays. Muons also subsequently decay into electrons and positrons. The flux of these particles is equivalent to about 1000 particles per minute passing through our bodies. However, the effect is only a small part of the natural background radiation. In space the flux can be considerably higher and cause damage to satellites or death to astronauts.

## 3.8 Supplements

> Can no one laugh? Will no one drink? I'll teach you physics in a wink.
>
> George Gamow, *Thirty Years that Shook Physics*, p. 190.

### 3.8.1 How do we describe motion?

Scientists and engineers study the motion of everything from golf balls, airplanes, blood cells through the body, chemicals through cell membranes, radiation in space, cars involved in a collision, and Earth around the Sun, just to name a few diverse examples. In order to understand motion of an object, several things need to be defined, such as what does it mean to move?

Mechanics is the study of moving bodies or objects. Motion is defined as the change in position of a body with respect to another body or to some reference frame (such as a room or Earth). It is often convenient to define a fixed reference frame (called an inertial reference frame) so that observers can describe to each other the motion of any object in a way that makes sense to both. The development of the concept of a reference frame is one of the great achievements of math and science. For example, you are probably reading this chapter sitting down. You would then describe your motion as zero, or in other words, you are not moving with respect to your surroundings. This is true if you define your surroundings as the room. However, what if you are sitting in the back seat of a car driving down a road? You are not moving relative to the driver, other passengers, or the car's interior, but to an observer on the side of the road, you have motion relative to the road and its surroundings. Therefore, when you describe motion (or lack of motion), you must also specify the reference frame you are using. A fundamental principle of physics is that the laws of physics are valid and identical in all reference frames. For space physics, the frame of reference is often the fixed Earth or the Sun. For example, we can say that we are sitting still with respect to the surface of Earth, but an observer fixed in space directly overhead will see us move towards the east as Earth rotates. An observer fixed with respect to the Sun would see not only us rotating with Earth, but also Earth receding as it orbits the Sun. An observer fixed with respect to the center of the Milky Way galaxy would see us rotating around the Earth's axis, orbiting the Sun, and the Sun in turn moving around the center of the galaxy. Of course, we can move farther out and use a system fixed not to our galaxy, but to the relative positions of the different galaxies in our local neighborhood (though to make things more complicated, those galaxies are also moving so sometimes it is difficult to clearly define where you are or even if you are moving). Any observer can easily transform calculations from one reference frame to another, though in each succeeding step we need more information about the relative motion and position of each reference frame to the succeeding reference frame.

In order to describe the position of an object in a reference frame, we must define a coordinate system. A coordinate system is a set of rules that describes quantitatively where an object is located from a specific point. The specific point is called the "origin". Figure 3.4 shows a two dimensional coordinate system you are probably familiar with. The $x$ and $y$-axes form a right angle (90°). This system, first developed by French mathematician René Descartes, is called a Cartesian coordinate system in his honor. This coordinate system can be used to describe the position of an object in a plane. A plane is a two-dimensional region

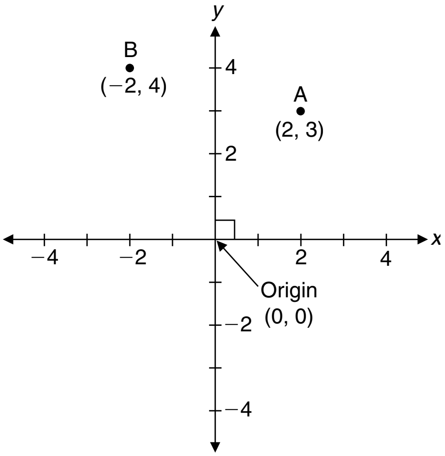


**Figure 3.4** A two-dimensional cartesian coordinate system.

(like the surface of a piece of paper) that can be described by two coordinates (in the example, $x$ and $y$). A three-dimensional coordinate system would need a third axis (usually denoted $z$) to specify the position above or below the $x$–$y$ plane. The tick marks on the axis can be used to describe a point anywhere on the coordinate system. The origin is defined as being at position (0, 0) and by convention we define distances to the right and up as increasing positive numbers and those to the left and down as increasing negative numbers. Point A in Figure 3.4 is therefore at position (2, 3); in other words, two tick marks over to the right of the origin on the $x$-axis and three tick marks up the $y$-axis. Point B in Figure 3.4 is at position (−2, 4) or two tick marks to the left of the origin and four tick marks up the $y$-axis. Road maps are usually drawn on a similar coordinate system. The axes are then labelled North, South, East, and West.

If we want to understand the motion of a parcel of solar wind on its trajectory to Earth, we often will use a coordinate system centered on the Sun. One such coordinate system is the solar ecliptic frame of reference. The center of the Sun is at the center of this system (or at the origin), and a line connecting the center of the Sun and the center of Earth is defined as the $x$-axis. The $z$-axis is the line perpendicular (or at a right angle) to this line and the plane that contains Earth in its yearly orbit around the Sun. You can visualize this system as a CD or DVD. The Sun is at the center of the CD/DVD and Earth travels about the Sun on the surface of the CD/DVD. The two-dimensional plane that contains the center of the Sun and Earth as it sweeps out its motion around the Sun is called the ecliptic plane. This plane also contains all the planets in the Solar System. We can now say this without the Pluto caveat. Pluto's orbit is inclined to the ecliptic by 17°. This large inclination, which is different from those of the other eight planets, contributed to Pluto's demotion to dwarf planet status in 2006.[2]

If we are interested in locating a place on the surface of Earth, we usually use a geographic coordinate system with degrees instead of distance units because Earth can be represented as a sphere or globe. In this system, we define the origin as the point at the equator where the meridian (or the north–south line that crosses the equator at right angles) goes through London. This has been defined as the 0 (or prime) meridian (remember the British Navy ruled the seas in the seventeenth and eighteenth centuries and hence were able to define the meridian that

---

[2] In 2006, Pluto was reclassified as a dwarf planet. This change was spurred by the discovery of another object larger than Pluto called Eris amongst thousands of other bodies that share orbital characteristics with Pluto. These bodies are called Kuiper Belt objects.

contained their main observatory at Greenwich as the reference line still in use today). Since there are 360° around a circle, the globe is marked off in degrees starting at the 0 meridian and going eastward all the way around Earth. This coordinate is called longitude. The position above or below the equator is also given in degrees. By convention we define the North Pole to be $+90°$ and the South Pole to be $-90°$. Do you know your present latitude and longitude?

### *Velocity*

Once we have defined a reference frame and a coordinate system, we can measure an object's change in position as a function of time. We can note the position of a car driving down a highway (designated as $x_1$) at some time (designated $t_1$) and then measure its position ($x_2$) at some later time ($t_2$). We can then say that the car moved $(x_2 - x_1)$ distance in $(t_2 - t_1)$ amount of time. The ratio of these differences is called a speed:

$$\text{speed} = \frac{(x_2 - x_1)}{(t_2 - t_1)} = \frac{\Delta x}{\Delta t}. \tag{3.1}$$

Speed has the dimensions of distance divided by time or, in standard metric units (Système International – SI), meters per second. In our cars we usually denote speed in miles per hour or kilometers per hour. Speed is an important parameter for understanding the dynamics of an object, but one additional parameter, the direction of motion, is needed for us to predict the future position of an object. We define an object's speed and direction of motion as its velocity. Velocity is a vector – a mathematical quantity that has both a magnitude (speed) and direction. Speed is a scalar (a parameter with a magnitude, but no direction). Vector mathematics is the rules that are followed when describing vector quantities. The simplest vector mathematics that we can do is adding two vectors. As an example, let us determine the velocity of a person walking down the aisle of a moving train. If the train is moving at 10 km per hour due north and the person is walking north towards the front of the train at 2 km per hour, then the total velocity of the person relative to the stationary tracks is $(\vec{v}_{\text{train}} + \vec{v}_{\text{person}})$ or 12 km per hour heading north. Note the arrows above the velocity symbol. These arrows signify that the quantity is a vector rather than a scalar quantity. Now consider a person walking south to the back of the train (the velocity of the train and the velocity of the person on the train are in opposite directions). If we define moving north as positive, then moving south would be described as having a negative velocity. Therefore the resultant sum of the velocity vectors of the train moving north at 10 km per hour and the passenger walking toward the back of the train (south) at 2 km per hour would be (10 − 2) or 8 km per hour northward. Note that the person is still moving north relative to the tracks even though he is moving south inside the train because the train is moving faster than the passenger is walking. This is similar to walking up an escalator. Your resulting speed is the sum of the escalator's speed and your walking speed. Now if you attempt to walk up a "downgoing" escalator you can still go up as long as you walk faster than the escalator is traveling down. However, your speed relative to the building will be less than your speed relative to the escalator steps.

In space physics we are often interested in the velocity of a parcel of plasma from the Sun on its way to Earth. In the solar ecliptic reference frame, the motion of the solar wind is radially outward from the Sun. For Richard Carrington to have observed a geomagnetic effect associated with a solar flare observed 17 hours earlier, the average speed of the plasma between the Sun and Earth can be calculated by looking at the ratio of the distance traveled to the time it took to reach Earth ($\vec{v} = (d/t)$). The Sun is 150 million km away from Earth, so the average speed the parcel of solar wind must have had if the solar flare and geomagnetic storm were related is: $150\,000\,000$ km/17 hours $= 2450$ km s$^{-1}$. The 1859 storm has been re-examined and found to be the strongest geomagnetic storm on record. This velocity is one of the highest ever estimated for a parcel of solar wind, which has an average velocity of about 400 km s$^{-1}$.

### *Acceleration*

To predict where an object will be in the future, we need to know not only its position at different times, but also its velocity at those times. By observing whether the velocity changes at each time step, we can learn if the object is slowing down, speeding up, or changing direction. If the object is doing any of these three things, we say that it is accelerating (or changing its velocity with time). Note that an object's speed can be constant, but if it is changing its direction, then the velocity vector is changing. Therefore, acceleration can be defined by comparing the difference in velocities at two different times. In a method similar to that used to calculate velocity, we can write this difference as

$$\vec{a} = \frac{(\vec{v}_2 - \vec{v}_1)}{(t_2 - t_1)} = \left(\frac{\Delta\vec{v}}{\Delta t}\right),$$

where the symbol "*a*" denotes acceleration. Note also that acceleration is a vector; it has a magnitude and direction. Acceleration tells us how much an object's velocity is changing as a function of time. If we observe an object's velocity increasing with time, we say it is accelerating. If an object is decreasing, we often say it is decelerating.

### *Force*

What makes an object speed up, slow down, or change directions? For objects we typically use – chairs, books, coffee mugs – we provide a push or pull that moves them. The formal name for an action that causes an object to change its velocity (or acceleration) is force. Sir Isaac Newton, one of the greatest mathematicians and scientists of all time, developed the laws of motion that accurately describe the motion of almost everything – from coffee mugs to planets and stars. (The "almost" is used because Einstein later found that if things move really fast – near the speed of light – then Newtonian mechanics or Newton's laws of motion need to be modified. This modification is called special relativity.) One of Newton's laws defines force. Force is equal to the mass of an object times its acceleration ($\vec{F} = m\vec{a}$). (Note the arrows above the force and acceleration variables. Force is a vector and therefore has a magnitude and direction.) A force will accelerate a mass. One force we deal with every day of our lives is the gravitational force between the Earth and us. The mass of Earth "pulls" us down to the surface. Therefore, if I drop a pencil from my desk, it will fall straight down to the floor (the gravitational force vector points to the center of Earth). As the pencil falls, it accelerates continuously (going faster and faster) until it hits the ground. Near the surface of Earth, the magnitude of the gravitational acceleration is 9.8 m $s^{-2}$. This means that an object dropped from some height will accelerate (or speed up) by 9.8 m $s^{-1}$ every second (neglecting wind resistance). So after two seconds the object is moving 19.6 m $s^{-1}$ ($\sim$ 70 km $hr^{-1}$).

## 3.9 Problems

3.1 Estimate the time required for a parcel of solar wind with a speed of 800 km $s^{-1}$ to travel from the surface of the Sun to Earth. Is Carrington's observation of possible cause and effect of a solar flare and geomagnetic activity consistent with this time?

3.2 How long does it take electromagnetic radiation moving at the speed of light to reach Earth from the Sun?

3.3 The Parker spiral of the IMF depends on the speed of the solar wind. For a 400 km $s^{-1}$ solar wind, what is the angle of the IMF with respect to the Sun–Earth line?

3.4 Which parameters determine the shape and size of the heliosphere?

3.5 If a parcel of solar wind were radially accelerated at 10 km $s^{-1}$ from rest, how long would it take to be going faster than the escape velocity of the Sun?

# Chapter 4

# Earth's space environment

It has now become possible to investigate the region above the ionosphere in which the magnetic field of the earth has a dominant control over the motions of gas and fast charged particles . . . it may appropriately be called the magnetosphere.

> *T. Gold, from the paper that coined the term magnetosphere*
> *(*J. Geophys. Res.*, **64**, 1219–1224, 1959). From the American Geophysical Union.*

## 4.1 Key concepts

- magnetic field
- magnetosphere
- Van Allen radiation belt
- magnetic reconnection
- geomagnetic storm

## 4.2 Introduction

At approximately 100 km (or about 60 miles) above Earth's surface, the amount of ionized gas becomes appreciable. Because ionized gas is electrically charged, it feels the effect of Earth's magnetic field, which plays an important role in guiding the motion of charged particles in near-Earth space. Through its interaction with the magnetized solar wind, Earth's magnetic field is intimately involved in coupling or transferring of energy and momentum from the Sun into our space environment. This chapter describes the magnetic field region surrounding Earth called the magnetosphere. The connection of the magnetosphere with the Sun is at the heart of space weather.

## 4.3 Dipole magnetic field

**Magnetic fields** are force fields around magnets, electric currents, or moving charged particles that exert a force on other magnets, electric currents, or moving charged particles. Due to the motion of molten iron inside Earth, a relatively strong magnetic field surrounds it.

Like the magnetic field in sunspot pairs or magnets, Earth's magnetic field emerges from one hemisphere with a certain direction and

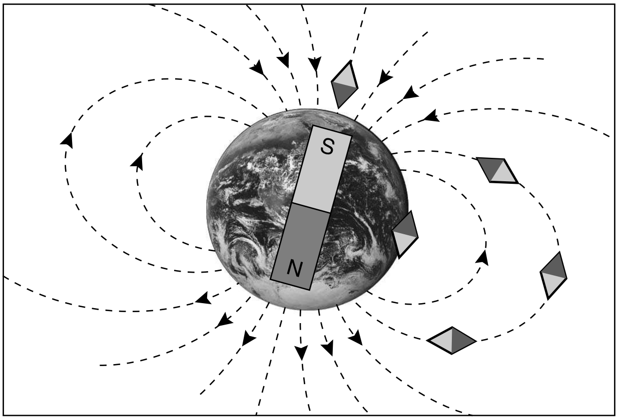


**Figure 4.1** Earth has a dipole magnetic field with the same shape as that of a regular magnet (from NASA StarGazers).

points towards the opposite hemisphere. In general, for magnetic fields emanating from things like magnets, the north pole is defined as the pole where the magnetic field points outward and the south pole where it points inward. Imagine a magnet inside Earth with the north magnetic pole pointing south and the south magnetic pole pointing north (see Figure 4.1). This is the configuration of Earth's magnetic field. To avoid confusion between Earth's north geographic pole and its magnetic north pole, the north magnetic pole is defined to be in the northern hemisphere. This potentially confusing difference between Earth and a magnet is done so that Earth's north magnetic pole is in the same hemisphere as its north geographic pole. We call the point where the magnetic field emerges straight out of Earth the south magnetic pole and the point where it goes directly into Earth the north magnetic pole. Such a magnetic field is referred to as a dipole ("di" being Latin for two). Earth's magnetic poles are not located in the same place as the geographic north and south poles, which are defined by Earth's spin axis. The magnetic dipole axis is tilted by about 11° with respect to the spin axis.

In Figure 4.1, the direction of each field line is indicated; the spacing between the lines represents the strength of the field. Note that the field is stronger at the poles than at the equator. The strength of a dipole magnetic field is two times greater at the magnetic pole than at the equator and falls off very quickly with distance; the strength of the field at the equator decreases as the cube of the distance:

$$\left(|B| \propto \left(\frac{1}{r^3}\right)\right).$$

Satellites with instruments that measure the strength and direction of magnetic fields have explored much of the space around Earth and all of the planets. These missions have verified the dipole nature of Earth and the other planets with magnetic fields (Mercury, Jupiter, Saturn, Uranus, and Neptune).

## 4.4 Structure of the inner magnetosphere

Figure 4.2 shows a cross section of the **magnetosphere** in the noon–midnight meridian, with north at the top and the Sun on the left. The regions of the magnetosphere are labeled. Note that the magnetic field resembles a dipole close to Earth. The dipole region of Earth's magnetosphere is called the inner magnetosphere. On the nightside at about geosynchronous orbit (6.6 Earth radii ($r_E$) from the center of Earth), the magnetic field lines become stretched into a long, tail-like configuration. The interaction of Earth's magnetic field with the solar wind is responsible for the distortion of its dipole field. The non-dipolar regions are called the outer magnetosphere.

Immediately surrounding Earth is a region of cold (about 1 electron-volt, eV), dense (tens to thousands of particles per $\text{cm}^{-3}$) plasma that essentially co-rotates with Earth. This region is called the plasmasphere. An eV is a measure of kinetic energy. For a proton, 1 eV corresponds to a velocity of about 14 $\text{km s}^{-1}$. Densities in space are much lower than at Earth's surface. The density of air at sea level is approximately Avogadro's[1] number. In the densest region of the magnetosphere – the plasmasphere – densities are billions and billions of times lower.

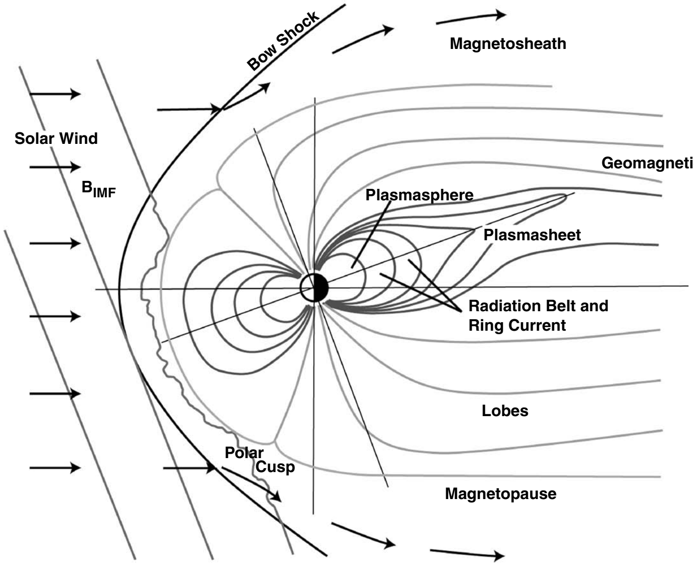


**Figure 4.2** A noon–midnight cross section of Earth's magnetosphere. Note the dipole shape of the inner magnetosphere. The Sun (hence noon) is to the left and north is up.

---

[1] Avogadro, Amedeo (1776–1856), Italian scientist and one of the founders of physical chemistry whose hypothesis that a molar volume (a volume whose mass is one gram molecular weight) of any gas contains the same number of atoms or molecules. This number, now called Avogadro's number, is equal to $6.022\,045 \times 10^{23}$.

The plasmasphere consists mostly of hydrogen and helium, but also an appreciable amount of oxygen, that have just enough energy to escape from Earth's ionosphere. The ionosphere, created by solar ultraviolet and X-ray radiation, will be discussed in much more detail in Chapter 6. As plasma drifts up the magnetic field line from below, it becomes trapped and co-rotates with Earth. There is often a very sharp boundary to the dense plasmasphere called the plasmapause. Plasma density frequently drops an order of magnitude within a very short radial distance (less than 0.5 $r_E$).

Often overlapping with the plasmasphere are the Van Allen[2] radiation belts and the ring current. These two regions are characterized by high-energy particles that are trapped in Earth's magnetosphere. The ring current is made up of particles with a peak energy of about 200 KeV, while the radiation belts consist of particles with energies extending into the relativistic regime. Relativistic particles have velocities near the speed of light and carry tremendous amounts of kinetic energy.

The ring current is so named because its charged particles produce an electric current that encircles Earth. Figure 4.3 is a schematic of

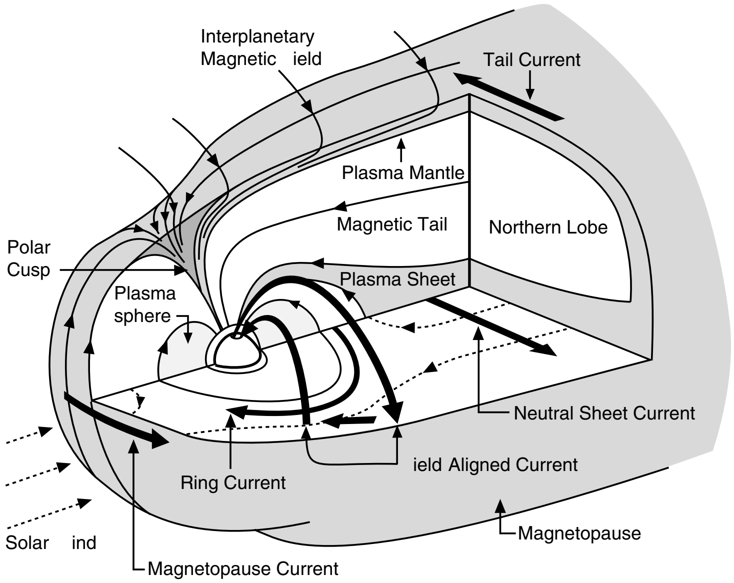


**Figure 4.3** A schematic of Earth's magnetosphere showing the equatorial and noon–midnight meridional planes. The electric currents flowing in the magnetosphere are shown as dark arrows. The regions of the magnetosphere are labeled (adapted from Figure 1.18 in Kivelson and Russell, 1995).

---

[2] Van Allen, James Alfred (1914–2006), American physicist and pioneer in the early development of the space program whose instrument on Explorer I (the first US satellite) discovered that Earth is surrounded by belts of trapped ionized high-energy particles. The region is now called the **Van Allen radiation belts**.

the magnetosphere showing both the noon–midnight meridian and the equatorial plane. The solid arrows indicate the directions of the different currents flowing in the magnetosphere. Because of the shape and strength of Earth's dipole magnetic field region, energetic ions flow from midnight to the dusk side, and energetic electrons flow in the opposite direction. This difference in flow directions of positively charged ions and negatively charged electrons gives rise to an electric current, a ring current that circles Earth. This ring current in turn gives rise to a magnetic field that points in the opposite direction to the dipole field at Earth's surface. Therefore, the ring current decreases the strength of Earth's magnetic field as measured on the surface. We use instruments near the equator to constantly measure the strength of the magnetic field. When the ring current intensifies suddenly, we see a rapid decrease in magnetic field strength. A magnetic index, called the Disturbed Storm Time Index (or Dst), measures the deviation or change in Earth's magnetic field from its normal quiet time value, the strength of Earth's internal magnetic field. When this index goes negative (indicating a decrease in Earth's field), it is due to intensification or increase in the strength of the ring current.

Note that in Figure 4.3 there are other currents, called field-aligned currents, that connect the ring current and plasma sheet to the ionosphere. These currents play a major role in aurora and other space weather phenomena.

The radiation belts, named after their discoverer James Van Allen, consist of two distinct regions of energetic particles. The outer belt, composed mostly of energetic electrons, has its inner edge around 3 $r_E$ and its highly variable outer edge usually just beyond geosynchronous orbit. The inner belt, which consists of energetic electrons and protons, extends out to about 2.5 $r_E$. The region between the belts (called the "slot") is generally kept clear of energetic particles by mechanisms that enhance the loss of the particles into the ionosphere. Plate 6 is a 3D schematic of the doughnut- or torus-shaped radiation belts. The radiation belts contain intense radiation that can kill astronauts and damage or destroy sensitive electronics on spacecraft. Understanding this region is one of the main efforts of space weather since many important satellites have their orbits in or through the radiation belts.

## 4.5 Interaction of the solar wind and magnetosphere

Within the magnetosphere the dynamics of charged particles (the plasma) are determined by the configuration of Earth's magnetic field, which looks less and less like a dipole farther from Earth because of its

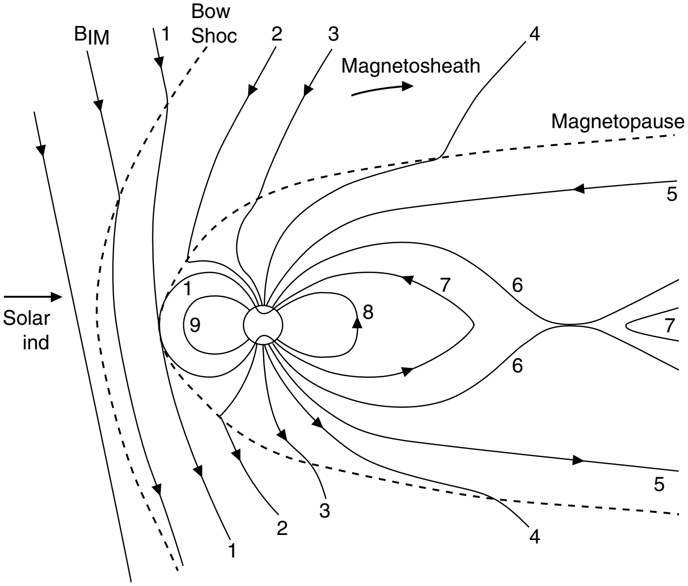


**Figure 4.4** A two-dimensional noon–midnight meridional cross section of Earth's magnetosphere showing magnetic reconnection on the dayside and in the magnetotail (adapted from Figure 9.11 in Kivelson and Russell, 1995).

interaction with the magnetized solar wind. The interaction of Earth's magnetic field with the magnetized solar wind is similar to that of a rock in a stream. The solar wind (stream) encounters Earth's magnetosphere (rock) as an obstacle and moves around it, leaving a wake behind. In the case of Earth and the solar wind, the interaction produces a long magnetotail. Figure 4.4 is a schematic of Earth's magnetosphere.

