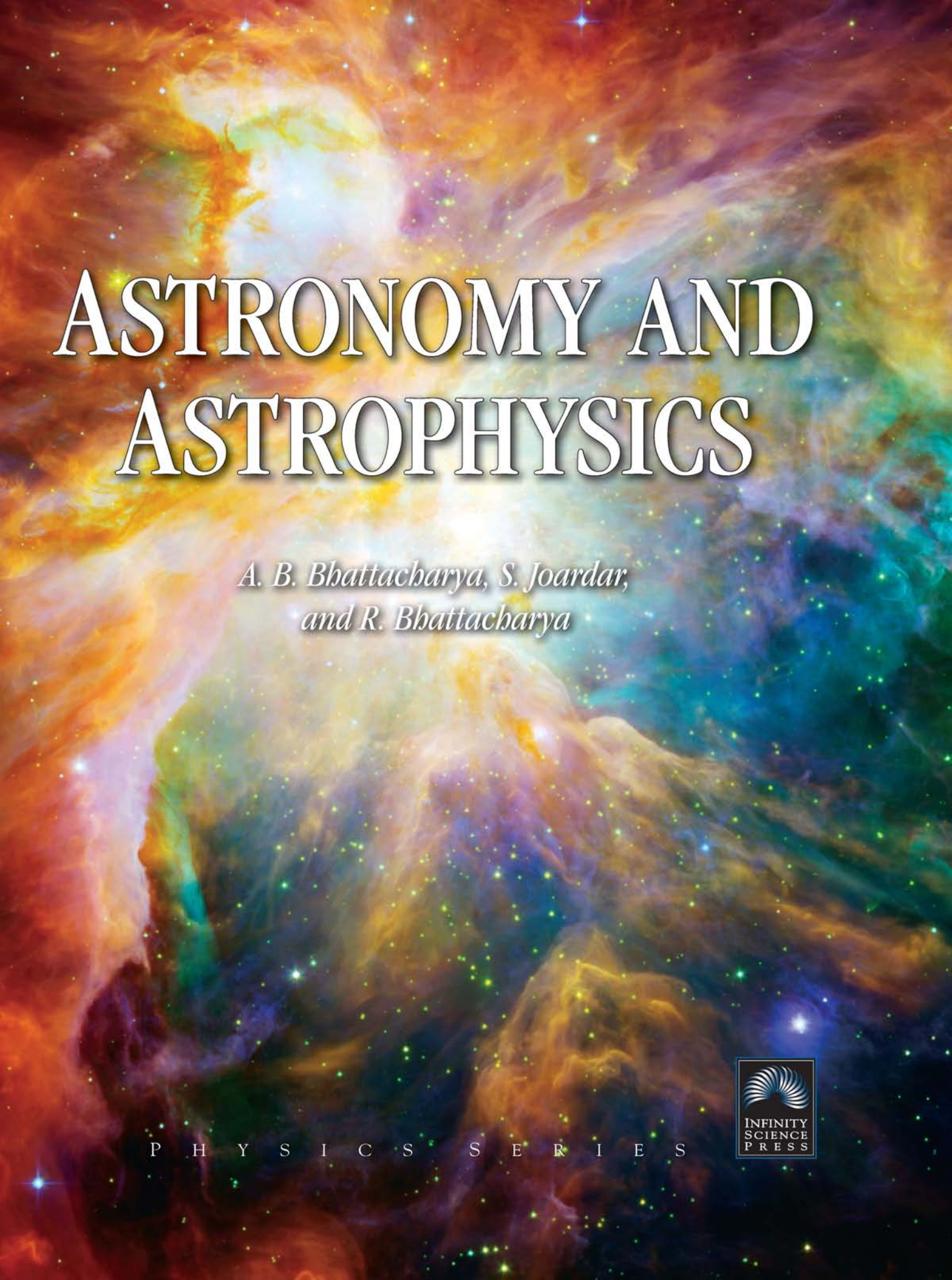
ASTRONOMY AND ASTROPHYSICS

A. B. Bhattacharya, S. Joardar, and R. Bhattacharya



INFINITY SCIENCE PRESS

PHYSICS SERIES

ASTRONOMY AND ASTROPHYSICS

LICENSE, DISCLAIMER OF LIABILITY, AND LIMITED WARRANTY

The CD-ROM that accompanies this book may only be used on a single PC. This license does not permit its use on the Internet