Because the solar wind is supersonic, a shock wave is formed upstream or on the dayside of the magnetosphere. This shock wave is called the bow shock. The bow shock slows the solar wind and begins to divert it around the magnetosphere. The region between the bow shock and the magnetosphere is called the magnetosheath.

The magnetopause is the boundary of the magnetosphere. The position of this boundary depends on the strength of the solar wind pressure, which is primarily due to solar wind density and velocity. As solar wind pressure increases, it moves the magnetopause earthward. When solar wind pressure decreases, the entire magnetosphere expands. The location of the magnetopause is determined by a balance between the solar wind pressure and the magnetic pressure of the magnetosphere.

## 4.6 Magnetic reconnection

The above discussion treats the solar wind like a fluid (water in a stream moving around an obstacle). However, in the case of the solar wind, the fluid is magnetized, which makes interaction between the solar wind flow and Earth's magnetosphere a little more interesting. When two magnetic fields are brought together, the fields combine. So if you bring two magnets close to each other and measure the strength of the field at some point, you would measure contributions from both. However, magnetic fields have direction as well as magnitude, and therefore if a magnetic field points in one direction and another in the opposite direction, the two fields subtract from each other. When this occurs in a magnetized plasma, such as in the solar wind and magnetosphere, the fields can interact in a new way – to form new field lines. In this process, called **magnetic reconnection**, energy is taken from the magnetic field and put into particle motion (magnetic energy is converted to particle kinetic energy). This process, which has been reproduced inside the laboratory, powers much of the activity observed on the surface of the Sun.

Magnetic reconnection occurs when a magnetized plasma parcel (called a flux tube) of one polarity is brought up against a magnetized plasma parcel of opposite polarity. When magnetic reconnection occurs, the field lines connect and change their topologies or connectedness. Figure 4.4 shows examples of how field topology changes on the dayside magnetosphere. This figure shows an idealized magnetopause with the southward magnetic field of the solar wind (line 1′) brought up against the northward-directed magnetic field of Earth (line 1). Note that there are two distinct field lines, one that has both ends in the solar wind and one connected to both poles of Earth. When they come together, they can reconnect, and in addition to converting some of the magnetic energy into particle kinetic energy, the two original field lines' topologies are converted into two new field line topologies. Field lines (lines 2 and 2′) still exist, but one end of line 2 is connected to the north pole of Earth and other end in the solar wind, and one end of line 2′ is connected to the south pole of Earth and the other end in the solar wind. Field lines with both ends connected to Earth are called "closed", and those with one end connected to Earth and the other in the solar wind are called "open". Plasma can become "trapped" on closed field lines, and therefore, densities can build up. The plasmasphere and radiation belts are found on closed field lines. Open field lines generally have much less plasma since the plasma can stream along the magnetic field away from Earth.

## 4.7 The magnetotail

Reconnection of the Sun's and Earth's magnetic fields creates open field lines with one end attached to Earth and the other end extending into interplanetary space. Because the solar wind end of this magnetic field moves away from the Sun with the rest of the solar wind, the field line gets swept back behind Earth. This is similar to a person seated in a moving convertible car whose hair is swept back in the direction of the wind. Earth's magnetic field is swept back (lines 3, 4, and 5 in Figure 4.4) into a long, cylindrically-shaped region called the magnetotail. The magnetotail consists of two magnetic lobe regions, one tied to the north polar cap (represented by line 5) and the other connected to the south polar cap (line 5′). These lobes contain "open" field lines of oppositely directed magnetic field – with the north lobe magnetic field pointing toward Earth and the south lobe magnetic field pointing away from Earth. These two lobe regions are separated by the plasma sheet – a region of higher plasma density than the lobes. This region carries the current that separates the two lobe fields called the cross tail neutral sheet. These regions are labeled in Figures 4.2 and 4.3.

## 4.8 Plasma sheet convection

The solar wind imparts an electric field across the magnetosphere that is directed from dawn to dusk in the plasma sheet (the same direction as the neutral sheet current shown in Figure 4.3). This electric field causes the plasma sheet flux tubes to move earthward in a motion called convection that completes the cycle brought about by magnetic reconnection on the dayside. Figure 4.4 shows the complete convection cycle or motion of a flux tube throughout the magnetosphere starting when a solar wind flux tube (line 1′) is first reconnected on the dayside magnetosphere (line 1). Because one end of the flux tube is connected to the solar wind, it is swept back over the polar cap (lines 2, 3, and 4). As the flux tube reaches Earth's nightside, the field line becomes part of the magnetotail (line 5) and is convected towards the central plasma sheet (line 6). At that point, magnetic reconnection takes two oppositely directed field lines – one from the north (line 6) and one from the south (line 6′) – and makes two new field lines (lines 7′ and 7). Line 7 on the earthward side of the reconnection site has both of its ends attached to Earth, while line 7′ now has both ends connected to the solar wind. Note that flux tubes represented by lines 7 and 7′ have the same magnetic topologies as lines 1 and 1′ at the start of the process. Line 7 is then convected earthward to the dayside magnetosphere, where it can participate in the convection cycle again.

Figure 4.5 shows the foot of the field lines projected into the polar cap of Earth. The figure represents a view down onto the polar cap with the numbers representing the position of the foot of each of the numbered field lines shown in Figure 4.4. The "convection cells" represent the motion of the field lines from the dayside over the polar cap and then back to the dayside at lower latitude. This convection motion, which is

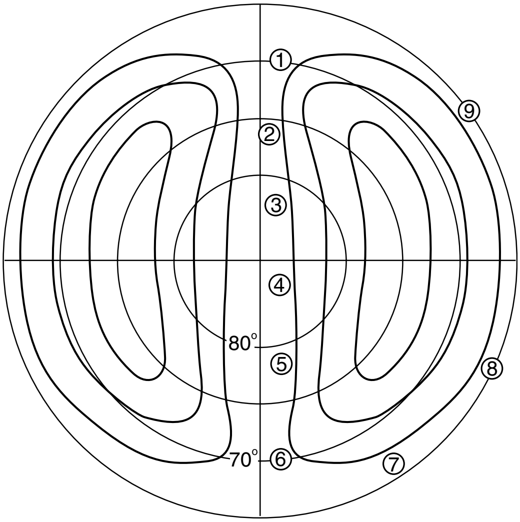


**Figure 4.5** The projection of the magnetic field lines shown in Figure 4.4 onto the ionosphere. The lines show the direction of plasma motion in the ionosphere due to magnetospheric convection (adapted from Figure 9.11 in Kivelson and Russell, 1995).

observed in the ionosphere, is the main dynamics of the ionosphere in the polar cap. The ionosphere and its dynamics are discussed more fully in Chapter 6.

## 4.9 Dynamics of the magnetosphere

Reconnection (and hence convection) in Earth's magnetosphere is not steady. Enhanced reconnection at the dayside brought about by strong southward IMF can lead to increased energy coupling and increase in the amount of magnetic flux transferred to the nightside. Increased magnetic energy density and pressure in the magnetotail lead to thinning of the current sheet, which enables magnetic reconnection to occur. This in turn converts the magnetic energy of the tail into the plasma kinetic energy associated with rapid flows observed in the magnetotail. In addition, an enhanced cross-tail electric field leads to enhanced convection into the inner magnetosphere. This changes both the motion of particles and the location of the plasmapause; the plasmapause moves earthward with increased convection and away from Earth with decreased convection. Reconnection at the dayside requires a southward component to the IMF. Since IMF polarity points north and south irregularly, the amount of reconnection and hence energy input to the magnetosphere is highly variable.

### 4.9.1 Storms

Occasionally, the amount of energy transferred into the magnetosphere from the Sun can increase rapidly. Such an increase is often associated

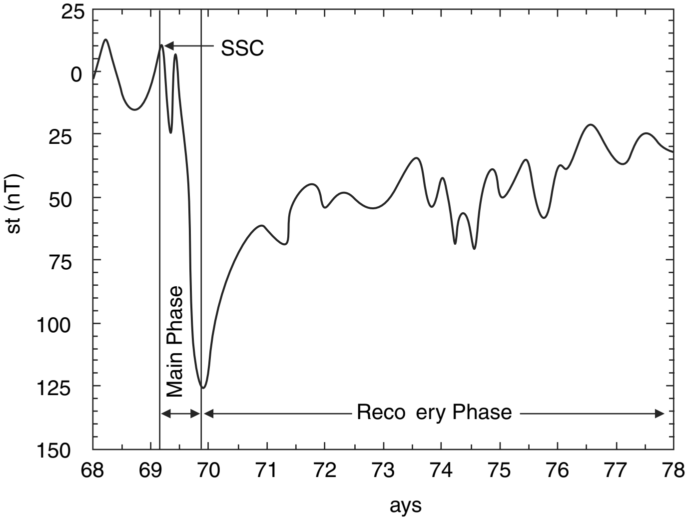


**Figure 4.6** The Disturbed Storm Time (Dst) index over a ten day interval showing the characteristic phases of a geomagnetic storm.

with the impact of Earth with southward IMF from a coronal mass ejection (CME). The southward IMF can reconnect with the northward field of Earth. This energy can increase the flow of energy and momentum into the magnetosphere and the rate of convection. Associated with this increased energy input is enhancement of the ring current. This is observed as a rapid decrease in the Dst. Figure 4.6 shows a time history of Dst during a storm. A **geomagnetic storm** is often characterized by three phases: sudden storm commencement (SSC), main phase, and recovery. The SSC is characterized by an enhancement of the Dst due to an increase in and motion earthward of the magnetopause currents (named the Chapman–Ferraro[3] currents for the scientists who first hypothesized their existence and their role in geomagnetic storms (see Chapter 1)). As the CME and increased solar wind dynamic pressure hit Earth, the magnetopause moves earthward. Chapman–Ferraro currents also increase to stand-off the solar wind by increasing the magnetosphere's magnetic pressure. The direction of the Chapman–Ferraro currents is such as to cause an increase in Earth's magnetic field as seen from the dayside low-latitude surface. Hence the Dst increases. This enhancement generally lasts for tens of minutes to hours, when suddenly the rapid increase in the ring current swamps the Chapman–Ferraro current signal and the Dst rapidly drops, signalling the main phase of the storm. The drop usually lasts several hours, at which time the Dst value

---

[3] Chapman, Sydney (1888–1970), British geoscientist who made significant theoretical contributions to terrestrial and interplanetary magnetism, the ionosphere and the aurora borealis. His ionospheric theory, called Chapman theory, explains the main structure of planetary ionospheres. Ferraro was a student of Chapman.

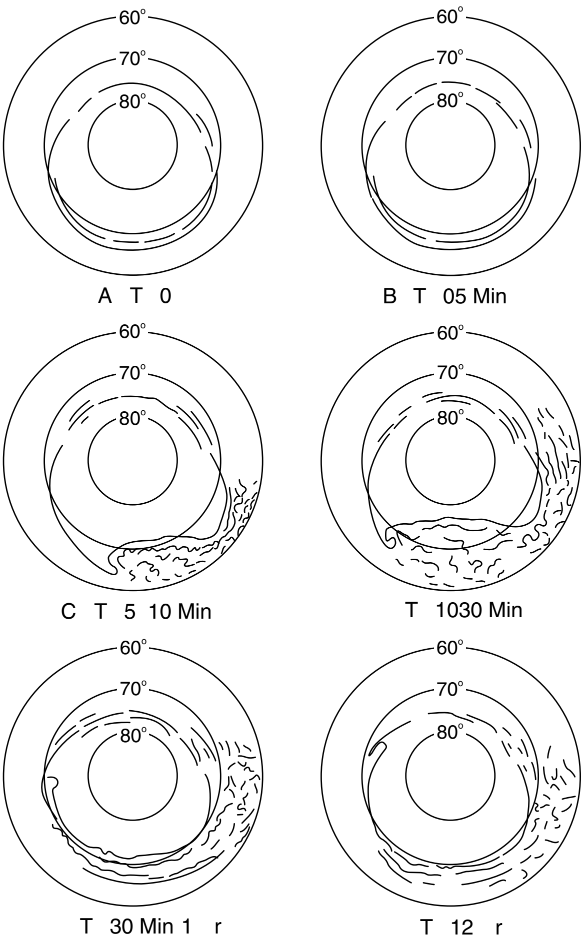
begins a slow recovery to pre-storm levels. The recovery phase can last for many days. The number of storms varies throughout the solar cycle, but typically is on the order of a few per month, with a great number and intensity of storms during solar maximum.

Associated with every storm is a rapid brightening and expansion of the entire auroral oval in both the northern and southern hemispheres. In many storms, the radiation belts intensify and the inner edge of the outer radiation belt moves earthward. In some major storms, the slot region can be completely filled and the inner belt can intensify dramatically. The overall level of geomagnetic disturbance measured by another geomagnetic index called Kp also increases. Kp measures the overall variability of Earth's magnetic field observed at mid-latitudes. It is a logarithmic-scale (like the Richter Scale for earthquakes) and goes from 0 (no activity) to 9 (major storm activity). The average, or most probable, Kp level is about 3.

### 4.9.2 Substorms

Another smaller, but much more common, disturbance in Earth's magnetosphere is called a substorm. It is so named because it was originally thought that a collection of substorms makes a storm. Substorms occur much more often than storms – four times per day on average. Substorms are defined by auroral behavior. During a substorm, the most equator-ward existing auroral arc suddenly brightens and expands poleward and westward. The enhanced aurora is associated with enhanced auroral ionospheric currents that are measured by a magnetic index called the Auroral Electrojet (AE) index. The AE index is a measure of the difference between the strength of two ionospheric current systems, the westward and eastward electrojets. These electrojets, but especially the westward electrojet near midnight, are intensified by the onset of a substorm. An enhancement of the westward electrojet leads to a decrease in the horizontal field at Earth's surface directly underneath the aurora, while an enhancement of the eastward electrojet gives rise to an increase in Earth's magnetic field directly under the aurora.

Also associated with substorms are a number of other signatures including the presence of magnetic waves called Pi2 and the sudden enhancement of energetic particles at geosynchronous orbit. Figure 4.7 shows a schematic sequence of a substorm as seen in the global aurora. The panels show the polar cap from above, with midnight at the bottom of the panel. Note that the auroral disturbance begins near midnight and then expands poleward, eastward, and westward.



A T 0

B T 05 Min

C T 5 10 Min

T 1030 Min

T 30 Min 1 r

T 12 r

**Figure 4.7** Configuration changes in the aurora, viewed looking down onto the polar cap, during a substorm. Noon is at the top and midnight at the bottom. Note that activity is concentrated on the nightside. Reprinted from *Planetary and Space Science*, **12**, S.-I. Akasofu, The development of the auroral substorm, 273–282, Pergamon Press, Ltd., 1964, with permission from Elsevier.

There are three main differences between storms and substorms: (1) their timescales – storms occur much less frequently and last for days, while substorms are common, occurring many times per day and generally having timescales of an hour; (2) their spatial extent – storms have global manifestations throughout the magnetosphere, while substorms are generally more localized to the nightside of Earth; and (3) their magnetic signature – by definition, storms are accompanied by an intensification of the ring current, whereas substorms are not. What adds to the complication and interest of storms and substorms is that all storms are accompanied by substorms, but not all substorms are associated with storms. Current research is attempting to distinguish the relationship between these two phenomena. For space weather research, the geomagnetic storm is of prime importance.

## 4.10 Supplements

Electricity and magnetism, which are intimately related, represent the two components of light and all other forms of electromagnetic radiation. The relationship between them can be described by a set of four equations called Maxwell's equations. For simplicity, electricity and magnetism will be discussed independently.

### 4.10.1 Electrostatics

Electric charge is an intrinsic property of matter. The sub-atomic building blocks of atoms are protons, neutrons, and electrons. Protons have a discrete positive charge; electrons have an equal but opposite negative charge; and neutrons have no charge (and hence are neutral). Most matter that we are familiar with on Earth (like the book you are holding and even you) is made up of atoms and molecules that have equal numbers of electrons and protons and hence are neutral. Electrons and protons are attracted to each other by their respective electric charge. This is the electric force or Coulomb[4] force that describes the force between electrically charged objects. Opposite (positive and negative) charges attract and like (negative–negative and positive–positive) charges repel. To follow a convention first proposed by Benjamin Franklin,[5] electrons carry negative charge and protons positive charge. The strength of the electric force is dependent on the net amount and sign (positive or negative) of charge and inversely proportional to the square of the distance between the charged objects:

$$F_{\mathrm{c}} = \frac{(kq_1q_2)}{r^2}.$$

where $k$ is a constant of nature, $q$ is the amount of net charge and $r$ is the distance between the objects.

Example

(a) What are the magnitude and direction of the electric force between an electron and proton in a hydrogen atom? ($k = 8.99 \times 10^9$ Nm$^2$/C$^2$, the charge on a proton is $1.60 \times 10^{-19}$ C, and the charge on an electron is equal, but with a negative sign. The average distance between the electron and proton in a hydrogen atom is $0.530 \times 10^{-10}$ m).

---

[4] Coulomb, Charles (1736–1806), French physicist who developed the theoretical understanding of electric charge. The SI unit of charge (coulomb – C) is named in his honor.

[5] Franklin, Benjamin (1706–1790), One of the founders of the United States and scientist whose work included proving the electrical nature of lighting (with his famous – and dangerous – kite flying during a thunderstorm experiment).

(b) If an electron mass is $9.11 \times 10^{-31}$ kg, what is the acceleration of the electron due to this force? (Recall Newton's law that states $F = ma$).

Answer

(a) $$F = k\frac{q_1 q_2}{r^2}$$

$$= 8.99 \times 10^9 \frac{\text{Nm}^2}{\text{C}^2} \frac{\left(1.60 \times 10^{-19}\text{C}\right) \times \left(-1.60 \times 10^{-19}\text{C}\right)}{\left(0.530 \times 10^{-10}\text{m}\right)^2}$$

$$= -8.19 \times 10^{-8}\ \text{N} \quad \text{toward each other}$$

(opposites attract).

(b) $$F = ma$$

$$a = \frac{F}{m}$$

$$= \frac{-8.19 \times 10^{-8}\text{N}}{9.11 \times 10^{-31}\ \text{kg}}$$

$$= -8.99 \times 10^{20}\ \text{m s}^{-2}$$

which is a huge acceleration.

### 4.10.2 Magnetostatics

Magnetism has been known for millennia since lodestones, which are naturally occurring magnets, were discovered in a region that is now part of western Turkey called Magnesia (hence the name magnet). Magnets have the property of attracting or repelling one another and being attracted to certain metals such as iron. As mentioned in Chapter 1, the first description of a compass was written in the eleventh century, and its similarity to a magnet was discovered in the years just prior to 1600.

All magnets attract iron, but because magnets have two poles, they either attract or repel each other. As discussed earlier in this chapter, magnetic poles are named north (+) or south (−) and opposite poles attract, like poles repel. Unlike electric charges (that come in either + or − varieties), magnetic "charges" always come together in a pair. It is impossible to separate the north pole from the south pole (i.e., cutting a magnet in half will not give you a separate north and separate south pole, but will give you two smaller magnets each with both a north and a south pole). Even at the atomic and sub-atomic level, magnetic particles are dipoles, containing both polarities. A dipole magnetic field has a strength that falls off faster with distance than an electric field, whereas an electric field falls off as an inverse square of distance $\left(E \propto (1/r^2)\right)$. Dipole magnetic fields fall off as the inverse cube of the distance, i.e., $B \propto 1/r^3$.

Experiments conducted by Oersted[6] in the late 1800s demonstrated that electric currents deflect a compass needle. This had immediate impact for space weather, as now there was a physical understanding of what could cause geomagnetic storms and deflection of compass needles by aurorae. This connection between magnetism and currents is what makes electromagnets and generators possible.

We now understand that all magnetism has as its source electrical currents. Within magnets, the currents are due to the alignment of individual atoms into domains. If these domains are ordered (aligned in a coherent manner with all the north poles pointing in the same direction), then the material is magnetized and will have a magnetic field associated with it. Electromagnets can be made by looping wire in such a way that the current flows in loops in the same direction. These current loops give rise to a magnetic field that is similar to a magnet – both have a north and south pole, and you cannot cut the loops into two to make a stand-alone north or south pole.

A magnetic field is a force field, similar to an electric field or a gravitational field, which exerts a force on an object at a distance from its source. Electric fields exert forces on electrically charged objects and particles. Gravitational fields exert forces on things with mass, while magnetic fields exert forces on other magnets and moving electrically charged objects or particles. Note the caveat with regards to electrically charged objects – they must be moving. One property of a moving charge is that it gives rise to an electrical current. Hence magnetic fields exert a force on a current.

### 4.10.3 Single particle motion in a magnetic field

A single charged particle moving in a magnetic field will feel a force due to that magnetic field. The force is not in the direction of the field, as with gravitational fields and electric fields, but in a direction that is perpendicular to both the direction of motion and the field. This is expressed mathematically as a vector cross product:

$$\vec{F} = q\vec{v} \times \vec{B}.$$

---

[6] Oersted, Hans Christian (1777–1851), Danish physicist who in 1820 demonstrated the connection between electricity and magnetism. This discovery led André Ampère and Michael Faraday to their major theoretical understanding of electromagnetism.

Note the vector notation (arrows) above the three parameters – force, velocity and magnetic field (magnetic field is written as $B$ since $M$ is used for another magnetic quantity called magnetization and $m$ is used for mass). The symbol $q$, which represents the charge on the particle, is a scalar that is either positive or negative depending on the sign of the charged particle. This cross product states that the direction of the force exerted on a moving charged particle in a magnetic field is perpendicular (or at right angles to both the velocity and magnetic field) and the magnitude is directly proportional to the component of the velocity that is perpendicular to the magnetic field. In other words, the magnitude of the cross product can be written $F = qvB \sin\theta$, where $\theta$ (Greek letter theta) is the angle between the velocity vector and the magnetic field vector. Note that if the velocity vector of the charged particle is along the magnetic field, the magnitude of the force is zero since $\sin(0) = 0$. The magnitude of the force is a maximum when the direction of the velocity of the charged particle is perpendicular to the magnetic field. The direction can easily be found using the "right-hand rule" (see Figure 4.8). This rule states that you point your right-hand index finger (one closest to thumb) in the direction of motion, point your middle finger in the

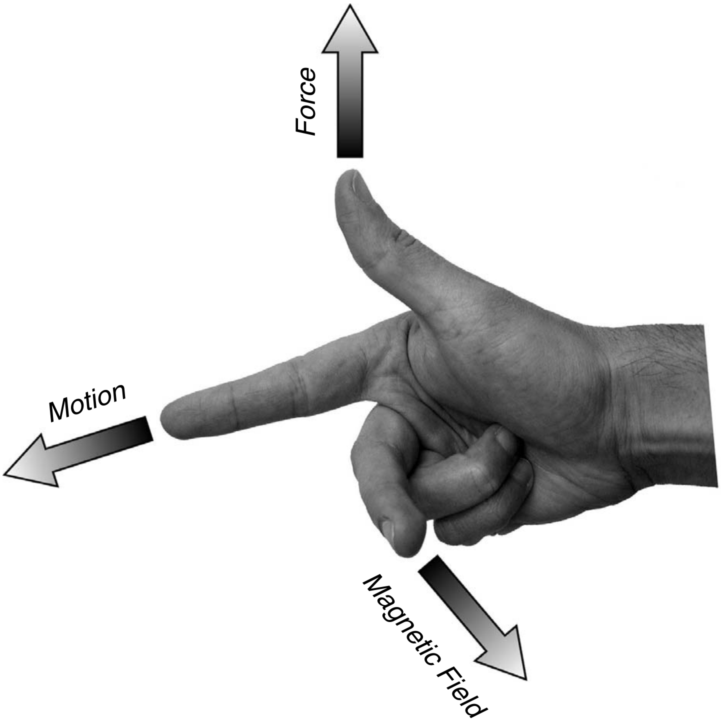


**Figure 4.8** The Lorentz force is described by a vector cross-product $\vec{F} = q\vec{v} \times \vec{B}$. The right-hand rule helps determine the direction of the resulting force.

direction of the magnetic field, and your thumb points in the direction of the force (for negative charges this is in the opposite direction).

Note that the force will cause the particle to be accelerated (its velocity vector will change with time) because there is a force pushing it perpendicular to its original velocity direction. This causes the particle to circle or spiral around the magnetic field line. Particles with a velocity vector along the magnetic field line will not feel the magnetic force and travel along the magnetic field. Particles with a velocity vector perpendicular to the magnetic field will circle around the field line. Particles with part of their velocity vector perpendicular and part parallel to the field line will spiral around the magnetic field line in a helical trajectory.

### 4.10.4 Magnetohydrodynamics

Space plasmas are a collection of different ionized elements (i.e., hydrogen, helium, oxygen, etc.) and equal numbers of electrons. Each of these particles can have its own velocity, charge, mass, and energy. Air in the atmosphere of Earth or in the room in which you are reading is made of neutral gases (nitrogen and oxygen primarily), each of which has its own velocity. One way of determining the motion of this collection of particles is to follow each individual particle in its path around the room, keeping track of its acceleration, instantaneous velocity, and position. This is a monumental task if you consider the number of particles in any given volume of space. Fortunately, gases and plasmas often behave in a collective way called the fluid approximation. Fluids move collectively – you can describe the motion of a river in terms of its velocity at a point instead of having to measure each and every molecule of water at that point. Because water molecules act collectively on scales large compared to their size, we can follow the motion with a set of simple equations that keep track of a volume of the fluid rather than the individual particles. The most important of these are the continuity equation and the momentum equation. The continuity equation is a statement that if the density of a fluid changes with time in a given volume, it can be due to two different processes – fluid has moved in or out of the volume (called transport) or fluid was created or destroyed in the volume (source or loss). The momentum equation describes the motion (or transport) of the fluid and describes how the velocity of a parcel of fluid changes with time. Since velocity can change with time (acceleration) because of an applied force, this equation essentially lists all the forces that act on the fluid parcel.

This treatment can be done for magnetized plasmas as well. But, unlike neutral fluids such as water in a stream, plasmas are made up of charged particles. Therefore, the forces felt by the plasmas include electric and magnetic forces. The simplest approximation for a plasma treats the entire collection of particles (ions and electrons) as a single fluid. Referred to as the magnetohydrodynamic (MHD) approximation, it describes the large scale dynamics of plasmas (such as the motion of the solar wind throughout the heliosphere and the plasma inside Earth's magnetosphere) very well. The main forces that act on a parcel of plasma are the pressure gradient force discussed in the next chapter more fully and the magnetic force discussed in the previous section.

## 4.11 Problems

4.1 Which magnetic index is most often used to define when a geomagnetic storm occurs? What magnetospheric current does this index primarily measure?

4.2 Name three differences between geomagnetic storms and substorms.

4.3 What is the dipole magnetic field strength at the equator at geosynchronous orbit? (The equatorial field at the surface is 30 000 nT).

4.4 Plasma is tied to magnetic flux tubes in MHD. What is the direction of the electric field required for plasma in the plasmasphere to co-rotate with Earth? What is the direction of the electric field in the plasma sheet needed to convect plasma sunward? ($\vec{E} = -\vec{V} \times \vec{B}$).

4.5 What is the frequency of motion of a typical ring current ion at an $L$ of 4? How does this change with distance from Earth? (Cyclotron frequency = $w_{\mathrm{c}} \equiv |q|B/m$ ($q$ is the charge, $B$ is the magnetic field strength, and $m$ the mass of the ion).

4.6 Estimate the location of the magnetopause at the subsolar point (a point directly between the Sun and Earth). Use the fact that Earth's magnetic field strength falls off as a dipole and realize that the magnetopause will be where the solar wind pressure $\left(\rho v^2\right)$ balances the magnetic field pressure $\left(B^2/2\mu_0\right)$ of Earth's magnetosphere (where $\mu_0 = 4\pi \times 10^{-7}$ H m$^{-1}$).

# Chapter 5
# Earth's upper atmosphere

*NY Times* Headline: WIRELESS SIGNALS ACROSS THE OCEAN; Marconi Says He has Received Them From England. Prearranged Letter Repeated at Intervals in Marconi Code. The Italian Inventor Will Now Leave St. John's, N.F., and Will Go to Cornwall to Continue the Transatlantic Experiments from His Station There.

New York Times. *New York, NY, Dec. 15, 1901.*

## 5.1 Key concepts

- satellite radio communication and navigation
- ionosphere
- aurora
- photoionization

## 5.2 Introduction

Earth's upper atmosphere plays an important role in ground-based and **satellite radio communication and navigation**, and its density determines the lifetime of low-Earth orbiting (LEO) satellites. The upper atmosphere is composed of primarily neutral atoms and molecules in a region called the thermosphere. Within the thermosphere the amount of ionized gas becomes appreciable and forms a region called the ionosphere (see Figure 5.1). The thermosphere and ionosphere overlap in altitude, but because they describe two different particle populations (neutral and ionized), they are often "divided" since what influences the structure and motion of one usually does not necessarily directly drive the other. However, the two populations are coupled through particle collisions (neutral–ion interactions), which means that you usually cannot neglect one or the other. Since the thermosphere–ionosphere system is so important to radio wave propagation and LEO satellite lifetimes, it is one of the crucial areas of study for space weather.

## 5.3 The thermosphere

Why does Earth have an atmosphere, while other planets, such as Mercury and Earth's Moon, have essentially none? A strong force due to a pressure gradient tries to push any atmosphere near the surface of a

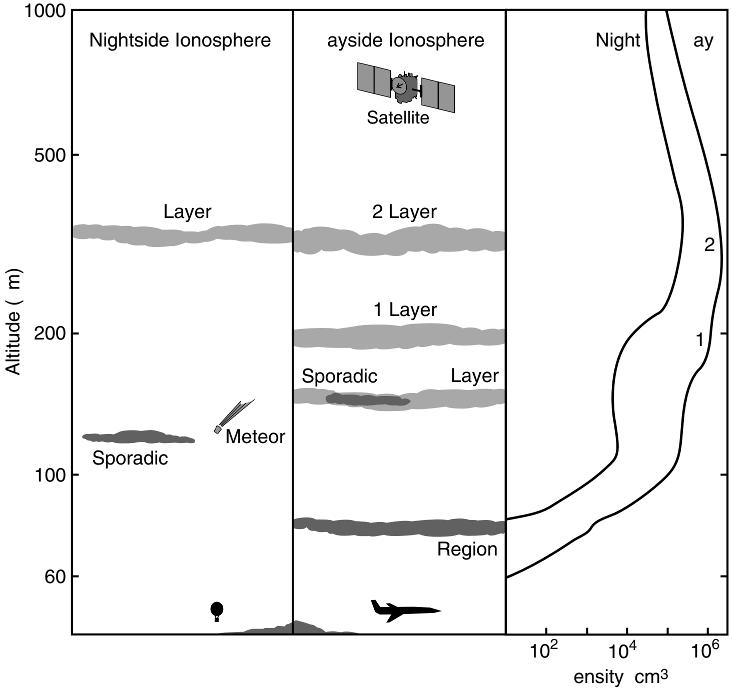


**Figure 5.1** The vertical structure of Earth's ionosphere. Note that the ionosphere  has several density peaks labeled as D, E, and F layers. At night the D region  normally essentially disappears, while the F region becomes a single layer (adapted from Radtel HF Radio Network).

planet or moon up and out into space. (Examples of this type of force are observed in a tire, balloon, or soda bottle. The pressure inside is higher than outside so there is a force wanting to push what is inside, air for the case of a balloon or tire, out.) Pressure is force per unit area; for a gas, it can be described by the familiar ideal gas law ($PV = nRT$ or $P = nkT$, where $P$ is pressure, $V$ is volume, $n$ is the number of moles of gas, $R$ is the gas constant, and $T$ is temperature. In the alternative form of the same equation, $n =$ the number of molecules per volume, and $k$ is the Boltzmann constant). The ideal gas law states that the amount of pressure a gas will exert on its surroundings is proportional to the amount of gas and its temperature. A gradient is a change in a quantity as a function of distance. A pressure gradient means that the pressure in one place differs from that at another nearby place. This can be expressed as a simple algebraic difference relation: $P_1(x_1) - P_2(x_2) =$ difference in pressure at two nearby points = pressure gradient. The difference in some quantity over a distance is given a special mathematical symbol, called the del operator, written as $\nabla$, so $-\nabla P =$ force. Note that the force of the pressure gradient points in the direction from high to low pressure (the reason for the negative sign is that the gradient points "uphill" always from low pressure to high pressure). On Earth the pressure near the surface is high compared to the relative vacuum of space, so there is a force trying to move the air near the surface into space (the phrase "nature abhors a vacuum" Spinoza (1677) is because of the pressure gradient force). So why doesn't Earth's atmosphere get pushed out into space? The simple answer is gravity. Earth's mass exerts a force on the mass of the atmosphere attempting to pull it down to the surface. The balance of the upward pressure gradient force with the downward gravitational force determines the atmospheric density structure. This relationship is called hydrostatic equilibrium. The prefix "hydro" is Greek for water or fluid and "static" means "not changing". The word equilibrium means that the two forces exactly balance. We can write this relationship as

$$\nabla P = -\rho g,$$

where $\rho$ is the mass density (units of mass per volume, or in SI, kg $\mathrm{m}^{-3}$), and $g$ is the gravitational acceleration (equal to 9.8 m $\mathrm{s}^{-2}$ at Earth's surface). Since pressure can be written in terms of density, $P = nkT = (\rho kT/m)$ (since $\rho = nm$, where $n$ is the number of molecules and $m$ is their average mass), the hydrostatic equilibrium equation can be written in terms of the gradient in mass density ($\rho$) or number density ($n$). The solution to this equation is that the mass density and number density fall off as a function of height. The density falls off in a special way – exponentially. Density as a function of height can then be written $n(\text{height}) = n_0 \exp(-\text{height}/H)$, where $H$, called the scale height, depends on the composition (or make up) of the gas (i.e., is it air, or pure oxygen, etc.), the temperature of the gas, and the acceleration of gravity, and $n_0$ is the density at the surface. The important thing to remember is that density falls off exponentially (quickly) with height. So with increasing altitude above Earth's surface, the amount of gas gets lower and lower.

Example: what fraction of sea-level air density is at the top of Mt. Everest? (Assume $H = 8$ km and height of Mt. Everest is 9 km.)

$$\begin{aligned} n &= n_0 \exp\left(-\frac{\text{height}}{H}\right) \\ \frac{n}{n_0} &= \exp\left(-\frac{9}{8}\right) = 0.32 \text{ or about } 1/3 \text{ the density at sea-level.} \end{aligned}$$

The Moon and Mercury are much smaller than Earth, and thus the effective gravity at their surfaces in much less. Therefore, the balance between the pressure gradient and gravity supports a much less dense atmosphere. Hence, if these Solar System bodies had a thick atmosphere at one time, much of it could have escaped over its long history because the gravitational force isn't strong enough to hold the atmosphere to Mercury or the Moon.

## 5.4 The ionosphere

When introducing the "spheres" in Chapter 1, we discussed briefly the temperature structure of Earth's atmosphere from the troposphere to the thermosphere and beyond. The upper atmosphere is usually defined as the region above 80 km. At this altitude the density of neutral particles is low enough that free electrons, which are created through the process of ionization, can survive for an appreciable amount of time before recombining with ions. Ionization is the process of making positively or negatively charged atoms or molecules by adding or stripping away one or more electrons. In Earth's upper atmosphere it is much more common to make positively charged ions by removing an electron than it is to make negatively charged ions by adding an electron. Ionization is accomplished when electrons are knocked free of their host ion by either solar high-energy photons (mostly UV and X-rays) or energetic particles that precipitate into the atmosphere and collide with the surrounding gas. In the traditional model of the atom (called the Bohr[1] model after the scientist who developed it), one or more electrons surround the nucleus, which is made of sub-atomic particles called protons and neutrons. Protons have positive charge, and electrons have negative charge of equal but opposite value. Opposite charges (positive and negative) have an attractive force called the electrostatic or Coulomb force, while the same charges (negative and negative or positive and positive) have a repulsive force. Almost all atoms and molecules in Earth's lower atmosphere are neutrals, meaning that there are equal numbers of protons and electrons in each atom. In the upper atmosphere, the number of charged particles (ions and electrons) becomes appreciable. At altitudes of about 300 km there is a peak in the number of free ions and electrons. The region surrounding this peak in electron density is called the **ionosphere**. Figure 5.1 shows the vertical structure of this region.

The production of the main part of the ionosphere is primarily due to solar electromagnetic radiation through a process called photoionization, and therefore the peak densities of the ionosphere are found on the dayside. However, at night the ionosphere does not completely go away since the recombination time of ions and electrons (the average amount of time needed for an ion and electron to come back together to form a neutral) is comparable to the rotation rate of Earth. The recombination

---

[1] Bohr, Niels Henrik David (1885–1962), Danish Nobel-Prize-winning physicist who developed the theory of atomic structure and explained the process of nuclear fission. The Bohr model of the atom describes the atom in terms of the nucleus (made of protons and neutrons) surrounded by orbiting electrons – analogous to a planetary system orbiting a star. Though quantum mechanics has changed our view of the atom, the Bohr model is still useful for understanding its basic structure.

rate is dependent on the background density; therefore it is high at low altitudes (where the density is high) and decreases with altitude along with the density. The amount of ionization that is present is determined by the balance of the source or production of the ions (photoionization) with the loss (recombination) of the ions.

## 5.5 Ionospheric structure

Photons of differing energies are able to penetrate and interact with atoms and molecules in Earth's atmosphere. Densities of atmospheric constituents (such as molecular nitrogen and hydrogen) also vary with height, and therefore the ionosphere forms a number of different regions at various altitudes above the surface of Earth. Figure 5.1 shows how the ionosphere is divided into different layers. Each region is characterized by a local maximum in the number density of ions. The D region, the lowest ionospheric layer, extends from approximately 50 to 90 km (therefore it extends down into the mesosphere – see Figure 1.1). The main sources of ionization in the D region are solar UV photons ionizing nitric oxide (NO) molecules. During solar maximum conditions, solar hard X-rays ionize air molecules (molecular nitrogen and oxygen). In addition, cosmic rays produce ionization at this altitude. Because the neutral density is relatively high in the D region, the amount of recombination is very great. Therefore, the D region is essentially only present during the day (though cosmic rays produce a residual level of ionization at night), and the level of ionization in the D region is the lowest of the different regions of the ionosphere. Solar storms can emit large amounts of X-rays that can cause rapid increases in D region ionization (sudden ionospheric disturbances (SIDs)). The D region is important with regard to high-frequency (HF) radio communication because it absorbs radio waves, which causes degradation of long-distance HF communication. During SIDs and intense polar cap precipitation of solar energetic particles, D region ionization can become so intense that HF radio communication is completely blacked out.

The next layer moving up in altitude is the E region (originally called the Kennelly–Heaviside or just the Heaviside layer). It extends from 90 to 120 km and is formed by both low energy (or soft) X-rays and UV solar radiation ionization of molecular oxygen ($O_2$). The peak density in the E region is over a 100 times greater than the peak density in the D region because recombination is less prevalent at these high altitudes. As with the D region, the E region decays away at night, which effectively raises its height as the faster recombination times decay the E region away more quickly at low altitudes than at higher altitudes. Besides solar photons, E region ionization also occurs due to energetic particles precipitating into the atmosphere. Particle precipitation is particularly important at high latitudes. Impact ionization causes visible light to be emitted (the aurorae). The aurorae, which appear as ovals in the high latitude northern and southern hemisphere, are some of the most beautiful natural color and light shows. Particle precipitation-induced ionization increases the E region ionosphere substantially, particularly at night when photo production is absent.

There are other, more transient sources of ionization at E region heights, including complex dynamics resulting from the effects of neutral atmosphere motion, auroral electric fields, and meteors entering the upper atmosphere that oblate (or burn up) and impact the surrounding neutral gas with enough energy to create an ionized trail. These sources produce narrow, short-lived regions of high-density ionization at E region altitudes, known collectively as Sporadic E. The mechanisms responsible for Sporadic E depend on latitude. Sporadic E can last from a few minutes to several hours. Ionization can be locally very high, and therefore, high-frequency radio waves can be reflected off these trails for long-distance communication.

The densest region of the ionosphere (and actually the entire magnetosphere) is the F region. It extends from 120 km and usually peaks at 300 km. In the region above the peak, called the topside ionosphere, the density slowly decreases and blends into the magnetospheric region called the plasmasphere. The transition between the topside ionosphere and the plasmasphere is typically at about 1000 km and is marked by the transition from oxygen as the dominant ion in the ionosphere to hydrogen as the dominant ion in the plasmasphere. (Plate 7 shows an image of the plasmasphere.) The F region is formed by extreme UV solar radiation ionizing atomic oxygen. F layer ionization decreases at night, but not as much as the E and D layer ionization because at this higher altitude recombination rates are lower, and the layer consists of atomic oxygen rather than the molecular ions that dominate in the D and E regions. Atomic ions have much lower recombination rates in general than molecular ions.

Figure 5.1 shows schematically the vertical structure of the ionosphere. The first panel shows the different layers that exist during the day. The F layer divides into two layers during the day because of the enhanced photoionization at high altitudes. The $F_2$ peak is more dense than the $F_1$ peak.

## 5.6 Ionospheric variations

The ionosphere varies in systematic ways because the main source of ionization – solar UV and X-ray intensity – depends on the position of the Sun in the sky at a particular location on Earth and on the Sun's absolute output. When the Sun is directly overhead, the intensity of sunlight reaching the upper atmosphere is greatest. As the observer either moves towards the poles or to the day–night terminator, the intensity decreases because the angle the Sun makes with the upper atmosphere is more oblique. As the observer moves into the dark or nightside hemisphere of Earth, the amount of sunlight goes to zero and production due to photoionization ceases. The rotation and curvature of Earth therefore give rise to variations in the ionospheric structure.

In addition, the Sun's output of energy is not constant in time. It changes rapidly (especially at the high-energy end of the electromagnetic spectrum) due to solar flares and over the solar cycle. Plate 3 shows the Sun in X-ray emission over the solar cycle. Note that during solar minimum there is little X-ray emission, while at solar maximum the Sun's atmosphere emits large amounts of X-rays. This gives rise to a solar cycle variation in the intensity of ionization of the ionosphere. During solar storms, the ionospheric structure can be drastically modified by energy input from the Sun. Therefore, during geomagnetic storms the ionosphere becomes most disturbed and the most space weather impacts are noted.

## 5.7 The aurora

As mentioned with regard to the E region of the ionosphere, energetic particles can precipitate into the atmosphere causing impact ionization that produces light. This light, called **aurora**, is visible from the ground with the naked eye during the winter months in Earth's polar regions. Where do these particles come from, and why do they primarily enter Earth at high latitudes? Recall from Chapter 4 that Earth's magnetic field is shaped like a dipole magnet. Field lines come out of the southern hemisphere and enter Earth in the northern hemisphere. Also recall that charged particles feel a force called the Lorentz force when moving through a magnetic field. This force causes particles to spiral around the field. If a particle is moving along a field line, there is no magnetic force exerted on it unless a component of its motion is perpendicular to or across the field line. Energetic particles trapped in Earth's magnetosphere are funneled down into the north and south poles where the field comes into and out of Earth. Therefore, impact ionization and creation of the aurora are most prevalent at high latitudes. The aurorae are also called the northern (or southern) lights, or Aurora Borealis (in the north) and Aurora Australis (in the south).

The aurorae appear as ovals around the poles (see Chapter 1, Figures 1.2 and 4.7). This is because the particles that cause the aurorae

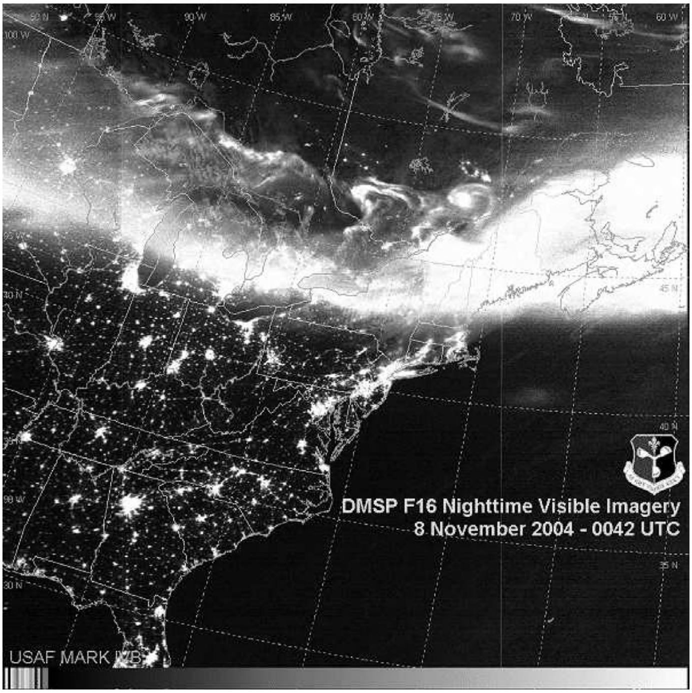


**Figure 5.2** The visible aurora observed from 800 km altitude by a camera on the DMSP satellite. Note that cities on the east coast of the United States are clearly visible from space. (Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the US Air Force Weather Agency.)

are from the plasma sheet that occurs only in a narrow section of Earth's magnetosphere (see Figure 4.3). Aurorae can be several different colors (greens, purples, and reds dominate) because they are primarily from nitrogen and oxygen molecules and atoms. The color depends on what atom or molecule is excited by the impact of plasma sheet electrons and the energy of the impacting electron. The most common auroral color is green, which comes from excitation of atomic oxygen. Plate 8 shows the aurora as seen from Alaska. Figure 5.2 is a black and white image of the aurora over the upper Midwest and Northeastern US. Individual bands or arcs as seen in Plate 8 make up the broader auroral oval, a part of which is seen in Figure 5.2.

## 5.8 Impacts on communication

Recall from Chapter 1 that in 1901 Marconi sent the first transatlantic radio message. English physicist Heaviside and Irish physicist Kennelly proposed the existence of an ionized layer in the upper atmosphere to explain how Marconi's radio waves could be reflected around the curvature of Earth. The English physicist Appleton soon verified this suggestion experimentally. The ionosphere is still sometimes referred to as the Appleton layer.

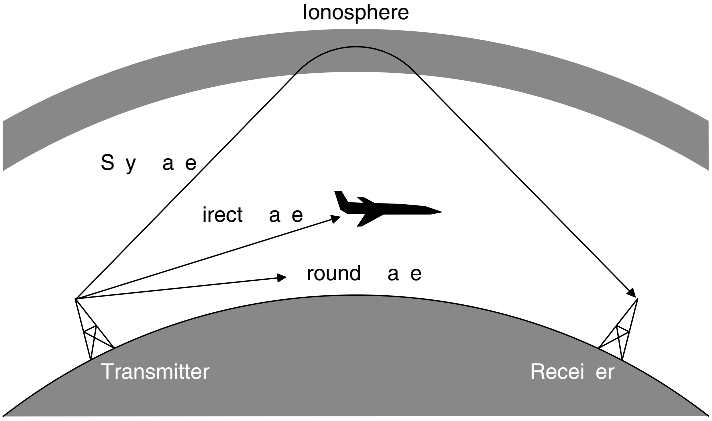


**Figure 5.3** The ionosphere can refract and reflect radio waves. Long distance radio communication is possible due to "bouncing" radio waves off the ionosphere (adapted from Radtel HF Radio Network).

How does the ionosphere interact with radio waves? Radio waves are electromagnetic energy with long wavelengths and low frequencies. As they propagate through an ionized medium, they become refracted or bent (Figure 5.3). The amplitude of the bending angle depends on the frequency of the electromagnetic wave and the density of the ionized gas. At a specific frequency – called the critical frequency – the wave will be perfectly reflected. This frequency, which is proportional to the density of the gas, is given by

$$f_{\mathrm{critical}} = 9\sqrt{n_{\mathrm{e}}},$$

where $n_{\mathrm{e}}$ is the electron number density in $\mathrm{m}^{-3}$ and the critical frequency is given in Hz. The peak ionospheric electron number density is typically $10^{12}\ \mathrm{m}^{-3}$ and therefore the critical frequency is $9 \times 10^6$ Hz or 9 MHz. This is in the high-frequency (HF) radio band. At frequencies less than this critical frequency, the ionosphere will reflect a signal back to the ground (or if a signal is coming from space, back out into space). At frequencies above this one, the radio wave can propagate through the ionosphere (it will still be refracted, but won't be completely reflected.)

## 5.9 Supplements

### 5.9.1 Photochemistry

The main source of ionization and particle excitation in planetary atmospheres, including that of Earth, is **photoionization**. In this process, a photon of electromagnetic radiation with sufficient energy (or a frequency greater than a specific threshold) interacts with an atom or molecule and knocks it into a higher energy level. In general, photochemical reactions can be written schematically as $A + hf \rightarrow A^*$.

This means that molecule $A$ absorbs an amount of energy from an electromagnetic photon (represented by $hf$, the energy of a photon from $E = hf$, where $h$ is the Planck constant and $f$ is the frequency of the EM radiation) and undergoes a change of energy to an excited molecular energy state $A^*$. This excited state could take the form of energy transferred to the molecule's motion, called rotational or vibrational energy. Molecules can spin or atoms in the molecule that are connected by chemical bonds can vibrate back and forth, analogous to atoms connected to each other via a spring. The molecule can also undergo an electronic transition, in which an electron is knocked to a higher orbital shell, can be dissociated (broken apart), or can become ionized (have an electron knocked off or added). What is special about excited molecules is that their chemical reactivity is enhanced. In biologic systems this can have detrimental effects on the organism (see Chapter 7 for a further discussion), and in the atmosphere of Earth these excited molecules can change the local equilibrium state. One potential outcome of an excited molecule or atom is that it can transition back to the ground state by emitting a photon. This photon will have a discrete frequency depending on the atomic or molecular species and so this emission is "line emission". That is, the light from these transitions is in specific wavelengths or wavelength bands.

Atoms and molecules can also be excited by impact of ions and electrons. In this case, the optical emissions that result are termed aurorae since they are generally confined in geographic coordinates to the poles, where the magnetic field lines of Earth can funnel the particles from outside the atmosphere down into the ionosphere. The field that studies optical emissions from the atmosphere is called aeronomy (from the Greek, for the study of air). Sydney Chapman coined this name, which is now used to describe the field of space physics that studies the upper atmospheres of the planets where ionization and dissociation (the breaking apart of molecules) is important.

## 5.10 Problems

5.1 The amount of photoionization in Earth's ionosphere depends on a number of factors. Neglecting transport (motion of an ionized gas from one place to another), what are the most important factors that determine the amount of photoionization in the ionosphere?

5.2 The temperature of the thermosphere can reach 2000 K. Why aren't astronauts cooked as they "walk" in space?

5.3 What is the density of air at 100 km if the atmosphere has a scale-height of 8 km?

5.4 Atmospheric drag is proportional to density. How does the magnitude of atmospheric drag change over a solar cycle at 100 km altitude?

5.5 What effect would a major solar flare, which emits a large amount of UV and X-ray radiation, have on the dayside ionosphere?

5.6 How would Earth's vertical density structure change if Earth were one half its size? Two times its size?

5.7 A satellite re-entering Earth's atmosphere will suffer a radio communications blackout because of the plasma created by the shock wave in front of it. If the satellite's radio operates at a frequency of 100 MHz, what is the minimum plasma density during the blackout?

# Chapter 6

# The technological impacts of space storms

However, I would like to close by mentioning a possibility of the more remote future – perhaps half a century ahead.

An "artificial satellite" at the correct distance from the earth would make one revolution every 24 hours; i.e., it would remain stationary above the same spot and would be within optical range of nearly half the earth's surface. Three repeater stations, 120 degrees apart in the correct orbit, could give television and microwave coverage to the entire planet. I'm afraid this isn't going to be of the slightest use to our post-war planners, but I think it is the ultimate solution to the problem.

*Arthur C. Clarke, Letter to the Editor, Wireless World, 58, February 1945. (The first suggestion of global communication using geosynchronous orbit satellites. The prediction came true in 20 years with the launch of the first geosynchronous satellite in 1965.)*

## 6.1 Key concepts

- atmospheric drag
- radiation effect on satellites
- radio wave propagation
- Faraday's law of induction

## 6.2 Introduction

Space weather has broad, everyday impacts on humans and technology. Spacecraft and astronauts are directly exposed to intense radiation that can damage or disable systems and sicken or kill astronauts. Radio signals from satellites to ground communication and navigation systems, such as the Global Positioning System (GPS), are directly affected by changing space environment conditions. What may be surprising is that many ground systems, such as power transmission grids and pipelines, and landline communication networks, such as transoceanic fiber-optic cables, are also susceptible to space weather impacts. Plate 9 shows the wide variety of systems that are affected by space weather, including astronauts and commercial airline crew and passengers as well as a host of satellite and radio communication devices. This chapter will describe how space weather affects these systems and describe the impacts space weather-related failures can have on technology and society.

## 6.3 Satellite orbits

We have become dependent on space technology, using satellites for a wide range of Earth-observing (such as weather) and communication (data, voice, television, and radio) purposes. Satellite technology is finding its way into a number of everyday activities. You probably used a satellite today. You did if you watched cable or satellite TV, listened to a nationally syndicated radio program, tracked a package being delivered to you by one of the major courier services, or used a credit card at a gas station pump or at a major retail store. To support these services, there are hundreds of satellites orbiting Earth. These satellites are in a variety of orbits, which means that each satellite has a unique path around Earth. Some satellites orbit close to Earth, others far from the surface. The orbit depends on the purpose of the satellite. There are four main important classes of orbits for Earth-orbiting satellites. The altitude above Earth that a satellite reaches defines the four main classes. These are low-Earth orbit (LEO), medium-Earth orbit (MEO), high-Earth orbit (HEO), and geosynchronous orbit (GEO). Figure 6.1 shows sample orbits for these four main classes of orbits. LEO satellites generally have circular orbits. A circular orbit means that the satellite's distance away from Earth's surface does not vary much during a complete orbit. Many satellites have elliptical orbits, meaning that the satellite moves closer and farther

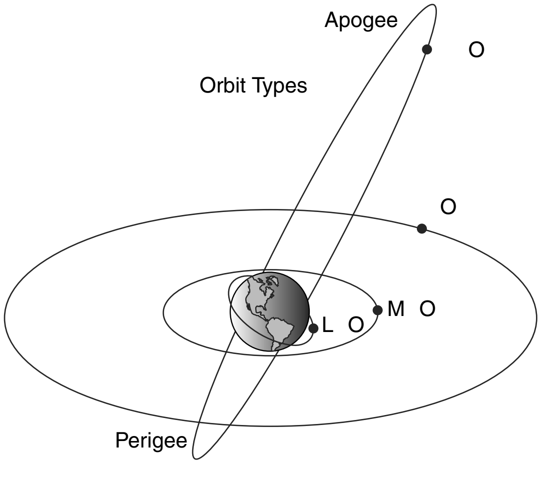


**Figure 6.1** The four main types of satellite orbits around Earth: low-Earth orbit (LEO), medium-Earth orbit (MEO), high-Earth orbit (HEO), and geosynchronous orbit (GEO).

away from Earth as it makes its way around Earth. In orbital dynamics, the distance farthest away from Earth that the satellite reaches is called apogee, while the closest is called perigee. LEO satellites have orbits from a few hundred kilometers to a few thousand kilometers in altitude. There are a number of advantages of LEO orbit. The first is that it is the easiest and least expensive orbit to get into. Depending on the size of the satellite, relatively small rockets can achieve LEO orbit. These launch vehicles, though costing tens of millions of dollars, are nearly a factor of ten cheaper than large rockets required to get some satellites into higher orbits. It takes lots of energy (and therefore money) to lift hundreds or thousands of kilograms into orbit. Another advantage of LEO orbit is that it is close to Earth, which means that small telescopes can see considerable detail on the surface, and low power radio transmitters can easily send signals back to Earth. Therefore LEO orbit is home to many Earth-observing satellites (like those that provide pictures to Google-Earth®). However, there are several significant disadvantages to LEO. The main space weather disadvantage is the lower the altitude, the more **atmospheric drag** there is on the spacecraft due to friction between the moving satellite and the tenuous atmosphere. Frictional force causes the spacecraft to lose altitude, which moves it into a denser neutral atmosphere (see Chapter 5 regarding the vertical structure of the atmosphere). The enhanced density causes increased drag, which lowers the satellite into even denser atmosphere. Eventually, the frictional force can heat up the satellite so much that it will begin to oblate or "burn up" in the atmosphere unless protected from the heat by strong insulating material. A satellite at less than 200 km altitude typically has a lifetime of only hours (one or two complete orbits). The space shuttle and International Space Station normally have altitudes between 280 and 460 km. They both have propulsion systems that allow them to continuously raise orbital altitude to prevent them from prematurely re-entering the atmosphere. Essentially all other LEO satellites do not have the capability to boost their orbits. Therefore the space shuttle must periodically dock with a satellite (such as with the Hubble[1] Space Telescope) to raise the satellite's orbit or the satellite will eventually burn up in the atmosphere. The lifetime of the satellite depends primarily on its initial altitude and the density of the upper atmosphere (as well as the satellite's cross-sectional area).

---

[1] Hubble, Edwin Powell (1889–1953), American astronomer whose work established the existence of galaxies and that the universe is expanding. This was found by observing that galaxies are receding from Earth and that their recession velocities are directly proportional to their distance from us. This is called Hubble's law. The first orbiting optical space telescope was named in his honor.

As discussed in Chapter 5, the density of the atmosphere is highly variable between solar minimum and solar maximum, and therefore a satellite's lifetime also depends on when it was launched. Space storms can cause rapid changes in the orbital altitude (and hence lifetime) of LEO satellites. The great storm of March 1989 caused thousands of space objects (including hundreds of operational satellites) to lose many kilometers of altitude. One satellite lost over 30 km of altitude (and hence a significant fraction of its orbital lifetime) during this storm.

The atmospheric orbital decay process has been extensively studied since it is one of the main parameters affecting satellite lifetime, and large satellites in uncontrolled re-entry could crash into populated areas. Most satellites are small enough that they will completely burn up in the atmosphere and not reach the ground. However, pieces from large satellites (like fragments from Skylab that fell to Earth in July, 1979) can survive re-entry and reach the ground. This would be similar to a large meteor entering Earth's atmosphere, and depending on the size of the fragment, could have tragic consequences if it hit an urban area. Large satellites are designed with propulsion capabilities so that their re-entries are controlled. Many do enter Earth's atmosphere and have pieces reach the surface, but by careful maneuvering at the end of the satellite's lifetime, the fragments are dumped into the ocean away from population centers. One such vehicle that is scheduled for completion in 2010 and then possibly decommissioned is the International Space Station (ISS). NASA has calculated that during solar maximum conditions, the ISS loses 400 meters of altitude per day (147 km $\mathrm{yr}^{-1}$). During solar minimum the loss is only 80 meters per day (28 km $\mathrm{yr}^{-1}$). Therefore, without periodic visits by the space shuttle and re-boosts, the ISS will relatively quickly re-enter Earth's atmosphere. Because the ISS is so big, large pieces will survive re-entry and hit the surface. Therefore, careful monitoring and control of the ISS is needed during its final orbits to make sure the pieces land harmlessly away from population centers.

## 6.4 Radiation impacts on satellites

Medium-Earth orbit (MEO), high-Earth orbit (HEO) and geosynchronous (GEO) satellites do not have significant satellite drag effects, but they have their own unique space weather concerns. These include spacecraft charging and high-energy radiation dose effects. Satellites in these orbits spend at least part of their orbit traversing the Van Allen radiation belts (discussed in Section 4.4), which contain trapped energetic particles that can severely damage or destroy sensitive electronic components.

There are a wide variety of **radiation effects on satellites**. These include surface charging, deep dielectric charging, single event upsets, and UV degradation of solar arrays. The next section describes each of these effects in detail.

### 6.4.1 Surface charging

Surface charging is caused by the interaction between a spacecraft and the low-energy electron environment of space. A satellite placed in space will be impacted by both positively charged and negatively charged particles. If the net transfer of positive or negative charge is not equal, net charging can take place. In addition to charged particle impacts, sunlight with enough energy can liberate electrons from any conducting surfaces via the photoelectric effect. The effect of these processes is that the spacecraft typically will become electrically charged. This is similar to scuffing your feet along a carpet and picking up electrical charge. If parts of the spacecraft, such as solar panels or the main spacecraft body, are made of different materials, those parts could charge up to different levels, possibly causing an electric discharge (spark) with potentially serious consequences. If the satellite is carrying any sensitive optical instruments, the spark could damage the detectors by overloading them with the bright flash. Another impact is that if the electric discharge is on a piece of sensitive electronics, the component could be damaged or “fried”. In order to prevent this from happening, care must be taken in the electrical design of the satellite. Despite careful designs, a large number of satellites have experienced space weather-induced failures and effects due to charging.

### 6.4.2 Deep dielectric charging

Deep dielectric charging and discharging is one of the most common and catastrophic issues regarding spacecraft electronics and radiation. Relativistic electrons in the Van Allen belts have enough energy to penetrate the spacecraft and deposit their charge on the insulating material (or dielectric material) of the circuit boards that make up the electronic “brains” of the satellite. Electric charge can build up to such a level that the dielectric material breaks down and charge can flow through these new pathways on the circuit board causing electrical shorts. Figure 6.2 shows a piece of plastic similar to that used in electronic circuit boards that has undergone dielectric breakdown. Had this contained electronic circuits, they would have been damaged. Satellite designers attempt to mitigate this effect by shielding sensitive components with thick aluminum covers or chassis and carefully grounding the circuit boards.

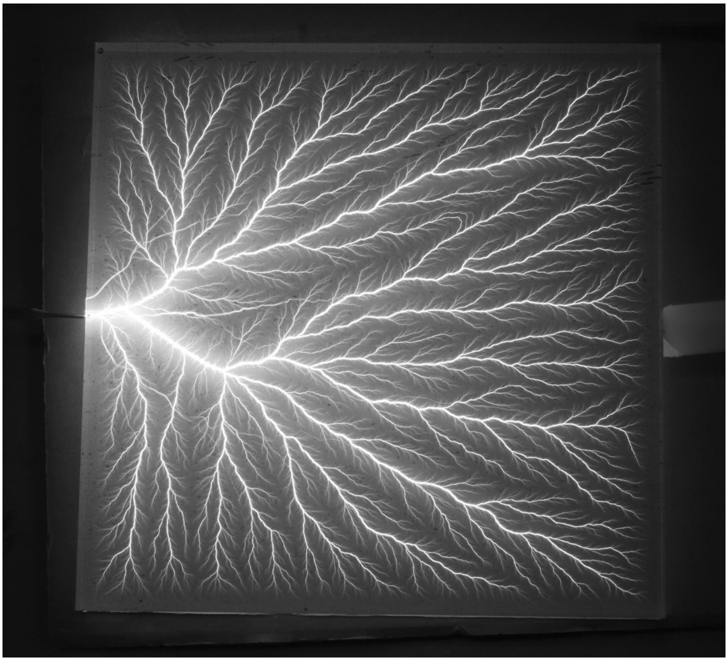


**Figure 6.2** A sheet of plastic (like the material used in printed circuit boards in computers) after it was exposed to a large electric field. The dendril features are defects in the material due to the dielectric breakdown (used with permission from Bert Hickman, Stoneridge Engineering) www.teslamania.com.

However, they must attempt to design the spacecraft to be not only tolerant of penetrating radiation, but also as lightweight and small as possible since launches are so expensive. Therefore trade studies are done to find the optimal shielding thickness for the environment over the expected lifetime of the satellite. This is similar to designing a beach house in Florida. Anticipating that the house may be subject to a category 3 hurricane in its lifetime, you design it to withstand very strong winds. You don't build the house to withstand extremely strong winds (a category 5 hurricane) because you don't want to live in a concrete bunker or pay for the extra heavy-duty construction required. Essentially you are conducting a trade study – what are the likelihoods of various categories of storms? How much can I afford? How does hurricane-proofing the house impact its design? The architect will attempt to optimize the construction – design the best house possible with the resources available within the risk profile.

### 6.4.3 Single event upsets

Single event upsets (SEUs) are due to penetrating ions that can "trigger" an electronic circuit. Since ions carry charge, a number of detectors, switches, and current and voltage regulators can observe a pulse of charge as the particle interacts with the circuit. This can cause a switch or a computer memory bit to “flip”, which could turn on or off or otherwise give an unintended signal to the spacecraft. These phantom commands can have catastrophic impacts. For example, a single event upset in a logic circuit could provide a phantom command to fire the satellite’s thrusters. Before ground controllers have figured out what happened, all the fuel could be spent, essentially ending the useful lifetime of the satellite.

### 6.4.4 Solar ultraviolet material degradation

Solar ultraviolet (UV) radiation is much more intense in space than on the surface of Earth. The ozone layer and oxygen in our atmosphere are very effective absorbers of UV (and X-rays and gamma rays), and therefore the surface of Earth is shielded from most electromagnetic radiation. Life on Earth would be very different without this atmospheric shield since UV light damages living cells. In addition to damaging living organisms, UV also can degrade certain materials, particularly plastics and other organic materials. UV light also contributes to solar cell degradation. In combination with energetic particle impacts (the main driver of solar cell degradation), UV light can make the solar panels less efficient. Satellite designers typically place solar arrays that are 25% bigger than needed for a particular mission, because over the lifetime of the satellite the efficiency of the arrays usually decreases by that amount. Individual solar storms can degrade solar cell efficiency by several percent and hence decrease the lifespan of a satellite by more than a year during just a single storm. Material selection for satellites, space stations, and future manned outposts on the Moon and Mars needs to take into account increased UV exposure, which limits the types of material that can be used, especially organic polymers and plastics.

## 6.5 Radio communication and navigation impacts

Space weather storms modify the density distribution of the ionosphere. Because **radio wave propagation** depends on the medium the waves move through, a time variable and spatially inhomogeneous ionosphere can severely perturb and degrade ground-to-satellite and satellite-to-ground communication. This can have serious impacts on different systems, but is particularly important for high frequency (HF) radio communication and Global Positioning System (GPS) navigation systems.

### 6.5.1 HF radio blackouts

High-frequency (HF) radio is used for ship-to-shore and ship-to-ship communication as well as by commercial airlines for air-to-ground and ground-to-air communication. This radio band is also popular with amateur radio operators. HF radio frequencies are between 3 and 30 Mhz. The ionosphere can reflect these frequencies, and therefore long-range communication is possible by bouncing your signal off the ionosphere several hundred kilometers above Earth. This phenomenon, called "skywave", allows for over-the-horizon communication and is how Marconi was able to make the first trans-Atlantic radio communication in 1901 (see Figure 5.3). The benefit of this frequency band – that it can interact with the ionosphere to permit long-range radio communication – is also its problem. Because the ionosphere is highly variable in space and time, HF radio communication can be severely degraded or even made inoperable depending on a wide variety of factors. Many of these factors are related to space weather and include the amount of solar activity (and hence sunspot cycle) and geomagnetic activity (particularly aurorae).

HF radio propagation depends on ionospheric density, which is controlled by sunlight and geomagnetic activity. Space weather degradation of HF radio has a particularly big impact on trans-polar airline flights. During large geomagnetic storms, HF radio communication can be rendered inoperable over the poles. Therefore, commercial airlines, which rely on HF radio communication, must base their flight schedules on space weather forecasts. Airlines will re-route trans-polar flights during large geomagnetic storms because of the impact on their HF radio communication ability.

Because of potentially serious impacts on HF radio communication, many users are switching to satellite phone communication (which uses much higher-frequency radio waves) and using HF as a backup system. However, since the cost of satellite communication is still relatively high, a significant number of industrial and government (maritime, aviation, and military) employees use HF radios, which are subject to space weather impacts.

### 6.5.2 GPS satellite errors

The Global Positioning System (GPS) allows users to accurately locate their position on Earth. The system consists of over 28 satellites in medium-Earth orbit arranged in such a way that at any given point on Earth at least four satellites are in view of an observer with an unobstructed view of the sky. These satellites have atomic clocks on board and continuously broadcast the time. A user on the ground with a GPS receiver can get this signal. By comparing the time broadcast by the satellite with the time at which the signal arrived, a distance (distance equals speed of radio signal divided by time for the signal to go from satellite to ground-user) to the satellite can be estimated. By triangulation (the process of determining the position of an object using three independent distance determinations), the exact location of the user can be estimated. Because the user does not have an atomic clock, a fourth satellite is used to acquire accurate time and the three other satellites are used to triangulate position. The speed at which a radio signal propagates through a vacuum is the speed of light (given as "*c*" in Einstein's famous equation). However, the speed at which an electromagnetic signal like a radio wave propagates through matter is less than the speed of light. This is called diffraction and has the effect of slowing down and bending the signal. The amount of bending and how much slowing occurs depend on the frequency of the signal and the properties of the medium. We experience this phenomenon when we look into water and see a rainbow. (Try this experiment: fill a glass with water and place a straw or pencil in the water. What happens to the straw or pencil when looked at from the side of the air–water boundary?) For a plasma, the property that determines electromagnetic propagation effects is the density. Therefore, because of ionospheric density, the radio signal from a GPS is slowed down. GPS systems attempt to account for this delay by using estimates or models of ionospheric density. For typical handheld single-frequency GPS measurements, positional errors on the order of 50 meters are common due to differences between the model ionosphere and the real ionosphere. This doesn't sound like much, but if GPS is used to fly an airplane, being 50 meters off the runway can make a big difference.

## 6.6 Ground system impacts

A number of technological systems on the ground are susceptible to space weather. During a large geomagnetic storm, large time-varying currents flow into and through the ionosphere. These currents can induce currents in long conductors on the ground, such as electric power lines, telephone lines, and pipelines. Induced currents in these systems can overload electrical components, causing failure, or can decrease the lifetime of the infrastructure by enhancing corrosion. The main principle behind these induced currents is called **Faraday's**[2] **law of induction**. This is

---

[2] Faraday, Michael (1791–1867), British physicist and chemist whose discovery of electromagnetic induction led to the invention of the electric generator and transformer.

a physical relationship that describes how a time-changing magnetic field can induce current and voltage in a conductor. Electricity can be described in terms of current or voltage. They are related through Ohm's law. In space, electrical currents flow into and through the ionosphere. These currents intensify and move to lower latitudes during geomagnetic storms. The time-changing and spatially varying currents create a time-changing magnetic field. According to Faraday's law, this time-changing magnetic field then can induce a voltage in long conductors. A wire is a good conductor designed to carry electrical signals long distances. On Earth, we have millions of kilometers of wire connecting buildings and houses with power plants and phone companies. These electric and communication grids are therefore susceptible to space weather effects.

### 6.6.1 Power grids

In the last few decades, power generation and distribution have become an interconnected continental-sized industry. Electricity produced by hydroelectric systems in Washington state in the United States is shipped to California. Power generated by HydroQuebec in eastern Canada can be shipped across the border to power homes in New York. Because of deregulation and this new interconnectivity, system vulnerabilities have increased. A power outage in one part of the grid can quickly propagate to other regions. Overgrown tree branches crossing a high-voltage line in Ohio triggered the power outage of 2003 that stretched from Detroit to New York City and left 50 million people in the dark. In March of 1989 a major geomagnetic storm caused an overload of a transformer in Quebec that quickly caused the collapse of the whole system. The transformer had been exposed to induced currents from the geomagnetic storm that exceeded its design capacity and it melted (see Plate 10). Transformers can convert high-voltage–low-current electricity into low-voltage–high-current electricity. It is more efficient to run high-voltage electricity long distances, but household appliances need high-current. Therefore the electrical system ships the electricity from the power plant to the user at high voltage, and transformers located near the user convert the electricity into useful household or industrial high-current electricity. If the transformer gets more voltage than it is designed for (like induced voltage from enhanced ionospheric currents during a geomagnetic storm), it can fail. Power grid operators therefore must watch the geomagnetic or space weather forecasts and reduce the load on their systems during geomagnetic storms. Of course, if a storm occurs during a heat wave or cold snap when electricity usage is high, the operators may not have the flexibility to handle the situation and then must institute planned rolling “brownouts” or potentially suffer catastrophic blackouts. It is estimated that if a perfect storm occurs during the next solar maximum (a large geomagnetic storm during a heavy electrical usage interval due to a cold or heat wave), hundreds of transformers could be damaged or destroyed. Replacement could take years because transformer manufacturing is expensive and fairly limited.

### 6.6.2 Pipelines

Metal will corrode when exposed to a variety of environmental conditions (like moisture and air). Corrosion is enhanced if there is an electrical current flowing through the metal. A long pipeline can be susceptible to enhanced corrosion if electrical currents are allowed to flow across it.

Pipelines carry natural gas and oil throughout the arctic region from their source region to terminals at lower latitudes. For example, the trans-Alaskan pipeline carries crude oil from Prudhoe Bay on the north slope of Alaska to the town of Valdez on the south coast of Alaska, traversing a distance of nearly 1300 km (800 miles). In Valdez the oil is loaded onto super-tankers for shipment to California and refineries elsewhere. The pipeline sits underneath the auroral oval, which is coincident with the largest ionospheric currents usually seen due to geomagnetic activity. These time-changing ionospheric currents can induce large currents in the pipeline. The Alaskan pipeline is specially electrically grounded to minimize this impact, but many pipes throughout the arctic region are not, and therefore their lifetime and potential for leaks is increased because of space weather.

The major disruption of oil production in Prudhoe Bay in 2006 was due to severe pipeline corrosion that may have been exacerbated by currents induced by auroral activity.

## 6.7 Supplements

### 6.7.1 Kepler’s laws and gravity

Johannes Kepler, using very precise data on the position of the planets in the night sky from Tycho Brahe, derived three laws of planetary motion that explained accurately the position of the planets about the Sun. Up until Kepler, the positions of the planets were described by the Ptolomeic model of planetary motion. Because this was a geocentric model of the Solar System (Earth was at the center), it required all types of geometric and mathematical tricks to describe the position of the planets. It was accepted because there was strong historical and religious motivation for having Earth (and hence mankind) at the center of the Universe. Kepler's laws are based on the assumption that Earth orbits the Sun and not the other way around. The success of Kepler's laws was the death knell of the Earth-centered model of planetary motion; they provided some of the strongest evidence up until that time in support of the Copernican heliocentric model. The first of the three laws states that the planets travel in ellipses with the Sun at one of the foci. An ellipse is like a squashed circle – like an oval – and its two foci are located on the long axis that connects the edges farthest apart. This is called the major axis. The position of the foci depends on the length of the axis that connects the edges closest together. This is called the minor axis. The more squashed the ellipse, the farther the foci are from the center. (See Figure 6.3 for a diagram of an ellipse.)

The second law states that an imaginary line connecting the Sun and a planet around its elliptical orbit will sweep out equal areas in equal amounts of time. This means that as the planet moves around its orbit, it speeds up and slows down depending on whether it is at the point of its orbit that is closest to or farthest from the Sun. When the planet is closest to the Sun, it is at perihelion and it moves faster. When it is farthest from the Sun, it is at aphelion and moves slower.

The third law states that the period of the orbit (time it takes the planet to go around the Sun once) is proportional to the distance the planet is away from the Sun. It is actually that the period squared is proportional to the distance cubed ($T^2 = ka^3$, where $T$ is the period, $a$ is the length of the semi-major axis, and $k$ is a constant of proportionality). The constant of proportionality is equal to 1 if the units of period are given in years and the units of distance are in astronomical units (AU).

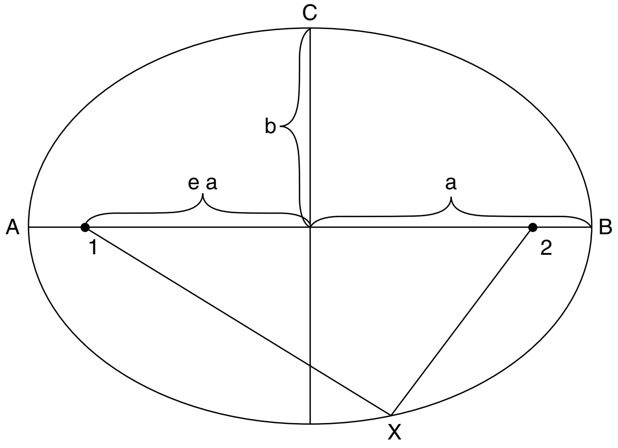


**Figure 6.3** An ellipse showing the major axis (between points A and B) and minor axis (between points C and D). F1 and F2 are the foci. The semi-major axis is "a", while "b" is the semi-minor axis. The ellipticity is given by "e".

For example, one can calculate the time it takes Jupiter (located 5 AU from the Sun) to make one orbit (one jovian year) simply by solving $T^2 = ka^3$ or $T = \sqrt{a^3} = \sqrt{5^3} = 11.2$ years.

Sir Isaac Newton developed the theory of gravity to explain the motion of objects falling to Earth and used it to explain the motion of the Moon about Earth and the planets around the Sun. The main concept of gravity is that two objects exert an attractive force (attempt to pull each other together) that is directly proportional to their masses and inversely proportional to the distance between the centers of mass squared. Algebraically, the magnitude of this force is written as

$$F_G = \left(\frac{Gm_1m_2}{r^2}\right),$$

where $G$ is the universal gravitational constant, $m_1$ and $m_2$ are the respective masses of the two objects, and $r$ is the distance between the centers of mass. The direction of this force is always attractive. Most objects have their mass spread out through the object (i.e., the mass of Earth is spread out inside the volume of the spherical Earth). However, Newton showed that the mass of a distributed spherical body (such as Earth) can be thought of as a point mass (all mass concentrated at a single point) at the center of mass of the object. The center of mass is the point that is the weighted average of the distribution of mass for that object. For example, the center of mass of a uniform sphere, such as a baseball or Earth – to good approximation, is at the center. For a book, the center of mass is a point essentially halfway from each edge. For an object like a baseball bat, the center of mass is nearer the thicker part of the bat, but along the long axis. Therefore, the force of gravity felt by a person on Earth is directly proportional to the mass of Earth and the distance the person is away from the center of Earth.

For a system of objects, such as Earth and the Moon, the system's center of mass lies at a point where the mass of one object times the distance to its center of mass is equal to the mass of the other object times the distance to its center of mass. Or, algebraically,

$$m_1d_1 = m_2d_2.$$

Figure 6.4 shows this relationship schematically. This is often called the "teeter-totter" ("see-saw") equation since it describes the point on the playground apparatus where you need to place the fulcrum to balance different-sized kids. The distance between the two masses is $d_1 + d_2$. For the Earth–Moon system, the center of mass is inside Earth, since the mass of Earth is so much larger than the mass of the Moon. Another important feature of the center of mass of a system is that the two objects

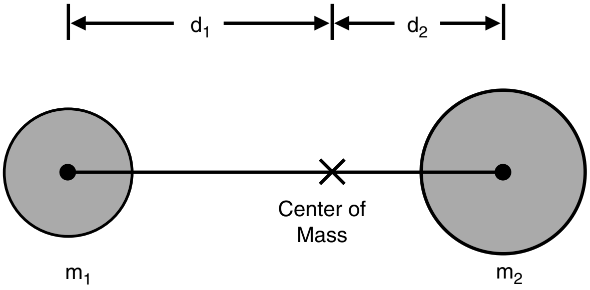


**Figure 6.4** Two bodies will orbit about a common center of mass whose location depends on the relative masses of the bodies.

will orbit each other about this point. Newton's law of gravity confirms Kepler's laws and provides the physical understanding for planetary motion. Gravity is the force that determines the orbital characteristics of the planets (and artificial satellites), and the motion of the planets can be predicted to very high precision based on this relatively simple force law.

### 6.7.2 Current and voltage in a circuit

The Western world is completely dependent on ready and reliable access to electricity. Electricity powers almost every modern appliance and system from cell phones, lights, computers, TVs, refrigerators, heating and air conditioning, traffic lights, to water pumps, etc. It is hard to imagine life without a cheap, ready supply of electricity. Electrical energy can be converted into mechanical energy to run fans and compressors, or can power electric circuits that control microprocessors and radios. Most commercial electricity is generated by spinning large turbines (essentially fans) that convert the mechanical energy of the spinning fan blades into electricity through an effect called induction. Induction is the process that can make electricity when a magnetic field changes near a coil of wire. Faraday's law of induction mathematically describes this phenomenon. The fan blades in a turbine can be spinning magnets and the turbines are surrounded by many coils of wire. Usually steam is forced through turbine blades, causing them to spin (this system is essentially a steam engine). The steam is produced from boiling water, by burning coal, natural gas, or oil. A nuclear plant uses the energy from splitting atoms to boil the water. Hydroelectric plants use rushing water instead of steam to spin the turbines (water wheels).

Since most power plants are large industrial sites, the power must be transported from the generating plant (usually located away from residential communities) to users. Electrical power is sent over long transmission lines as an alternating current (AC). Alternating current is electricity that oscillates in a sinusoidal fashion. In the United States the AC power oscillates at 60 Hz. AC is in contrast to direct current (DC), which is a steady state (not changing with time) or constant current. Batteries generate DC current. Electricity can be described by two main variables – current and voltage. They are related through the resistance of the material through which electricity is flowing. Resistance is essentially a measure of how easily electricity can flow. Metals and water have very low resistances (electric current flows easily), while wood, rubber and air have high resistances (electric current has difficulty flowing). The relationship, called Ohm's[3] law, is that the voltage[4] equals the current times the resistance:

$$V = IR.$$

One property electricity flowing through a wire (like that flowing the hundreds of kilometers from power plants to your house) is that part of the energy goes to heating the wire because of the wire's resistance. The amount of power dissipated in a wire depends on the amount of current flowing in the wire. The filaments inside a toaster are high-resistance wire with large currents flowing in them. The current causes the wires to glow red-hot and toast your bread. For power transmission wires, we want to make the resistance as low as possible and make the current as low as possible. Therefore power plants send the electric power out along the transmission grid as high-voltage–low-current AC electricity. Before it gets to your house, a device called a transformer converts it into a low-voltage–high-current source that is needed for appliances, computers, and lights. Because the power was transmitted with low current, the amount of energy loss due to heating of the transmission wires is minimized. Faraday's law of induction is the principle behind both electricity generators and transformers that change the voltage from high-to-low or low-to-high. The amount of voltage and current that a transformer can carry depends on its construction. Often transformers and power transmission grids operate near their peak capacity. If a large geomagnetic storm occurs during times of peak electrical usage, the system can become overloaded and the transformers can be destroyed.

[3] Ohm, Georg Simon (1789–1854), German physicist who deduced the theoretical explanation of electricity by careful experimentation and quantification of electrical current through a wire. The SI unit of resistance (ohm) and conductivity (the inverse of resistivity called the mho – his name spelled backwards) are named in his honor.

[4] Volta, Alessandro Giuseppe Antonio Anastasio (1745–1827), Italian physicist credited with discovering how to make electricity and building the first battery. The SI unit of electric potential or electromotive force (volt) is named in his honor.

## 6.8 Problems

6.1 Describe how space weather impacts airlines.

6.2 What are the orbital period and velocity of an astronaut in orbit at 300 km altitude? What is the orbital period of a satellite in geosynchronous orbit ($r = 6.6\, r_{\mathrm{E}}$)? How long does it take the Moon to orbit Earth if it is at 60 $r_{\mathrm{E}}$ from the center of Earth? (Use $k = 1.69$ and Kepler's third law with period ($T$) in hours and semi-major axis ($a$) in $r_{\mathrm{E}}$.)

6.3 How does the critical frequency of Earth's ionosphere change from noon to midnight?

6.4 The type of orbit (LEO, MEO, HEO, or GEO) a satellite is in depends on its purpose. What are the advantages and disadvantages of LEO and GEO for communication and Earth-observing purposes?

6.5 An auroral substorm produces tens to hundreds of billions of watts of electrical power. Why isn't it feasible to harness this power? (Hint: calculate the power per area and compare it with the amount of power the Sun provides per square meter at Earth.)

6.6 From 35 000 feet (typical altitude of a commercial airliner), how far can you see (or how far can line-of-sight radio communication travel)? Compare this to the size of an ocean or of the polar regions.

6.7 Where is the center of mass of the Earth–Moon system? ($m_{\mathrm{Earth}} = 6 \times 10^{24}$ kg; $m_{\mathrm{Moon}} = 7 \times 10^{22}$ kg; distance between Earth and Moon = 385 000 km).

6.8 What is the gravitational attraction of someone on Earth due to the Sun? How does this compare with the gravitational attraction of the Moon? The Earth? Of a classmate standing 0.1 m away?

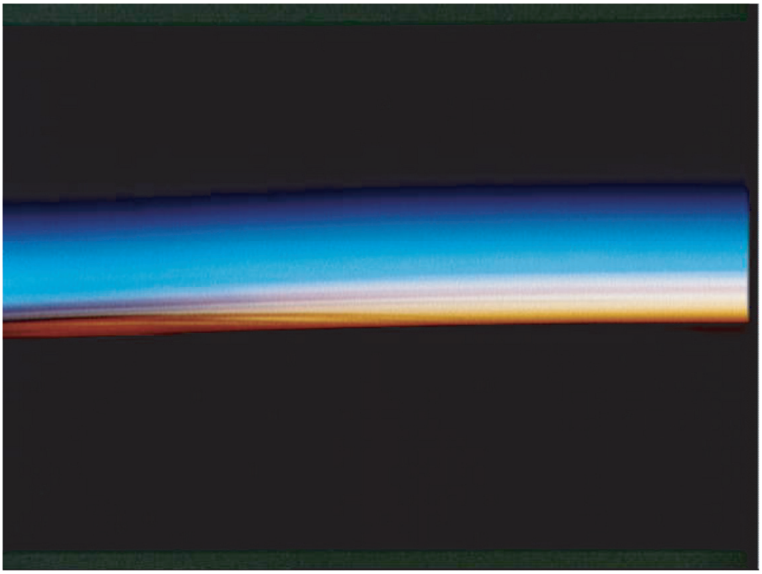


**Plate 1** The limb of Earth taken from the space shuttle. Notice the sharp edge to the blue of the atmosphere against the black of space (from NASA).

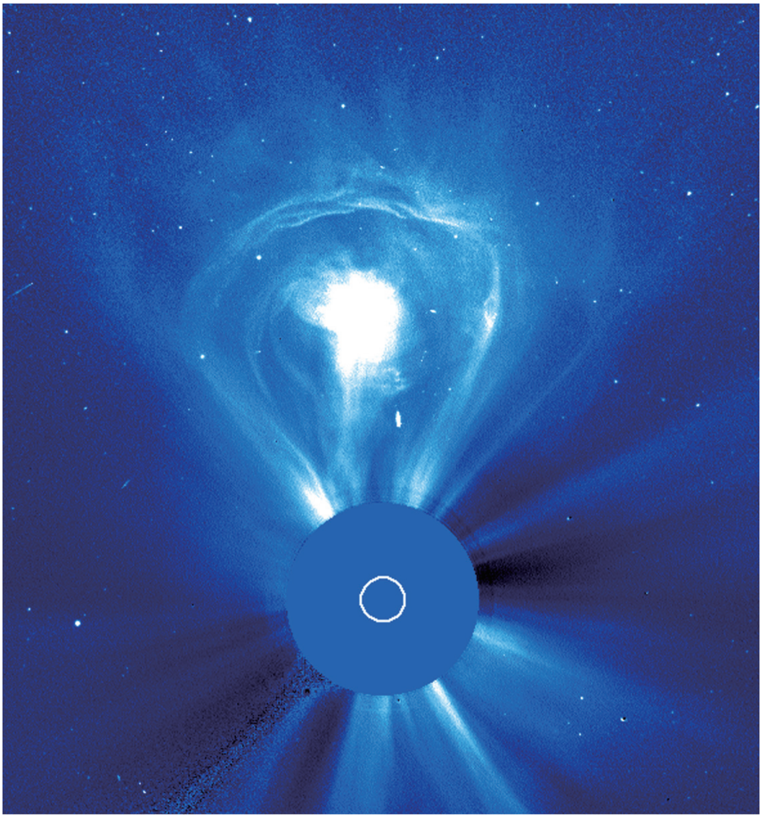


**Plate 2** A coronal mass ejection erupting from the Sun as observed by a white light coronagraph aboard NASA's SOHO satellite (Courtesy of SOHO/LASCO consortium. SOHO is a project of international cooperation between ESA and NASA).

**Plate 3** The Sun in x-rays observed over a solar cycle by the Yokoh satellite. Note that during solar maximum, the Sun is bright and structured in x-rays, while at solar minimum, the Sun is essentially dark in x-rays. (The solar x-ray image is from the Yohkoh mission of ISAS, Japan. The x-ray telescope was prepared by the Lockheed-Martin Solar and Astrophysics Laboratory, the National Astronomical Laboratory of Japan, and the University of Tokyo with the support of NASA and ISAS.

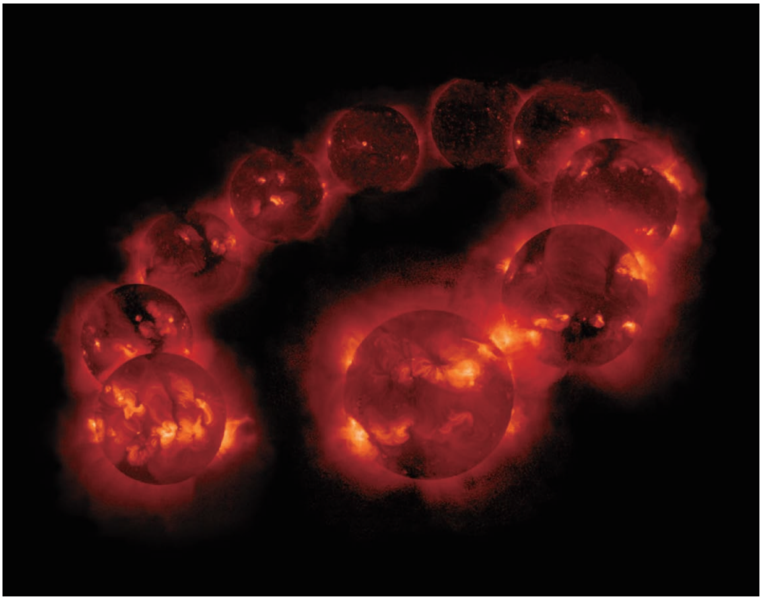


**Plate 4** During a total solar eclipse, the Moon passes directly in front of the Sun, completely blocking out the photosphere. The chromosphere and corona then become visible for the few minutes the eclipse lasts (from NASA).

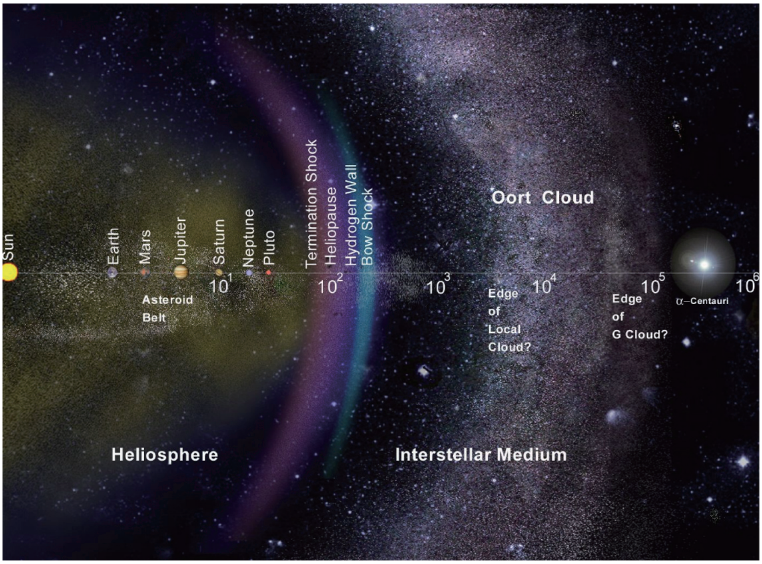


**Plate 5** The heliosphere and local interstellar medium out to our nearest stellar neighbor – *α* Centauri. Note that the scale is logarithmic in astronomical units (AU) with each tick mark 10× farther from the Sun than the previous one (from NASA Interstellar Probe Science and Technology Definition Panel, 1999).

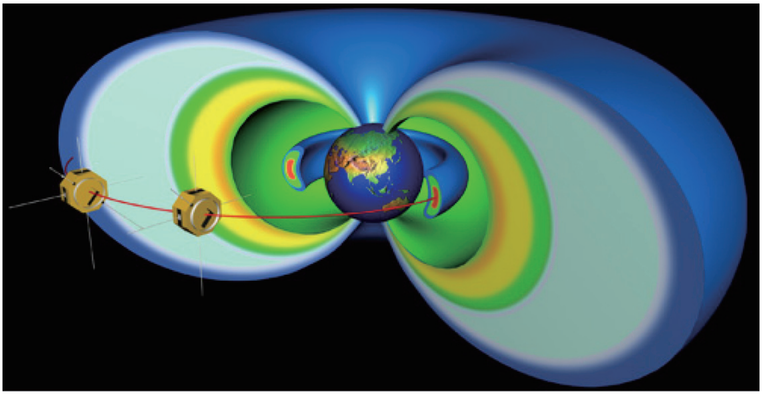


**Plate 6** Earth's radiation belts as measured by different NASA satellites. This 3D false color image shows the structure of the inner and outer radiation belts inside Earth's magnetosphere (from Johns Hopkins University Applied Physics Laboratory).

**Plate 7** The plasmasphere as observed in extreme ultraviolet light from the IMAGE spacecraft. The blue is a false color image of the high-density plasmasphere that extends several Earth radii from the surface of Earth. Note the sharp density boundary called the plasmapause is clearly visible in this image. The view is from over the North polar cap with the Sun to the upper right corner. The auroral oval is visible at the center of the image (from IMAGE Mission, courtesy of Bill Sandel).

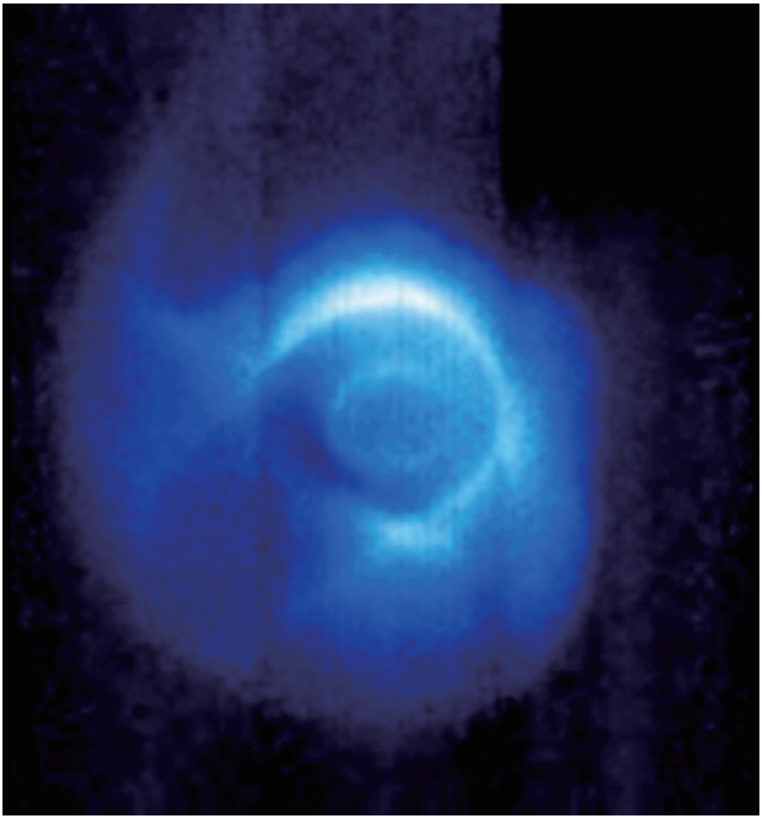


**Plate 8** The aurora as seen from Alaska (Photo: Jan Curtis).

**Plate 9** Different systems affected by space weather include satellites, astronauts, radio communication, and electric power grids (courtesy of Bell Laboratories, Lucent Technologies).

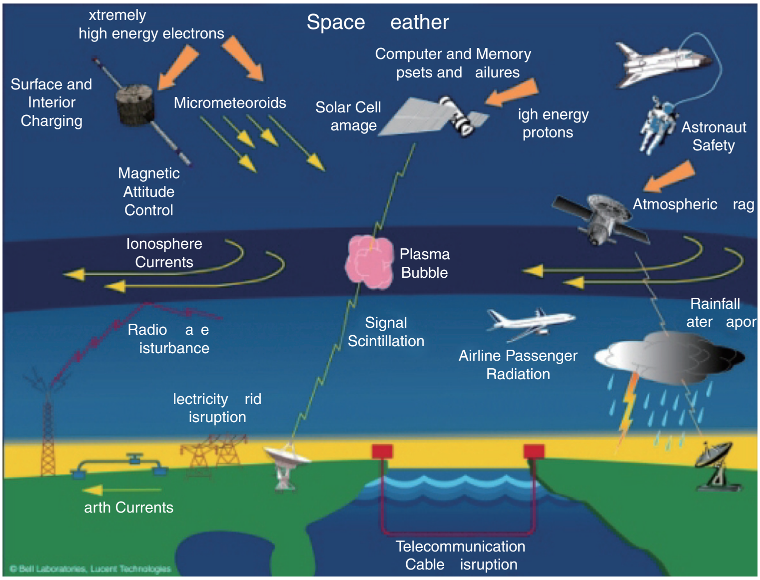


**Plate 10** A closeup picture of one of the transformers melted by the electrical overload brought about by the great geomagnetic storm of March 1989 (from the National Space Weather Program Strategic Plan, Office of the Federal Coordinator for Meteorological Services and Supporting Research, FCM–P30–1995, Washington, DC, 1995).

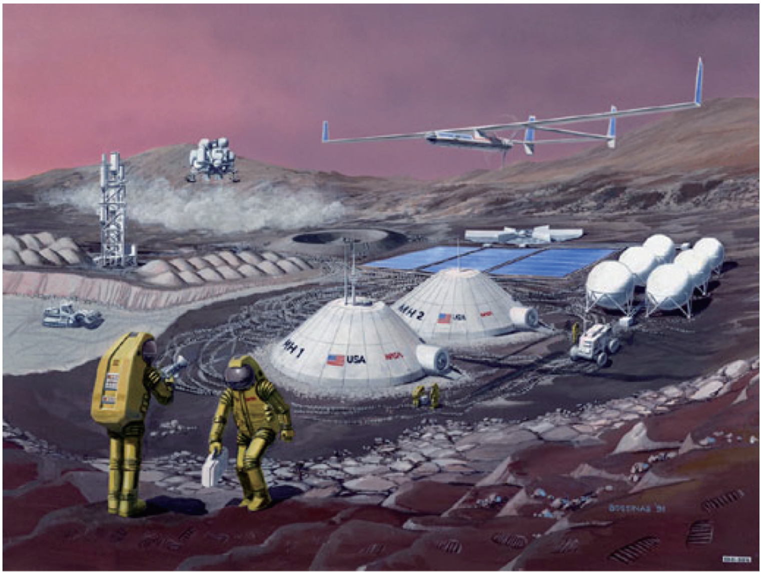


**Plate 11** An artist’s conception of a Mars outpost (from NASA).

**Plate 12** Meteor crater in northern Arizona formed when an asteroid 50 m in diameter hit about 25,000 years ago (aerial photo by Dr. David Roddy, courtesy of NASA).

# Chapter 7

# The perils of living in space

"My God, Space is radioactive!" Ernie Ray's exclamation after seeing the data returned from Explorer 1, the first US satellite launched in 1958. The radiation was space radiation in the Van Allen radiation belts.

*Quoted in Hess (1968).*

## 7.1 Key concepts

- protected by the atmosphere and magnetosphere
- types of radiation – electromagnetic and corpuscular
- rems (radiation equivalents in man)

## 7.2 Introduction

Life on Earth has an over 3.5 billion year history. It had a beginning and it will have an end. The longest that life can possibly exist on Earth into the future is approximately 4 billion years, though a number of catastrophes can happen on much shorter timescales (some of these are discussed in Chapter 8). In approximately 4 billion years, our Sun will run out of nuclear fuel and enter what is called the Red Giant phase of stellar evolution. It will expand perhaps past the orbit of Earth, vaporizing Mercury, Venus, and Earth. If humans are to survive the end of Earth, we will have to develop the technological capability to move to another star system. Humans have made the first tentative steps off Earth. We routinely fly astronauts into low-Earth orbit, and the US and China plan to have a manned presence on the Moon and Mars in the relatively near future. However, the technological obstacles to space travel are daunting – some say even insurmountable.

On the surface of Earth, we are **protected by the atmosphere and magnetosphere** from deadly electromagnetic and corpuscular radiation coming from the Sun and outer space. If we leave the protective cocoon of Earth's atmosphere and magnetosphere, we will have to bring this protection with us. In addition, our bodies are not adapted to the extreme temperatures, the extreme radiation, the intense vacuum, and impacts from high-velocity micrometeoroids that exist in space. This chapter describes how the space environment impacts living things (especially humans) and our current efforts to live and work on the surface of the Moon and Mars and ultimately around another star.

## 7.3 Radiation

There are two main **types of radiation – electromagnetic and corpuscular**. Electromagnetic (or EM) radiation consists of massless particles of pure energy called photons. Photons also act as waves, and so they have corresponding wavelengths and frequencies. Wavelengths describe the distance between the crests or troughs of a wave, and frequency tells how many times a wave passes a point each second. The entire electromagnetic spectrum is composed of radio, microwave, infrared, visible, ultraviolet, X-ray, and gamma rays, going from long wavelength to short wavelength. Figure 2.7 shows the entire EM spectrum schematically. Photons exist at different energies, with radio waves the lowest energy and gamma rays have the highest energy. Energy and frequency have a simple relationship:

$$E = hf,$$

where $E$ is energy, $h$ is the Planck constant, and $f$ is the frequency. Frequency and wavelength are related to the propagation speed of electromagnetic radiation – that is, the speed of light in a vacuum. The relationship is $\lambda f = c$, where $\lambda$ (the Greek letter "lambda") is wavelength and $c$ is the speed of light.

Radio waves have the longest wavelengths, the lowest frequencies, and the lowest energies, while gamma rays have the shortest wavelengths, the highest frequencies, and the highest energies. Visible light is made up of multiple frequencies that we perceive as different colors. Red has the longest wavelength and hence the lowest energy and frequency, whereas violet has the shortest wavelength and therefore the highest frequency and highest energy.

EM radiation from the Sun is mostly visible light, but light at essentially all wavelengths is also emitted. Much of the radiation does not reach the ground because of atmospheric absorption and reflection, particularly the high-energy photons such as UV and X-rays. However, in space the full intensity of sunlight is felt without the protection of the atmosphere.

Corpuscular or particle radiation is primarily sub-atomic (protons and electrons) or atomic or molecular particles. A natural background radiation environment comes from Earth, not only from space. Many elements are radioactive, meaning that they decay spontaneously from one element to another element by releasing different types of radiation. Madame Marie Curie[1] won Nobel Prizes in Physics and Chemistry for

---

[1] Curie, Marie (1867–1934) and Pierre (1859–1906). Madame Curie was a Polish-born French physicist. She and her husband shared the 1903 Physics Nobel Prize with Becquerel for their work in understanding radioactivity. Madame Curie went on to

her work in understanding radioactive decay processes. These decay processes, both electromagnetic and corpuscular radiation can be released. Alpha (helium nuclei) and beta (electrons) are two common types of radioactive decay by-products. These are also part of the solar wind and cosmic ray populations, though they originate from the ionization of helium and other gases and not from radioactive decay.

Both EM and corpuscular radiation can be ionizing radiation – that is, they carry enough energy to ionize an atom or molecule. In biology, this radiation can also interact with living cells by either damaging or destroying the cell or DNA contained in the cell. The building blocks of all life on Earth are cells. Cells consist of mostly water and the elements hydrogen, carbon, nitrogen, and oxygen with a little phosphorus and sulfur. Radiation, both corpuscular and high-energy EM radiation, can directly interact with these elements by removing an electron or ionizing the elements. This dramatically changes the chemical reactivity of the atom. Reactivity is the probability or likelihood that the atom or molecule will combine with another or take part in a chemical reaction with another atom or molecule. Molecules containing ionized atoms can react with other cells in a way that is detrimental to the living organism. Depending on the type and intensity of the radiation, the organism can suffer a wide range of health effects. For humans these can include white-blood-cell count reduction, nausea and hair loss, development of cancer, or immediate death. Cells that are most sensitive to radiation include white blood cells and the cells that make the white and red blood cells. Therefore, radiation exposure can have significant impacts on the body's immune system. The intensity of the radiation is very important – not only the absolute amount, but also the energy. Very energetic particles can impact a large number of molecules, hence causing significant ionization. High-energy EM radiation reacts differently with living tissue than energetic particles. EM photons lose all of their energy with a single interaction. This is opposed to how energetic particles lose their energy. Energetic particles lose their energy through collisions with large numbers of molecules. However, EM radiation (particularly X-rays and gamma rays) can create secondary electrons and photons, which then can interact with nearby atoms, creating more secondary electrons and photons, etc. The two processes are distinguished by calling particle radiation interactions with living tissue direct ionization radiation, while the process involving electromagnetic radiation is called indirect ionization radiation. This is because the same particle can interact directly with a

---

study the chemical and medical applications of radioactivity and won the 1911 Chemistry Nobel Prize for her work. A unit of radioactivity (the curie – Ci) named in their honor is equal to $3.7 \times 10^{10}$ decays per second.

large number of atoms or molecules, whereas a photon only directly interacts with one atom or molecule but indirectly affects others by producing other photons or electrons.

One possible outcome of the interaction from both types of radiation is that instead of ionizing the atom, the radiation knocks an outer electron into a higher energy level, creating what is called a free radical. Free radicals are highly reactive and can damage surrounding molecules. Since living tissue contain mostly water molecules, one of the most common free radicals produced is excited hydroxide ($OH^*$, where the star or * is the symbol used to distinguish that it is in an excited state. This is identical to the symbol used to distinguish an excited atom or molecule discussed in Chapter 5 with regard to aurora and photochemistry). Hydroxide, a strong oxidizing agent, can cause abnormal chemical reactions in living cells.

Free radicals can rupture cell membranes, causing the destruction of the cell. If enough cells are killed, the biological function associated with them (i.e., white and red blood cells, internal organs) could cease. Death can then occur either due to the loss of the organ's functionality or from infections that overwhelm the system due to the shut down of the body's immune system (decreased white blood cell counts). Another way in which radiation can damage living cells is through direct interaction with complex molecules such as proteins and nucleic acids (which make up the genetic DNA). Radiation can break DNA strands or break apart the protein, preventing their proper functioning. DNA has the ability to repair itself, and because much of the genetic code is highly redundant, damage at a few sites is not detrimental. However, if the intensity of radiation is great enough, the living cells' ability to repair themselves could become overwhelmed and cause permanent damage. This damage could lead to cancer, genetic mutations in offspring, or decreased cell function.

The outcome to a biological system exposed to radiation depends on the type of radiation (particle or EM), the energy of the radiation, the amount of radiation and the time period of exposure. For humans, various types of radiation impact different organs differently, and there are differences between the sensitivities to radiation among men and women. Most electromagnetic radiation is benign to humans, though high-energy EM radiation (UV, X-rays, and gamma rays) can be dangerous or lethal. Most of you have had experiences with medical X-rays, often at the dentist's office. The dental technician leaves the room when the X-ray machine is turned on. This precaution is made due to the potential long-term damage that X-ray exposure can have on living tissue and the high penetration power of X-rays. We also are taught to wear sunscreen and hats when we are out in the Sun to prevent sunburn. Sunburn is direct damage to the skin by solar ultraviolet radiation. Too much exposure can cause short-term pain and swelling and also lead to long-term health effects such as skin cancer. About one million Americans are diagnosed with non-melanoma skin cancers each year, and about 2000 Americans are killed by this disease each year. Melanoma, the most deadly form of skin cancer, kills an estimated 8000 Americans a year. The direct cause of this cancer is UV damage to the largest organ of the human body – the skin.

Sources of possible confusion regarding radiation are measuring how much radiation there is and how to measure the effect of the radiation on material or living organisms. One way to describe the level of radiation is using the units curies (Ci) and becquerels (Bq). Henri Becquerel[2] discovered spontaneous radioactive decay, and the Curies (husband and wife team) explained and measured it. All three shared the 1903 Nobel Prize in physics for this work. The Ci and Bq measure the number of radioactive decays per second in different radioactive materials (such as uranium). This describes a property of the activity of the source of the radiation, but does not say anything about the type of radiation or effect the radiation has on any material or biological system.

Rads and grays (Gy)[3] are used to measure the amount of energy absorbed by a specific material. They are given in units of energy per mass or joules per kilogram in SI units. Sieverts (Sv) or **rems (radiation equivalents in man)** measure what are called dose equivalents, taking into account the effects different types and energies of radiation have on human tissue. The units of sieverts are also joules per kilogram. Rems or sieverts are the units used when estimating radiation dose and effects on humans.

What happens to a human exposed to higher and higher doses of radiation? A complicating factor is that equal exposure to different types of radiation does not lead to equal biological impacts. A gray of alpha-particles (helium nuclei that are by-products of radioactive decay and the second most abundant ions in the Sun) does not produce the same effect as a gray of beta (electron) radiation or gamma ray radiation. That is because human tissue absorbs various types of radiation differently. Therefore, the units of sieverts or rems are used when discussing human biological effects of radiation because these units take into account the

[2] Becquerel, Antoine Henri (1852–1908), French Nobel-Prize-winning physicist who discovered radioactivity. The SI unit of radioactivity (Becquerel – Bq) named in his honor describes the number of radioactive decays of a certain amount of material per second (therefore, the larger the number of Bq, the more radioactive the substance).

[3] Gray, Louis Harold (1905–1965), British physicist who was instrumental in the development of radiation biophysics (radiology). The SI unit of radiation dose (gray – Gy) is named in his honor.

quality factor of the different types of radiation with respect to their impact on living tissue.

Typically, an American is exposed to approximately 0.36 rem per year. This dose comes from a variety of sources, as shown in Figure 7.1. The biggest source of radiation is from naturally occurring radon gas from the radioactive decay of radium, which has as its source uranium-238. These sources occur naturally in Earth's crust. Radon itself has a short half-life and decays into radioactive polonium, which is an alpha-emitter. Radiation from naturally occurring radioactive isotopes inside your body provides the next highest dose.

Exposure to high doses of radiation can lead to acute radiation effects (in medical terminology, acute means sudden or lasting a short time. It is the opposite of chronic). At an exposure level of 25 rems, subtle, hard-to-detect reduction in white-blood-cell counts (WBC) can occur. At 50 rems, reduction in WBC is easily detectable. WBC return to normal after a few weeks. White blood cells are an important component of the body's immune system. Therefore, reduction in WBC can lead to death if there are other illnesses or infections. At 75 rems there is a one-in-ten chance of nausea. Nausea is a symptom of radiation sickness because crypt cells that line the intestine are especially sensitive to radiation. Damage to these cells can trigger nausea, vomiting, and dehydration. At 100 rems there is also a ten percent chance of temporary hair loss. These two symptoms (nausea and hair loss) are often associated with cancer radiation treatment, which takes advantage of the ability of radiation to penetrate healthy bone and tissue and destroy cancerous cells. Even

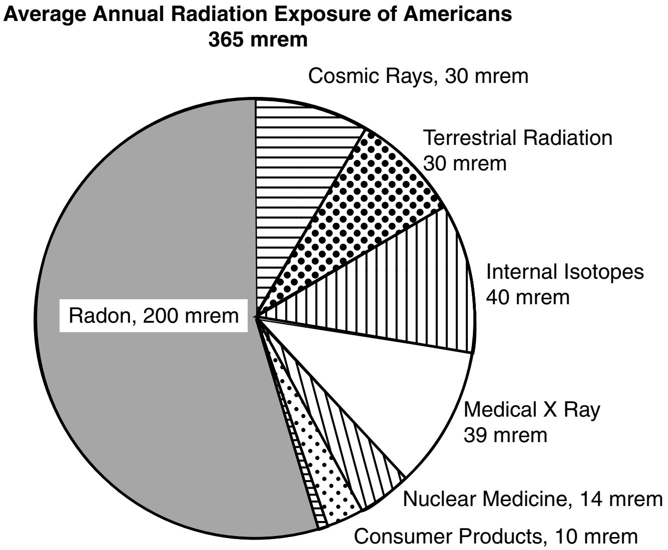


**Figure 7.1** Sources of radiation exposure for a typical American. The unit of radiation exposure is millirem (0.001 rem = mrem) (adapted from data from the US Department of Energy).

though the radiation is highly focused, it impacts not only cancerous tissue, but also can damage surrounding healthy tissue. At 200 rems there is a 90% chance of radiation sickness and moderate WBC reductions. At 400 rems there is a 50% chance of death within 30 days. Six hundred rems is lethal to most people within three to 30 days. Exposure to over 10 000 rems is almost instantaneously lethal, with death coming in less than a day. Unfortunately, human understanding of health radiation effects at these high limits has come from studying the aftermath of the Hiroshima and Nagasaki atom bomb attacks and the 1986 Chernobyl Nuclear Power Plant disaster.

What types of exposure are humans subjected to in space? An astronaut in low-Earth orbit, such as on the space station or space shuttle, receives about 0.1 rem per day (or a typical American's yearly dose in under four days). For comparison, a chest X-ray gives an exposure of 0.010 rem (10 millirem); therefore, an astronaut's exposure is equivalent to about 10 chest X-rays per day. If an astronaut is on a space walk during a solar storm, the dose at LEO could be 10 to 1000 times higher than the normal background. Therefore, astronauts need to be concerned about solar storms, even within the protective cocoon of Earth's magnetosphere. For trips to the Moon and Mars, large solar energetic particle events can give lethal doses to unprotected astronauts.

During the Apollo missions to the Moon, the US space program and the astronauts got lucky. In the 1970s, radiation effects from the Sun were not fully appreciated. Apollo 16 was launched in April of 1972. Apollo 17 (the last of the six lunar landings) was launched in December. In August of 1972 one of the largest solar proton events ever measured occurred. If the astronauts had been walking on the Moon during this event, there is estimated to be a 50–50 chance that one of them would have received a lethal dose.

In addition to acute radiation effects, there are effects of long-term or chronic radiation exposure, such as genetic defects in offspring and development of cancer. Effects of chronic radiation exposure usually are slow to develop, with latencies (medical term for time period) of effects to develop usually measured in years if not decades. These effects are characterized by the total dose exposure, which is just the sum of all the radiation received over a certain time frame. There are annual and career limits on the total radiation exposure that a radiation worker is allowed. The Occupational Safety and Health Administration (OSHA) in the United States provides these limits for those who work in high radiation environments, such as radiation technicians, airline pilots, nuclear power plant operators, and astronauts. The number of radiation workers has increased in recent years because of such things as increased luggage

X-ray inspection at airports and the use of ionizing radiation on the US mail to neutralize biological agents such as anthrax.

The career limits depend on gender and age, with men and older workers allowed exposures to higher doses. Developing embryos and fetuses are especially vulnerable to radiation effects, and therefore more stringent exposure limits are placed on women. For a 25 year-old woman, the career dose exposure limit is 100 rems. It is 150 rems for men the same age. These limits essentially increase by a factor of three at age 55.

An astronaut who spends three months on the International Space Station receives a typical dose of less than 10 rems. Spending the same amount of time working inside a laboratory on the surface of the Moon gives similar exposure rates. However, these exposure rates do not include effects of solar storms or exposures possible during spacewalks. Astronauts are particularly vulnerable when they are outside the protective shielding of their spacecraft or lunar base.

A 3-year mission to Mars (2-year round-trip flight time and 1 year on the surface) would give approximately 100 rems due to typical background radiation. Again, with a few solar storms astronauts could be exposed to lethal doses of radiation. Currently, providing radiation protection to astronauts is one of the biggest technological hurdles needed to be overcome before manned missions to the Moon and Mars could be possible.

How can we shield astronauts from radiation? The ability to penetrate matter depends on the type and energy of the radiation. Low energy alpha particles can be stopped relatively easily. A sheet of paper or a shirt can stop low-energy alphas. Energetic particles – especially energetic electrons – can penetrate deeply through most matter. Spaceships are usually designed with thin metal walls to save weight. Astronauts on the ISS are instructed to go to the center of the space station during space storms to provide the maximum protection available from the spacecraft structure. There is a proposal to fly a large magnet with a spaceship to Mars. The magnetic field would then act as a shield against some of the energetic charged particles. One proposal is to build subsurface labs in order to allow the lunar "dirt" (called the regolith) to act as shielding.

## 7.4 Problems of long-duration space travel

Besides radiation, there are a number of potentially fatal space environment impacts on human exploration of space. These include the vacuum of space, low gravity, and micro-meteoroids. Each of these problems will need to be addressed for humans to work and live in a space environment.

### 7.4.1 Vacuum of space

Humans living at sea level are subject to an atmospheric pressure of 1 atmosphere (atm) or 1 bar (b), or in SI units about 100 000 pascal (Pa). This pressure, due to the force of the atmosphere pushing down on the surface, is equivalent to 14.7 lbs per square inch (or one kilogram per square centimeter). If an unprotected astronaut is exposed to the vacuum of space, death will usually result unless she can be rescued within approximately 90 seconds. Within approximately 10 seconds she will lose consciousness. If the exposure to vacuum is sudden, problems with explosive decompression (the rupture of lungs if the astronaut attempts to hold her breath) can occur. There have been several instances where astronauts, pilots, and high-altitude parachutists have experienced rapid decompression. One symptom of rapid decompression is decompression sickness, more commonly called "the bends". This is the same problem that deep-sea divers must be worried about. The bends is caused by gases being released from the blood stream and tissue (similar to opening a bottle of seltzer water. The carbon dioxide gas is dissolved in the water under high pressure. When you open the bottle quickly the gases are released from the solution and cause the water to fizz out).

A myth perpetuated by Hollywood (such as in Arnold Schwarzenegger's movie, *Total Recall*) is that your body will explode and your blood will boil if you are suddenly exposed to the vacuum of space. Though bad things occur, they aren't that dramatic (unless they happen to you or your crewmate!). Due to the release of water vapor in your tissue, you can swell to up to two times your normal volume (unless constrained by a pressure suit). Arterial blood pressure will fall within a minute and effectively shut down blood circulation. After the initial rush of air from the lungs, continued release of air and water vapor will cool the nose and mouth near freezing due to the expansion of the gas. Death will come within 90 seconds. Decompression can happen for astronauts if their spaceship, spacesuit, or lunar or Martian habitat gets breached, and therefore designs for these suits and structures must incorporate features that will prevent or fix any leaks.

### 7.4.2 Reduced gravity

Another physiological issue that astronauts face when traveling in space is the greatly reduced gravity environment. On the surface of Earth we are subject to an acceleration of 9.8 m $s^{-2}$. This gravitational acceleration of Earth is written as $g$. The force with which Earth pulls down on an object on its surface is equal to this acceleration times the mass of the object (force = mass × acceleration). We call this force "weight" when referring to measuring how much force an object exerts on Earth due to its mass. It is also often referred to as a *g* force. So if a roller coaster subjects you to 4 *g*s, you are subject to a force four times the normal force of gravity. As you move away from Earth, the acceleration due to Earth decreases since the gravitational force is inversely proportional to distance between two objects raised to the second power (or squared). Also, if you are in orbit around Earth, you are constantly accelerating due to your motion (called centripetal acceleration). You are "free falling" toward Earth along with your spacecraft. Therefore, there is no gravity relative to your surroundings. In a sense, you become "weightless". If you were at the altitude of the space shuttle (about 300 or 400 km) and not in orbit around Earth, the force of gravity (which is only down about 10% from its surface value) would pull you back to Earth quickly. Astronauts in LEO experience micro-gravity because they are orbiting at such a speed that they continually "free-fall" around Earth.

The human body is designed to live in a one *g* gravitational environment. If you live on the space station or a spaceship going to Mars, your muscles and bones are not subject to the normal stresses of living on Earth and can begin to atrophy. In a month, an astronaut can lose up to 1–2% bone density and even more muscle mass. After a half-year mission, an astronaut can lose over 40% of muscle mass and over 10% of bone density. This makes astronauts very weak in Earth's one-*g* environment and even could lead to difficulty on Mars (1/3 *g* gravity). If astronauts spend significant amounts of time on the Moon (1/6 *g* gravity) or Mars, muscle and bone loss could have serious impacts on the success of the mission. Imagine being an astronaut landing on Mars after an eight-month trip and not being able to stand up against the Martian gravity because of muscle loss suffered on the journey. Astronauts currently spend much of their time on the ISS and shuttle exercising in order to attempt to slow the effect of low gravity on muscles and bones. Several spaceship designs, such as in Arthur C. Clarke's *2001: A Space Odyssey* or *Rendezvous with Rama*, incorporate artificial gravity. The outer wall of a spinning spaceship would exert centripetal force on an astronaut. If the ship is large enough and spins fast enough, the one-*g* environment of Earth could be mimicked.

### 7.4.3 Meteoroids

Finally, micro-meteoroids are present in the Solar System and can be from comets or asteroids. Though these can be very small (the size of a grain of sand or smaller), they can have large relative velocities compared to an astronaut in LEO or on the surface of the Moon. These velocities can be tens of thousands of miles per hour (tens of m $s^{-1}$), so even a speck of dust can have a tremendous impact on a spaceship or astronaut. A 17 000 mph grain of sand has about the same kinetic energy as a bowling ball moving at 60 mph. Impacts of meteoroids on the Moon have been observed by telescopes as bright flashes. If a Moon base happens to be hit, there is little that can be done except perhaps develop advanced radar detection systems and build the laboratories deep inside existing craters. Building them underground would protect astronauts not only from meteroid impacts, but also from some of the space radiation.

## 7.5 Living on the Moon and Mars

Astronauts would face a myriad of potential health issues traveling to and working on the Moon and Mars. Mars is currently the most problematic because a round trip would be of several years duration. A possible solution is to plan a one-way trip! Although some planetary scientists have volunteered, the ethics and potential political and legal issues involved will not allow such an endeavor. The technical hurdles of an up to 3-year mission, which include radiation protection, bringing enough fuel, food, air, and water (or the technology to extract these resources from Mars), overcoming the physiological effects of traveling and working in reduced gravity environments, and finally the psychological stress involved in living in such a harsh environment without any chance for rescue are daunting. A tooth infection or broken limb could be catastrophic for a limited crew depending on the entire team for survival. NASA, the Europeans and the Japanese are currently sending spacecraft and rovers to study the Red Planet, search for water, and develop technologies to extract resources (fuel, oxygen, and water) from the Martian soil or from the polar ice caps. Rocket fuel is essentially hydrogen and oxygen, and therefore finding frozen water near the surface of Mars could provide fuel, water, and breathable oxygen for the crew to live on Mars and return safely to Earth without having to bring everything with them from Earth (see Plate 11 for an artist's conception of a possible Martian outpost). The physical health of the astronauts due to radiation and low gravity are the main issues that currently do not have tenable solutions.

## 7.6 Interstellar travel

Space weather occurs not only in the context of our Solar System, but also around other stars and in the interstellar and intergalactic media. Therefore, to travel between the stars would require the same technological protections needed to go to our closest cosmic neighbor, the Moon. The greatest challenge to interstellar travel for humans is the vast distances between the stars. To get to our nearest stellar neighbor, even traveling at the speed of light, would require a four-year journey. To reach the center of the Milky Way galaxy would take 30 000 years at the same speed. Our current technology is capable of possibly going about 10% the speed of light, and so it would still take at least 40 years to visit Alpha Centauri (and four years to send back a radio message to mission control that you made it).

Stars are constantly being born, emitting high levels of UV during the initial phases of their evolution; energetic particles are constantly being accelerated in huge blast waves during the death throes of stars; and cosmic collisions between binary stars and stars and black holes fill the Universe with their radiation. Other stars also show variability, in some cases much stronger and shorter timescale variability than our Sun. Therefore, not only is any interstellar journey long, it is perilous.

Therefore, if there is life elsewhere in the Universe, even if it has evolved into intelligent life, it is hard to convincingly argue that these beings will have the capability to visit our corner of the galaxy. However, in the last century we have begun to beam EM radiation out into space from our radio and TV broadcasts and in the last 25 years we have sent probes – the Pioneer and Voyager spacecraft – out towards interstellar space. Voyager I is expected to leave our heliosphere and pass into interstellar space by approximately 2020. It should continue to fly across the Galaxy for millennia. Who knows if Voyager I will ever come close to another star system that has a spacefaring civilization. It carries a plaque and recordings of the civilization that created it with a star map pointing out our Sun and a map of our Solar System showing Earth. Will someone find this, decipher it, and have the capability to come look for its owner? Will we be still here?

## 7.7 Supplements

### 7.7.1 Special relativity

Special relativity describes the motion of objects that are moving close to the speed of light. The speed of light is $3 \times 10^8$ m $s^{-1}$ (or 300 000 km per second). This is really fast compared to the normal speeds we are used to on Earth – a photon of light could circle Earth nearly 7.5 times in one second. Albert Einstein developed the special relativity theory during his *annus mirabilis* (Latin for extraordinary year) of 1905. In that year, he published four papers, each of which is Nobel-Prize-worthy, though only one (his study of molecular motion called Brownian motion) was cited in his 1921 Nobel Prize announcement. The other two papers described the photoelectric effect, which ushered in the development of quantum mechanics, and his paper on the equivalence of matter and energy contained his famous $E = mc^2$ equation.

Special relativity describes electricity and magnetism for moving bodies. It explains the famous Michelson–Morley experiment that found that the velocity of light is independent of Earth's relative motion and hence light (unlike sound waves) does not require a medium to propagate. The original idea was that light propagated like a sound wave through a medium called the æther. The Michelson–Morley experiment attempted to measure the relative motion of Earth through this æther by looking at the propagation of light split into two beams moving orthogonal to one another. One beam pointed in the direction of motion of Earth in its orbit around the Sun and the other at right angles. They found no difference in the velocity of the light regardless of the direction of the light beam. The essential point of Einstein's paper was that there is not an absolute reference frame. In order for this postulate to work, Einstein needed to make the conjecture that the speed of light is independent of reference frame and moves at a constant velocity (called $c$) regardless of the motion of the observer or object emitting the electromagnetic radiation. Essentially he modified classical mechanics (the study of the motion of objects that was developed by Isaac Newton and therefore called Newtonian mechanics) so that it was consistent with Maxwell's electromagnetic theory.

This leads to some pretty interesting (or even bizarre) conclusions for objects moving very fast, including time dilation, length contraction, and the relativity of mass. Time dilation is the fact that a clock moving with respect to an observer runs slower than it does when it is at rest. Therefore a clock onboard a fast-moving spacecraft will run more slowly than the same clock on the ground, as observed by an observer on the ground. This has been demonstrated experimentally. Therefore, time is relative. An astronaut traveling near the speed of light will "age" differently from those left on Earth. Imagine going on a round-trip interstellar journey at nearly the speed of light and coming back to Earth to find that your children are now older than you! This completely non-common-sense conclusion was subjected to intense scrutiny by scientists. Over the years, all of the main predictions of special relativity have been tested and confirmed. Special relativity is considered one of the most well-established concepts of physics.

Length contraction states that the length of an object in motion with respect to some observer appears to the observer to be shorter than its length $L_0$ when it is at rest with respect to the observer ($L_0$ is called the rest length). Length contraction only occurs in the direction of the relative motion. So this means that to an astronaut on a fast-moving spacecraft, objects on Earth appear shorter than they did from the ground (and to an observer on the ground, the spacecraft appears shorter than when it was waiting for lift-off on the ground).

The relativity of mass is another interesting consequence of special relativity. Objects moving at high speeds relative to an observer have larger masses than the same object at rest. Therefore a relativistic electron in Earth's radiation belts is more massive than an electron at rest. In order to understand a relativistic electron's interactions with other particles and its dynamics, special relativity needs to be taken into account.

The magnitude of all three of these effects depends on the relative speed of the particle compared to that of light. This is expressed as a ratio $v/c$. When $v/c$ is close to 1, special relativity effects are appreciable. Therefore, objects with $v/c$ close to 1 are called relativisitic. The actual equations showing the magnitude of this effect are:

$$t = \left( t_0 \middle/ \sqrt{1 - v^2/c^2} \right) \text{ (time dilation)},$$

where $t_0$ is time on a clock at rest relative to an observer, and $t$ is the time on a clock in motion relative to the same observer. The $\surd$ symbol is the notation for square-root.

$$L = L_0\sqrt{1 - v^2/c^2} \text{ (length contraction)},$$

where $L_0$ is the length at rest relative to an observer, and $L$ is the length of an object in motion relative to the same observer.

$$m = \left( m_0 \middle/ \sqrt{1 - v^2/c^2} \right) \text{ (relativity of mass)},$$

where $m_0$ is the mass at rest relative to an observer, while $m$ is the mass of an object in motion relative to the same observer.

Though it is not intuitive that the measurement of time, length, or mass should be dependent on the relative motion of the observer, these effects only appear when the relative velocity approaches the speed of light. For almost all motion that is observed on Earth and in the Solar System, regular Newtonian mechanics works well. One of the big implications of special relativity is the cosmic speed limit of $c$. Note that if $v > c$, the observed times, lengths, and mass become imaginary numbers (and if $v = c$, mass goes to infinity and length contracts to zero). The implication of special relativity is that no object can go as fast or faster than the speed of light. This has tremendous implications for interstellar communication and travel, because distances between the stars are measured in light-years and distances across the galaxy in 100 000 light-years.

### 7.7.2 Estimation and the Drake equation

One of the tools of science is estimation. Critical thinkers are able to estimate essentially anything. Enrico Fermi[4] was famous for his ability to make quick and usually accurate estimates based on clearly stated assumptions, but little or no data. These types of estimations are now often called Fermi approximations. One of the famous Fermi approximations is estimating how many piano tuners there were in Chicago. To start, he estimated how many people lived in Chicago ($P_c$), how many people live in an individual household ($H$), what fraction of households had pianos ($p_h$), what fraction of them had their pianos tuned yearly ($f$), how long it would take a piano tuner to tune a piano ($t$), and how many hours a piano tuner worked each year ($T$). From this you can estimate how many pianos are tuned in a given year (population of Chicago/number of people per household × fraction of households with pianos × number of piano tunings per year) = number of pianos tuned in Chicago each year (or using our symbols – $P_c/H \times p_h \times f$). We can also estimate the number of pianos an individual piano tuner can tune each year (number of hours per day worked × number of days per week worked × number of weeks per year worked × number of pianos tuned per hour). Then after dividing the number of pianos tuned each year by the number of pianos an individual piano tuner can tune each year, and you have an estimate of the number of piano tuners in Chicago.

The Drake[5] equation is a Fermi approximation that attempts to estimate the number of advanced technological communicable civilizations in the Galaxy at any one time. It is a series of estimates of the number of star systems, the fraction of those that have habitable planets, the fraction of those that have life, the fraction of those that have advanced civilizations, and the average life-span of a technological civilization. The Drake equation essentially addresses each of the questions involved with the Search for Extra-Terrestrial Intelligence (SETI). SETI asks the question "Are we alone in the Universe?" Since the development of radio astronomy, we have been able to scan the heavens looking for any form of radio communication that may have been beamed towards Earth. Because of the vast distances of even our nearest stellar neighbors, the

---

[4] Fermi, Enrico (1901–1954), Italian-born American Nobel-Prize-winning physicist who developed the first controlled chain fission reaction that led to the development of the atomic bomb and nuclear power. The element with atomic number 100 (fermium), which was discovered the year after his death, was named in his honor.

[5] Frank Drake (1930– ), American radio astronomer who in 1961 developed an estimate of the number of civilizations in the Milky Way Galaxy and helped usher in the new field of astrobiology, the study of the origin and evolution of life in the Universe.

radio signals left their source hundreds to thousands of years ago. In the Drake equation,

$N$ is the number of technical civilizations in our galaxy that can communicate with us.

$$N = R^{*} \times f_{\mathrm{p}} \times n_{\mathrm{e}} \times f_{\mathrm{l}} \times f_{\mathrm{i}} \times f_{\mathrm{c}} \times L,$$

$R^{*}$ is the rate of star formation in our Galaxy,
$f_{\mathrm{p}}$ is the fraction of stars that have planets,
$n_{\mathrm{e}}$ is average number of planets that can potentially support life per star that has planets,
$f_{\mathrm{l}}$ is the fraction of planets that develop life,
$f_{\mathrm{i}}$ is the fraction of planets with life that develop intelligent life,
$f_{\mathrm{c}}$ is the fraction of planets with intelligent life that are willing and able to communicate,
$L$ is the expected lifetime of such civilizations.

Determining many of these parameters are active areas of research, especially the star formation rate in our Galaxy and the planetary formation rate. A new field of space science called astrobiology is attempting to address the origins and evolution of life on Earth and estimate how common it is elsewhere. The latter is being done currently by looking for evidence of life – past and present – on Mars and potentially some of the moons of the outer planets. Of course most of the parameters are not well constrained and therefore their values have large possible ranges (for all the "fractions" many are not constrained at all by current observations and can be anything from 0 to 1). Our single data point for life in the Universe is ourselves. So we know that the probability is not exactly zero, but there are some scientists who believe that the conditions required over the last 4.5 billion years of our own development are so particular that intelligent life elsewhere is highly unlikely. Therefore, estimates of intelligent civilizations in the Galaxy range from 1 (us) to 1000s. One difficulty with estimates of large numbers of advanced civilizations existing at the same time in the Galaxy is the fact that we haven't found each other yet. If advanced civilizations last for 100 000 or millions of years, why haven't we found evidence of them yet? This is often called the Fermi paradox because after hearing about early high estimates from the Drake equation, Fermi asked "Where are they?"

One of the largest unknowns as well as one of the most important parameters in the Drake equation is $L$, the lifetime of a civilization. The lifetime of our technological civilization capable of sending messages across space is just 100 years old. In the last 50 years, we have begun to understand some potential manmade and natural disasters that would end our civilization (e.g., global nuclear war, impact by a large asteroid). So how long does a typical advanced civilization last? Are we alone?

## 7.8 Problems

7.1 What is the wavelength of an FM radio wave ($f = 100$ MHz)?

7.2 Name and describe two ways in which radiation is used for medical purposes.

7.3 Using a scale height ($H$) of 8 km, at what altitude does atmospheric pressure fall to 10 mbar?

7.4 At space shuttle altitudes, what is the strength of gravity (alt = 400 km, $g = 9.8\ \mathrm{m\ s^{-2}}$)?

7.5 What is the mass of a radiation-belt electron moving at $0.9c$?

7.6 Estimate the number of college students in the United States. State assumptions and list your approximations algebraically before solving numerically in a form similar to the Drake equation.

# Chapter 8

# Other space weather phenomena

There is good evidence that within the last millennium the sun has been both considerably less active and probably more active than we have seen it in the last 250 years. These upheavals in solar behavior may have been accompanied by significant long-term changes in radiative output. And they were almost certainly accompanied by significant changes in the flow of atomic particles from the sun, with possible terrestrial effects.

*Eddy, J. A.* Science*, 192, 1189–1202, 1976.*

## 8.1 Key concepts

- global climate
- asteroid impact
- supernova

## 8.2 Introduction

Can space weather have impacts on **global climate**? We know that the amount of energy striking Earth from the Sun is the main driver of weather and climate. Temperature differences from day to night and from season to season are explained by the intensity of sunlight falling onto the surface and into Earth's atmosphere. For the most part, variation in intensity of sunlight is due not to changes in the luminosity or brightness of the Sun, but instead due to the daily rotation of Earth and the regular annual variation in hemispheric tilt toward or away from the Sun. In northern hemisphere summer, the Earth's axis and the northern hemisphere are tilted toward the Sun. Six months later, northern hemisphere winter occurs – Earth is on the opposite side of the Sun, and the northern hemisphere is tilted away from the Sun.

We know that Earth's climate has changed often in the past. Most of these changes occurred gradually – over thousands, if not millions, of years. One regular climate cycle is when Earth moves from periods of global cooling and Ice Ages (with capital letters), to periods between ice ages and relatively warm temperatures. Ice Ages have come and gone regularly in Earth's history. Two to three million years ago we entered the most recent Ice Age – called the Quaternary. Ice Ages are marked by periods of extensive ice cover (or ice ages with lower case letters) and interglacial periods with climate similar to today's.

The rise of agriculture and human civilization corresponds climatically to Earth leaving the last ice age (or entering an interglacial period) about 10 000 years ago. The present era, called the Holocene, is characterized by relatively mild global temperatures and year-round ice cover restricted to high elevations and polar regions. Some of these global climate changes correspond to changes in Earth's orbit (its eccentricity and its inclination) in cycles that last tens of thousands to hundreds of thousands of years called Milanković[1] cycles. Other large-scale changes brought about by plate tectonics have changed Earth's topology and landscape – creating and destroying seas, oceans, and mountains. In addition, life itself has changed Earth's climate by changing the chemical composition of the atmosphere. The first such large-scale change occurred soon after the arrival or formation of photosynthetic life that released oxygen as a by-product of respiration. More recently, Earth over the last century has experienced warming that is unprecedented over the last 1000 years and is due to the increase in greenhouse gases since the Industrial Revolution. Model predictions indicate that Earth's climate will be drastically different than it is today in less than 100 years because of the burning of hydrocarbon fuel for transportation and energy. This is very quick compared to the normal timescale of climate change.

Does the Sun's luminosity or other aspects of its energy output change or is the 11-year solar cycle a permanent property of the Sun? We know that the Sun is constantly changing. These changes have been observed minute-to-minute, day-to-day, and year-to-year and have also over the centuries. The most dramatic change is called the Maunder Minimum, when solar activity as indicated by the number of sunspots apparently disappeared. This period coincided with a time of extremely cold winters and cool summers in Europe and North America now called the Little Ice Age. Is there a cause and effect link between solar variability and climate?

The Sun isn't the only source of energy into Earth's space environment. Cosmic rays, asteroids, comets, and electromagnetic energy from other stars in our stellar neighborhood can also have an influence on Earth. Though events that can have significant impacts are very rare, they have happened in the relatively recent history of Earth. As human population continues to grow, the impacts of some of these extreme space weather effects could be devastating. This chapter describes three long-term space weather effects that could have significant and potentially catastrophic impact on civilization.

---

[1] Milankovic, Milutin (1879–1958), Serbian engineer and mathematician who demonstrated that many periods of climate change correspond to changes in the orbital characteristics of Earth, such as its inclination and orbital eccentricity.

## 8.3 Climate variability and space weather

How does changing solar activity as indicated by the number of sunspots affect our weather and climate? We know that the luminosity of the Sun hardly changes at all over the 11-year solar cycle (it is called the solar constant for good reason). The total luminosity changes by about 0.1% between solar minima and solar maxima. Models of Earth's climate clearly show that changes at this level have a very small impact on global mean temperatures or large-scale weather patterns. However, there is considerable evidence that shows that there are significant correlations with long-term solar activity and climate. If not the total amount of energy coming from the Sun, what could the physical mechanism be that connects climate and the Sun? With the advent of space-based observations of the Sun in the 1970s and 1980s, we have learned that the amount of solar UV radiation can change by 6 to 8 percent over a solar cycle. The amount of high energy UV and X-ray emission from the Sun is correlated with the number of sunspots, so during solar maximum, solar activity and the amount of high-energy radiation from the Sun are at their peak.

High-energy EM radiation absorbed in the upper atmosphere influences the chemical reactions that can take place. Therefore, changes in UV and X-ray intensity can have important consequences on the total energy balance of the atmosphere. The increase in high-energy solar radiation can increase the amount of ozone in the upper atmosphere, causing warming. We know that the thermosphere's temperature changes by a factor of two (from about 1000 K to 2000 K) over the solar cycle due to enhanced high-energy solar radiation, but we don't yet understand how energy is coupled from the upper atmosphere to the troposphere.

One suggestion is that solar activity and climate are linked by cloud formation. Clouds play an important role in climate because of their effect on the total amount of sunlight that reaches the ground and their greenhouse properties. Clouds are formed by water vapor condensing onto small particles called aerosols. Aerosols have their source from both below and above. There are several large natural sources of aerosols from below such as from dust storms, sea spray, and volcanoes. In addition, sunlight can cause chemical reactions with various gases, such as from swamps and exhaust from cars and industrial smokestacks, to make aerosols. From above, cosmic rays can interact with the atmosphere through various chemical reactions that can create aerosols and hence encourage the formation of clouds. Recall that the number of cosmic rays that hit Earth depends on the strength of the solar magnetic field and hence the interplanetary magnetic field – IMF. Strong solar magnetic fields produce a strong IMF in the heliosphere, deflecting cosmic rays from entering deep into the Solar System. A decrease in the number of cosmic rays reaching Earth would favor a decreased amount of cloud cover, which would allow more sunlight to reach the surface. During solar minimum – and during the Maunder Minimum – the amount of solar activity is low and hence the IMF is weak, which would allow more cosmic rays to hit the atmosphere of Earth. This would increase the amount of cloud cover and decrease the amount of sunlight falling on Earth's surface.

So is solar activity – through the interaction of the IMF, cosmic rays and clouds – another driver of climate change? Because of the very short timescales of our solar observations and accurate climate data, many of these correlations are very hard to confirm. Therefore a healthy amount of skepticism is present in the climate community. However, testable physical mechanisms now being proposed linking space weather and climate and new observations in the paleo-climate (paleo means ancient or pre-historic and is from the Greek word meaning "long ago") indicate that there are correlations between a number of climate changes with solar activity changes. Some of the most recent dramatic climate changes, such as the warm period known as the Medieval Climatic Optimum that lasted from 900 to 1250, were associated with a period of intense solar activity. This period saw increased temperatures in at least the region around the North Atlantic. During this time period the Norse people settled Greenland and gave it the name that seems peculiar now since it is covered by one of the world's largest ice sheets year round.

There are a number of causes of climate change, and Earth's climate has changed dramatically in the past. We are now coming to realize that our Sun is not a constant star and its variability can have a significant influence on Earth.

## 8.4 Asteroid and comet impacts

The Sun and cosmic rays are not the only energy inputs from space into Earth's atmosphere. Comets and asteroids have impacted Earth since its formation. One of the largest impacts occurred early in Earth's history when a large asteroid (with the same diameter as Mars, about half the size of Earth) hit Earth. Out of this collision, the Moon formed. Since then, Earth has been hit by asteroids and comets continuously, though the rate of impacts has dropped off considerably since about 3.8 billion years ago. The epoch of Earth from its formation 4.5 billion years ago to about 3.8 billion years is called the Hadean (the Greek word for Hell), because of continuous bombardment by asteroids and comets that were present in the early Solar System. Over the last 3.8 billion years, many of these objects have impacted planets and moons and each other, and have been swept out of the Solar System, or confined to quasi-stable orbits, such as the asteroid belt. However, still quite a few cross Earth's orbit, and there is a small, but good, chance each year that one will hit Earth again. These impacts could have global climatic effects.

Some of the most rapid climate change events correspond to impacts of large asteroids with Earth. The most famous **asteroid impact** is the one that occurred 65 million years ago and is thought to have led to the end of the dinosaurs. Could such an impact happen again? The short answer is yes. Extraterrestrial dust, meteoroids, and asteroids continuously bombard Earth. Some make it to the surface of Earth. On any given day, about 100 tons of extraterrestrial material settle onto Earth's surface. (Some of the dust on your windowsill might be interplanetary material.) A few of the dust or sand-sized objects "burn up" in the atmosphere as meteor trails or shooting stars. We know that a number of large objects with the potential to hit Earth cross its orbit. Those that have diameters greater than 1 to 2 km can cause global effects. The number of near-Earth asteroids of this size is estimated to be about 1100, and Earth has the probability of being hit by one of these asteroids about once every million years. Asteroids smaller than this are much more likely to impact Earth, and though they probably wouldn't cause long-term global climate change, they can have detrimental and deadly local impacts. Plate 12 shows a picture of Meteor Crater in northern Arizona. It is about 20 000 to 50 000 years old and is 1200 meters wide and 170 meters deep. Meteor Crater was created by an iron–nickel meteor about 50 m in diameter. It is estimated that asteroids of this type and size hit Earth every few thousand years.

The most recent on the order of 50 to 100-m diameter-scale asteroid impact was in 1908 when an asteroid slammed into the forests of Siberia near Tunguska. Asteroids less than 100 meters in diameter generally don't hit the surface, but explode in the air (called an airburst). Fortunately the 1908 asteroid hit well away from populated areas, but it knocked down trees over a 25 km diameter area and released energy equivalent to an approximately 50 megaton explosion (size of a large nuclear bomb). If this had hit a populated city, it would have caused one of the most catastrophic natural disasters in history.

When a large asteroid hits Earth, the atmospheric and climatic effects depend on whether it hits in the ocean or on land. Ocean impacts are more probable because oceans cover nearly three-quarters of Earth's surface. Impact of a large asteroid with the ocean would raise large amounts of water into the stratosphere and launch a large-scale tsunami. Some models predict that mid-Atlantic ocean impact of a 154-km diameter-sized asteroid would blast a hole in the ocean 11 miles across all the way to the seafloor and create a giant tsunami as water rushes back into the cavity. This tsunami could be a 100 m (400 foot) wave when it hits the continental shelf off the coast of the US. Obviously this wave would have impacts many kilometers inland and would devastate coastal communities. The presence of stratospheric water vapor could have short-term climate impacts similar to a large volcanic explosion. It is estimated that Earth has been hit about 600 times by a km-sized asteroid since the time of the dinosaurs. Impacts by asteroids with diameters of greater than 2 km in diameter would create a global impact that could directly kill several billion people.

Impacts on land can have global climate effects because of the amount of dust ejected into the stratosphere and the potential for widespread fires. Stratospheric dust can cool Earth's surface temperature by blocking sunlight and changing the atmospheric composition. This cooling effect would be global and could last for years. Recent volcanic eruptions such as that of Mt Pinatubo had significant global temperature effects that lasted for several years.

Two Hollywood movies (*Deep Impact* and *Armageddon*) that came out in the late 1990s imagined our response to a massive asteroid impact. Currently we do not have a complete inventory of all the asteroids that could potentially impact Earth nor would we be capable of doing anything about it if we knew one was heading our way. NASA has a program called Near-Earth Objects (neo.jpl.nasa.gov/) that is attempting to identify all asteroids and comets that cross Earth's orbit and hence have some probability of hitting Earth. The relative yearly probability of being killed by an asteroid impact is very small, about 20 000 to 1 over a lifetime. However, the potential global impacts make these highly unlikely but very destructive events important in Earth's history. An asteroid impact is the only natural hazard that has the potential to destroy civilization.

## 8.5 Nearby supernova

There are two main types of **supernova**, Type I and Type II (astronomers sometimes are not very imaginative when it comes to naming things). Type I supernovas occur in binary star systems when a white dwarf star (the remnants of a Sun-like dead star) accretes or attracts material from its main sequence companion. The material can build up on the white dwarf star and the resulting increase in pressure cause such an increase in internal temperature that the surface layer of the white dwarf star explodes. Type II supernovas occur when a star much more massive than the Sun exhausts its supply of nuclear fuel, thermonuclear reactions cease, and the star collapses under its own gravitational force. This collapse triggers a massive explosion that releases tremendous amounts of energy (about a billion billion times more energy than the Sun emits and generally several times more energy than a Type I supernova). This energy is in the form of electromagnetic radiation (gamma rays) and energetic particles that expand out into the Universe. What happens if there is a supernova in the neighborhood of Earth? If a supernova went off within about 50 or 100 light-years of Earth, the energy input into Earth's space environment and atmosphere could be enough to dramatically change the photochemistry of the atmosphere and destroy the ozone layer. This would expose the surface to high doses of UV radiation from the supernova and the Sun. The effect could last for years, and the UV radiation dose could be 10 000 times normal levels. This would obviously have significant impacts on the biosphere. The size of the Milky Way Galaxy is about 100 000 LY across, so a star within 100 LY is in our immediate neighborhood.

How many stars are within 100 LY of Earth? The estimate is about 14 000 stars, though most of these are unknown since many stars even within our stellar neighborhood are too dim to detect. How many of these stars can go supernova? In our Galaxy a star goes supernova about once a century. The last one occurred in 1680. More recently, Supernova 1987A occurred in the Large Magellanic Cloud, which is a companion galaxy to the Milky Way and can be seen from a dark place in the Southern Hemisphere with the naked eye. The amount of time between supernovae within 100 LY of Earth is estimated to be a few hundred million years.

Therefore, although close supernova explosions are extremely unlikely, they have the potential for global impacts on climate and civilization. These very unlikely but highly-destructive space weather events could have globally catastrophic consequences.

## 8.6 Supplements

### 8.6.1 Kinetic energy and conservation of energy

Energy comes in many different forms: mechanical, chemical, and electrical, to name a few. Mechanical energy is the energy that an object has due to its motion (called kinetic energy) or its potential energy due to its stored energy of position. For example, a spring can contain potential energy if it is compressed or stretched. When the spring is released, the potential energy is converted to kinetic energy. Another example is holding a ball above the ground. The ball has gravitational potential energy due to its position above the ground. If dropped, the ball's potential energy will be converted to kinetic energy as it speeds up or accelerates to the ground.

Chemical energy is the potential energy contained in the bonds between atoms. When chemical reactions – such as the oxidation of molecular oxygen and methane gas – occur, energy is released as heat, light, and work (the gas expands). The amount of energy released depends on the type and amount of reactants in the chemical reaction. Reactions that release energy are called exothermic ("exo" is Greek for "out of", while "thermic" is from the Greek for heat). Other reactions, which take energy from an external source and absorb it into the product of the reaction, are called endothermic ("endo" is Greek for "within"). Biological systems use chemical energy to sustain life. The chemical reactions that sustain animal life involve the oxidation of sugars and other hydrocarbons. Humans also use chemical energy for transportation (burning of gasoline converts the chemical energy inside the hydrocarbon fuel into the mechanical energy of the moving pistons to the driveshaft to the wheels) and environmental control (air conditioning and heating).

The concept of kinetic energy – the energy of motion – is important for understanding asteroid impacts because the amount of kinetic energy an asteroid has determines how destructive its impact would have on life on Earth. Kinetic energy is proportional to the mass of the object (the more mass a moving object has, the more kinetic energy it has) and the square of its velocity (the faster an object is moving, the more kinetic energy it contains). This is written as

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

The SI unit of energy is the joule.

The conservation of energy states that the total amount of energy in a closed system must be conserved. In other words, you can convert the system's energy from one form to another (from potential to kinetic or chemical to mechanical), but cannot get more energy out of the system than it already contains without adding it. This principle allows easy calculation of a number of important variables regarding the motion of the object. For example, one may wonder how fast a one kilogram mass is moving just before it hits the floor after being dropped from a height of one meter. We know from the conservation of energy that the potential energy of the system can be converted to kinetic energy. One way to express the potential energy of a mass lifted above the ground is where $PE$ is potential energy, $m$ is mass, $g$ is the acceleration of gravity at Earth's surface, and $h$ is the height above the ground:

$$PE = mgh.$$

This can be set equal to the kinetic energy of the object to find the object's velocity just before it hits the ground since the potential energy will be converted to kinetic energy. Therefore,

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

$$v = \sqrt{2gh}.$$

The one-kilogram mass will be moving at 4.4 m $s^{-1}$ just before it hits the floor. Note that the mass of the object actually cancels out of the equation. The velocity of an object dropped from a height is independent of the object's mass and depends only on the acceleration of gravity and the height from which it is dropped.

For the asteroid problem, we can estimate the velocity and mass of the object and therefore can estimate the total amount of kinetic energy that will be released. We use the conservation of energy to understand into what forms of energy this kinetic energy will be converted (i.e., heat, sound, the movement of dirt, rock, or water at the impact site, the heat of vaporization, etc.). For asteroids more than 2 km across, the amount of energy released would be globally catastrophic. Even sub-2-km-sized asteroids can have major regional impact and could directly affect millions of people and cause untold deaths and hundreds of billions of dollars in economic damage (see Chesley and Ward, 2006 for a description of the probabilities and damage estimates of various sized asteroids and their human and economic impacts).

### 8.6.2 Correlation and causation

One of the goals of science is to understand how and why things work. Knowing the cause of an effect can help us predict a future outcome (i.e., we know that cigarette smoking causes lung cancer and a number of other illnesses. Therefore, if we prevent young people from smoking, we can reduce the overall health costs and impacts of a number of illnesses). One way to attempt to find the cause of a phenomenon is to study a number of parameters and see whether any are correlated with the phenomenon of interest. For example, scientists studying hurricanes examined the role of sea-surface temperature, atmospheric pressure, and wind velocities as a function of altitude. Meteorologists found that warm sea-surface temperatures can give rise to the formation of tropical storms and hurricanes and that upper-level winds can be detrimental to the development of a storm. With this information, forecasters and meteorologists can make physical models in order to predict the development and evolution of a hurricane. Of course, scientists must make some assumptions or initial hypotheses about which parameters to examine when looking for correlations. This is done by thinking of possible physical mechanisms that would relate one parameter to a certain outcome or phenomenon. For example, in space weather, one may suspect that the direction of the interplanetary magnetic field would have an impact on the magnitude of a geomagnetic storm. This follows from our understanding of magnetic reconnection – southward IMF can directly couple with Earth's northward-directed magnetic field and transfer energy from the solar wind into Earth's magnetosphere. The excellent correlation between southward IMF and geomagnetic activity provided a strong piece of evidence that magnetic reconnection plays an important role at the dayside magnetosphere. Of course, scientists could have examined a number of things to see if they are correlated with geomagnetic activity. These include the price of gold, the performance of a sports team, the crime rate, the phase of the Moon or literally an infinite number of other things. What may be fairly surprising is that if you looked at enough random variables, there is a good chance that some of them WOULD be correlated with geomagnetic activity (or whatever else you are examining). Most things tend to have periodicities or cyclic behavior. And sometimes, merely by chance, two phenomena can appear to be correlated.

Another complication is that one phenomenon can be correlated with another and not be the direct cause even if there may be a physical mechanism that potentially connects the two phenomena. This is because many factors may be correlated with each other, even though they are independent. For example, there is a strong correlation between illiteracy and high mortality due to a number of diseases. There isn't a direct causative link between deadly disease and being able to read, but there is a clear and strong correlation between these two phenomena. The solution to this puzzle is that there is a strong correlation between low socioeconomic status and illiteracy. There is also a strong correlation between low socioeconomic status and access to healthcare and good nutrition. Therefore illiteracy isn't the direct cause of deadly diseases, and neither is being poor. The cause is lack of access to health care and good nutrition.

Therefore, when attempting to find the cause of an effect critical thinkers are careful not to confuse correlation and causation. One must always have a clear physical mechanism in mind when linking two variables, and one must always test whether a correlation is pure coincidence or actually related.

## 8.7 Problems

8.1 Some recent studies suggest that the strength of the solar magnetic field is increasing. What would be the potential climate impact of a prolonged period of strong solar magnetic field?

8.2 What role does the ozone layer play in protecting life on Earth's surface? What role does Earth's magnetosphere play?

8.3 Kinetic energy is the energy of motion. How much kinetic energy does a spherical asteroid with a 1 km diameter have if it is made out of iron (has a density of about 8 kg $m^{-3}$) and hits Earth at a relative velocity of 10 km $s^{-1}$? Convert your answer to megatons of TNT (1 megaton of TNT $= 4.18 \times 10^{15}$ joules).

8.4 Should society be concerned about low-probability events such as asteroid impacts or nearby supernova explosions? We know that there are many natural drivers of climate change. Should society be concerned about global climate change that can take decades to centuries to manifest itself?

8.5 Would there be any warning of a nearby supernova? Why or why not?

8.6 If rapid global impacts (and consequent mass extinction events) have happened in the past every few hundred million to billion years due to asteroid impacts and supernova, how can we explain our existence today?

# Appendix A
# Web resources

A number of excellent websites relate to space weather. Below is a selection of sites that are useful for exploring different aspects of the Sun–Earth relationship.

A regularly updated page giving news and information about the Sun–Earth environment:

`www.spaceweather.com`.

The National Oceanic and Atmospheric Administration's Space Environment Center continually monitors and forecasts Earth's space environment; provides accurate, reliable, and useful solar–terrestrial information; conducts and leads research and development programs to understand the environment and to improve services; advises policy makers and planners; plays a leadership role in the space weather community; and fosters a space weather services industry. The Space Environment Center is the United States' official source of space weather alerts and warnings:

`www.sec.noaa.gov/.`

A listing of satellite outages and failures:

`www.sat-index.com/failures/.`

*Space Weather: The International Journal of Research and Applications* is an online publication devoted to the emerging field of space weather and its impact on technical systems, including telecommunications, electric power, and satellite navigation:

`www.agu.org/journals/spaceweather/.`

NASA's web-page dealing with the Sun–Earth connection:

`sec.gsfc.nasa.gov/.`

The High Altitude Observatory (HAO) explores the Sun and its effects on Earth's atmosphere and physical environment, in partnerships extending throughout the national and international scientific communities for research, observational facilities, community data services, and education:

`www.hao.ucar.edu/`.

The European Space Agency's Space weather web server:

`esa-spaceweather.net/`.

The National Solar Observatory web page provides solar images and data from a number of solar observatories:

`www.nso.edu/`.

To understand a science it is necessary to know its history – Auguste Comte.

A timeline of the major achievements in our understanding of the solar–terrestrial relationship:

`measure.igpp.ucla.edu/solar-terrestrial-luminaries/timeline.html`.

# Appendix B
# SI units

Table B.1 *Fundamental SI units* (*see physics.nist.gov for more information about SI units*)

| Base quantity | Name | Symbol |
|---|---|---|
| length | meter | m |
| mass | kilogram | kg |
| time | second | s |
| electric current | ampere | A |
| thermodynamic temperature | kelvin | K |
| amount of substance | mole | mol |
| luminous intensity | candela | cd |

Table B.2 *Derived SI units*

| Derived quantity | Name | Symbol |
|---|---|---|
| area | square meter | $m^2$ |
| volume | cubic meter | $m^3$ |
| speed, velocity | meter per second | $m\ s^{-1}$ |
| acceleration | meter per second squared | $m\ s^{-2}$ |
| mass density | kilogram per cubic meter | $kg\ m^{-3}$ |
| current density | ampere per square meter | $A\ m^{-2}$ |
| magnetic field strength | ampere per meter | $A\ m^{-1}$ |
| luminance | candela per square meter | $cd\ m^{-2}$ |

For ease of understanding and convenience, several SI-derived units have been given special names and symbols. A selection of those used in space weather studies is given below.

Table B.3 *SI-derived units with special names and symbols*

| Derived quantity | Name | Symbol |
|---|---|---|
| frequency | hertz | Hz |
| force | newton | N |
| pressure, stress | pascal | Pa |
| energy, work, quantity of heat | joule | J |
| power, radiant flux | watt | W |
| electric charge, quantity of electricity | coulomb | C |
| electric potential difference, electromotive force | volt | V |
| capacitance | farad | F |
| electric resistance | ohm | Ω |
| electric conductance | siemens | S |
| magnetic flux | weber | Wb |
| magnetic flux density | tesla | T |
| inductance | henry | H |
| Celsius temperature | degrees Celsius | °C |
| luminous flux | lumen | lm |
| activity (of a radionuclide) | becquerel | Bq |
| absorbed dose, specific energy (imparted), kerma[a] | gray | Gy |
| dose equivalent | sievert | Sv |

*Note:* [a] Kinetic energy released per unit mass (J $\text{kg}^{-1}$ or gray).

# Appendix C
# SI prefixes

Table C.1 *The 20 SI prefixes used to form decimal multiples and submultiples of SI units*

| Factor | Name | Symbol | Factor | Name | Symbol |
|---|---|---|---|---|---|
| $10^{24}$ | yotta | Y | $10^{-1}$ | deci | d |
| $10^{21}$ | zetta | Z | $10^{-2}$ | centi | c |
| $10^{18}$ | exa | E | $10^{-3}$ | milli | m |
| $10^{15}$ | peta | P | $10^{-6}$ | micro | μ |
| $10^{12}$ | tera | T | $10^{-9}$ | nano | n |
| $10^{9}$ | giga | G | $10^{-12}$ | pico | p |
| $10^{6}$ | mega | M | $10^{-15}$ | femto | f |
| $10^{3}$ | kilo | k | $10^{-18}$ | atto | a |
| $10^{2}$ | hecto | h | $10^{-21}$ | zepto | z |
| $10^{1}$ | deka | da | $10^{-24}$ | yocto | y |

# References

Akasofu, S.-I. (1964). The development of the auroral substorm. *Planet. Space Sci.* **12**, 273–282.

Ben-Ezra, Y., Pershin, Y. V., Kaplunovsky, Y. A., Vagner, I. D., & Wyder, P. (2000). Magnetic fractal dimensionality of the surface discharge under strong magnetic fields. Tech. report, Physics and Engineering Research Institute, Emek Hefer, Israel.

Carrington, R. C. (1860). Description of a singular appearance seen in the Sun on September 1, 1859. *Mon. Not. Roy. Astron. Soc.*, **20**, 13–15.

Chesley, S. R. & Ward, S. N. (2006). A quantitative assessment of the human and economic hazard from impact-generated tsunami. *Nat. Haz.* **38**, 355–374.

Clarke, A. C. (1945). V2 for ionosphere research? *Wireless World*, February, p. 58, Letter to the editor.

Galilei, G. (1613). *Istoria e Dimostrazioni Intorno Alle Macchie Solari e Loro Accidenti*. Rome: Giacomo Mascardi.

Gold, T. (1959). Motions in the magnetosphere of the Earth. *J. Geophys. Res.*, **64**, 1219–1224.

Hess, W. N., ed. (1968). *The Radiation Belt and Magnetosphere*. Waltham, MA: Blaisdell Publishing Company.

Jokipii, J. R. & Thomas, B. (1981). Effects of drift on the transport of cosmic rays. IV. Modulation by a wavy interplanetary current sheet. *Astrophys. J.*, **243**, 1115–1122.

Kivelson, M. G. & Russell, C. T., eds. (1995). *Introduction to Space Physics*. Cambridge, UK: Cambridge University Press.

Loomis, E. (1869). Aurora borealis or polar light. *Harper's New Month. Mag.*, **39**, 1–21.

Spinoza, B. (1677). *Ethics*. Everyman's Library, No. 481, 1963, London: Dent. Translated by A. Boyle.

# Historical bibliography

Akasofu, S.-I. (1989). The dynamic aurora. *Sci. Am.* **260**, 90–97.

Alexander, P. (1992). History of solar coronal expansion studies. *Eos Trans. AGU* **73**, 433, 438.

Baker, D. N., Allen, J. H., Kanekal, S. G. & Reeves, G. D. (1998). Disturbed space environment may have been related to pager satellite failure. *Eos Trans. AGU* **79**, 477, 482–483.

Biermann, L. F. & Lust, R. (1958). The tails of comets. *Sci. Am.* **199**, 44.

Bone, N. (1996). *The Aurora: Sun–Earth Interactions*. John Wiley & Sons, 2nd edition, published in association with Praxis Publishing Ltd, Chichester.

Burch, J. L. (2001). The fury of space storms. *Sci. Am.* **284**, 86–94.

Carlowicz, M. J. & Lopez, R. E. (2002). *Storms from the Sun: the Emerging Science of Space Weather*. Washington, DC: The Joseph Henry Press.

Chapman, S. (1967). History of aurora and airglow. In *Aurora and Airglow: Proceedings of the NATO Advanced Study Institute held at the University of Keele, Staffordshire, England, August 15–26, 1966*, ed. B. M. McCormac. New York: Reinhold Publishing Corporation, pp. 15–28.

Cliver, E. W. (1994a). Solar activity and geomagnetic storms: the first 40 years. *Eos Trans. AGU* **75**, 569, 574–575.

Cliver, E. W. (1994b). Solar activity and geomagnetic storms: the corpuscular hypothesis. *Eos Trans. AGU* **75**, 609, 612–613.

Dessler, A. J. (1967). Solar wind and interplanetary magnetic field. *Rev. Geophys.* **5**, 1–41.

Dooling, D. (1995). Stormy weather in space. *IEEE Spectrum* **32**, 64–72. doi:10.1109/6.387145.

Dwivedi, B. N. & Phillips, K. J. H. (2001). The paradox of the Sun's hot corona. *Sci. Am.* **284**, 40–47.

Foukal, P. V. (1990). The variable Sun. *Sci. Am.*, **262**, 34–41.

Gillmor, C. S. & Spreiter, J. R., eds. (1997). *Discovery of the Magnetosphere*, volume 7 of *History of Geophysics*. Washington, DC: American Geophysical Union.

Hewish, A. (1988). The interplanetary weather forecast. *New Scientist* **118**, 46–50.

Holzworth, II, R. H. (1975). Folklore and the aurora. *Eos Trans. AGU* **56**, 686–688.

Joselyn, J. A. (1992). The impact of solar flares and magnetic storms on humans. *Eos Trans. AGU* **73**, 81, 84–85.

Kappenman, J. G. & Albertson, V. D. (1990). Bracing for the geomagnetic storms. *IEEE Spectrum*, **27**, 27–33.

Lerner, E. J. (1995). Space weather. *Discover* **16**, 45–61.

Meadows, A. J. & Kennedy, J. E. (1981). The origin of solar–terrestrial studies. *Vist. Astron.* **25**, 419–426.

Parker, E. N. (1964). The solar wind. *Sci. Am.* **210**, 66–76.

Rishbeth, H. (2001). The centenary of solar-terrestrial physics. *J. Atmos. Sol. Terr. Phys.* **63**, 1883–1890.

Silverman, S. (1997). 19th century auroral observations reveal solar activity patterns. *Eos Trans. AGU* **78**, 145, 149–150.

Siscoe, G. (2000). The space-weather enterprise: past, present, and future. *J. Atmos. Sol. Terr. Phys.* **62**, 1223–1232.

Stern, D. P. (1989). A brief history of magnetospheric physics before the spaceflight era. *Rev. Geophys.* **27**, 103–114.

Stern, D. P. (1996). A brief history of magnetospheric physics during the space age. *Rev. Geophys.* **34**, 1–31.

Suess, S. T. & Tsurutani, B. T., eds. (1998). *From the Sun: Auroras, Magnetic Storms, Solar Flares, Cosmic Rays*. Washington, DC: American Geophysical Union.

Van Allen, J. A. (1975). Interplanetary particles and fields. *Sci. Am.* **233**, 161–173.

# Index

f indicates a figure, t a table, p a color plate, and definitions and key concepts are in bold

$r_E$, 28
Kp, 128

acceleration, 15, **48**
accretion, **18**
active regions, **28**
Alfvèn waves, **12**
Alfvèn, Hannes, 12
Alpha Centauri, f6, 42, 106
alpha radiation, 97
apogee, **81**
Apollo 11 Mission, 13
Apollo 16 Mission, 101
Apollo 17 Mission, 101
Appleton layer, 75
Appleton, Edward, 12, 75
Archimedes, 39
Archimedian spiral, **39**, 40
asteroid, 104, 111, 113, 115–117, 119, 120, f122
astronaut, 44, 54, 79, 80, 95, 101–105, 107
astronomical unit, f6, **15**, 42, 90
atmosphere, 2, **3**, f4, 5, 6, 9, 11–13, 20, 22–24, 44, 66, 68, 70–77, 81, 82, 85, 95, 96, 103, 112–116, 118
  Earth's upper, 68, 71
  solar, 3, 11, 20, f21, 26, 29, 31, 38
  stellar, 42
aurora, 2, 4–6, f7, 9, 12, 54, 59, 60, f61, 73, **74**, 75, 77
aurorae, 64, 86
Auroral Electrojet (AE) index, **60**
Avogadro's number, **52**

bar (b), 103
Becquerel, Henri, 11, 99
becquerels, 99
beta radiation, 97
Biermann, Ludwig, 12
blackbody radiation, **35**
Boltzmann constant, **18**, 69
bow shock, **42**, 55
Brahé, Tycho, 5, 89

Carrington, Richard, 10, f10, 12, 25, 48
Cavendish, Henry, 5
Celsius, Anders, 9
Chapman, Sydney, 12
Chapman–Ferraro currents, 59
chromosphere, 22, **26**
Clarke, Arthur C., 104
climate, **1**, 112–115, 117, 118, 121
  change, 112, 113, 115, 116, 122
climate cycle, 112
comet, 9, 12, 37, f38, 104, 113, 115, 117
compass, 4, 6, 7, 9, 63, 64
conduction, **19**
conservation of energy, 119, 120
convection, **19**, **21**, 25, 57, 58, f58, 59
  plasma sheet, 57
  solar, 25, 26, 28
convection zone, 21, f25
Cook, Captain James, 6
Copernican heliocentric theory, 5, 6, 90
corona, 31, **37**, 41
coronagraph, 26, 38, 41
coronal mass ejections (CME), **3**, **41**, 59
corpuscular radiation, **11**, 95–97
cosmic rays, **43**, 44, 72, 114, 115
critical frequency, 76
Crookes, William, 11
Curie, Marie, 96
curies, 99
current sheet, 58

d'Ortour de Mairan, Jean-Jacques, 5
density, **18**, 32, 41
  chromospheric, 26

density (cont.)
  ionosphere, 86
  magnetic energy, 58
  mass, 70
  number, 18, t41, 70
  of air, 31, 32, 52
  plasma, 53, 57
  solar wind, 55
Descartes, René, 5, 45
dielectric discharging, 83
dipole, **51**, 63
dipole magnetic field, 7, 30, 50, f51, **51**, 52, 54, 63, 74
direct ionization radiation, 97
Disturbed Storm Time Index (Dst), **54**, 59, f59
Doppler effect, **22**, 25, 33
Dungey, James, 12

Einstein, Albert, 23, 24, 33, 49, 107
  special relativity theory, 106
electric current, 50, 53, f53, **54**, 64, 88, 89, 93
electric generator, 87
electromagnetic (EM) radiation, 2, 10, 11, 18, 20, 22, 24, 29, 31–35, 62, 71, 76, 85, 95–98, 107, 118
electromagnetic radiation spectrum, 31, 33, 34, f35, 74, 96
electrons, 2, 10, 11, 23, 34, 37, 38, 43, 44, 62, 66, 67, 71, 77, 83, 96–98, 102
  energetic, 54
energetic particles, 29, 41, 43, 54, 60, 71, 74, 97, 102, 106, 118
  and E region ionization, 73
  solar, 43
  trapped in radiation belts, 82
equilibrium, 19, **70**
  hydrostatic, **18**, **70**
escape velocity, 38

Faraday's law of induction, 87, 88, 93
Ferraro, Vincenzo, 12, 59
Fitzgerald, George, 11
flare, solar, 10, f10, 12, 28, 29, 48, 74
  energy of, 29
flux ropes, 41
flux tube, **56**, 57
force, **49**
Franklin, Captain John, 6
free radical, 98
frequency, 22, **31**, 32–36, 76, 77, 87, 96
  of a wave, 33
  of EM radiation, 77
  relationship to energy, 96
  relationship to period, 32
  SI unit, 32
  sound wave, 22
  spectral, 22, 27

Galilei, Galileo, 5, **7**, f8, **8**, 13, 24, 25
gas constant, 69
Gassendi, Pierre, 5
geomagnetic field, 6, 7, 9
geomagnetic storm, 9, f11, 12, 42, 48, f59, **59**, 61, 64, 74, 86–89, 93, 121
  enhanced ionospheric currents, 88
  main phase, 59
  of 1859, 48
  of 1989, 82
  recovery, 59
  recovery phase, 60
  sudden storm commencement, 59
geosynchronous orbit, 52, 54, 60, 80, f80
Gilbert, William, 6
Giovanelli, Ronald, 12
Global Positioning System (GPS), 13, 79, 85–87
Graham, George, 6, 9
granulation, 25
  solar, 25
gravitational constant, 91
grays (Gy), 99
Greenwich Observatory, 10, 47

Hadean, 115
Halley, Edmond, 9
heat, **19**
heat transfer, 19
Heaviside, Oliver, 11, 75
heliopause, **42**
helioseismology, **22**
heliosphere, 2, 37, 38, 41–43, 67, 106, 114
heliospheric current sheet, f40, **40**
helium, 20, 43, 53, 66, 97
high-Earth orbit (HEO), 80, 82, f80
high-frequency (HF) radio communication, 72, 85, 86
Hiorter, Olaf, 9

Ice Age, 112
ideal gas law, **18**, **69**
immune system, 97, 98, 100
inner magnetosphere, 52, f52, 58
International Space Station (ISS), 82, 102, 104

interplanetary magnetic field (IMF), **12**, 38, f40, 114, 115, 121
interstellar medium (ISM), 42, 105
interstellar space, 2, 42, 43, 106
ionization, **18**, 71–74, 76, 77, 97
ionosphere, 12, 53, 54, 58, f58, 59, **68**, f69, 71–76, f76, 77, 85–88
  D region, f69, 72, 73
  E region, 72, 73, 74
  F region, f69, 73
  sporadic E, 73

Jupiter, 5, 15, 52, 91

Kelvin, Lord, 10
Kennelly, Arthur, 11, 75
Kepler, Johannes, 89, 90, 92
kinetic energy, 43, 52, 53, 56, 58, 105, 118–120, 122
Kua, Shen, 6

light-year, 42, 108, 118
Little Ice Age, 113
Lockyer, J. Norman, 11
Lodge, Sir Oliver, 11
Loomis, Elias, 6
Lorentz force, 74
Lorentz, Hendrik Anton, 27
low-Earth orbit (LEO), 68, 80, f80, 81, 82, 95, 101, 104
luminosity, **23**, 112–114
  solar, 20, 23, 24

magnetic cloud, 41
magnetic field, 2, 4–7, 9–12, 18, 19, 24, 26–30, 37–39, f39, 40–43, 50–54, 56, 57, f58, 59, 60, 64–66, 74, 77, 88, 92, 102, 114, 121
magnetic field merging, 12
magnetic index, 54
magnetic lobe, 57
magnetic poles, 28, 51, 63
  of Earth, 51
magnetic reconnection, f55, **56**, 57, 58, 121
magnetopause, **55**
magnetosheath, **55**
magnetosphere, 12, 16, **50**, 52, f52, 53, f53, 54, 55, f55, 56–61, 67, 73–75, 95, 101, 121, 122
magnetotail, f55, 57, **57**, 58
Marconi, Guglielmo, 11, 75, 86
Maunder Minimum, 113, 115
Medieval Climatic Optimum, 115
medium-Earth orbit (MEO), 80, f80, 82, 86
Mercury, 8, 52, 68, 70, 95
mesosphere, 72
meteor, 73, 82, 116
Meteor Crater, 116
meteorology, **1**
micro-gravity, 104
micrometeoroids, 95
Milanković cycle, 113
Muncke, Georg Wilhelm, 6
muons, 44

nebula
  solar, 18, 20
  stellar, 42
Neckham, Alexander, 6
Neptune, 52
neutral atmosphere, 73, 81
neutrino, 44
Newton, Sir Isaac, 49

Occupational Safety and Health Administration (OSHA), 101
open-field region, 40
outer magnetosphere, 52
ozone layer, 85, 118, 122

Parker, Eugene, 12
perigee, 81
photoelectric effect, 34, 83, 106
photoionization, 71–74, 76, 77
photon, 21, 24, **34**, 71, 72, 76, 77, 96–98, 106
photosphere, 21, 22, **24**, 25, 26, 28, 29, 38
pions, 44
pipeline, 13, 79, 87, 89
  corrosion, 87, 89
plages, 26
Planck constant, **33**, 77, 96
plasma, 11, 12, 37, 38, 40–42, 48, 52–54, 56, 58, f58, 66, 67, 87
plasma sheet, 54, **57**, 75
plasmapause, **53**, 58
plasmasphere, f8, **52**, 53, **53**, 56, 73
Pluto, 42, 46
polar cap, 57, 58, 60, f61, 72
polarity, 28, 31
  IMF, 58
  of Sun's field, 29
  of sunspot pairs, 31
positrons, 43, 44
potential energy, 118–120

power grids, 3, 13
pressure gradient, **69**, 70
protons, 2, 18, 23, 24, 38, 54, 62, 71, 96

radiation, **20**
radiation sickness, 100, 101
radio waves, 5, 11, 32, 72, 73, 75, 76, f76, 87, 96
rads, 99
recombination, 72
reconnection, 12, 56, 58, 59
Red Giant, 95
reflection, 5, 11, 22, 73, 75, 76, f76, 86, 96
refraction, 22, 34, 76, f76
relativistic electrons, 83, 108
rems, 99
ring current, f52, 53, **53**, 54, 59, 61
rotation
  Earth, 71, 74, 112
  solar, 25, 28, 38, f39
    differential rate, 25, 28, f28, 30

Sabine, Edward, 9
satellite drag effects, 82
Saturn, 52
Schwabe, Samuel Heinrich, 8
scientific notation, **15**
shock wave, 55
SI units, 2, **14**, 15, 32, 47, 62, 93, 103, 119, a125
sieverts (Sv), 99
single event upset (SEU), 83, **84, 85**
skywave, 86
solar constant, **22**, 114
solar cycle, p4, **29**, 31, 36, 60, 74, 114
  11-year, 8, 24, 29, 113, 114
  sunspot cycle, 8, 29, f30, 86
solar eclipse, f5, 11, 26, 37, 38
solar prominence, 11, 26, 28
solar spectrum, 20, f35
solar storm, 72, 74, 85, 101
Solar System, 1, 6, 20, f38, 43, 89, 104–106, 108, 114, 115
solar wind, **2**, 12, 16, **26**, 31, 36, 37, f38, **38**, 39–41, t41, 42, 46, 48, 50, 52, 54–57, 59, 67, 97, 121
sound speed, 41
southward IMF, 58, 59, 121
space weather, **1**, 2–5, p10, 13, 17, 29, 50
spectrograph, solar, 11
speed, **47**
spicules, 26
spiral
  Archimedian, f39
Sputnik, 12
standard solar model, **20**, 22
stratosphere, 3, 116, 117
Størmer, Carl, 5
substorm, 60, **60**, 61, f61
sudden ionospheric disturbances (SID), 72
sunspots, 4, 7, 8, f8, 9, 10, **24**, 25, **25**, 28, 29, 113, 114
supergranulation, 25
supergranulation cell, 26
supernova, 43, 117, 118, 122
  Type I, 117, 118
  Type II, 117
surface charging, 83

terrella, 6
thermodynamics, 10, 17, 18, 35
thermonuclear fusion, 18, 43
thermonuclear reactions, 18, 20, 22, 117
thermosphere, f4, **68**, 71, 114
Thomson, J. J., 11
topside ionosphere, 73
transformer, f11, 87–89, 93
transition region, 26
triangulation method, 5, 87
tropopause, 3
troposphere, **2**, 3, 71, 114
Tunguska, 116

upper atmosphere
  density, 68
Uranus, 52

Van Allen, James Alfred, 53
Van Allen radiation belts, **53, 54**, 82, 83
velocity, **47**
  of a wave, **33**

Wein, Wilhelm Carl Werner, 35
white blood cell counts, 98, 100
Wilke, Johann, 9
Wolf, Rudolph, 9
work, 119

Yohkoh satellite, 29
Young, Charles, 11

Zeeman effect, **27**
Zeeman, Pieter, 27
Zurich Observatory, 8, 9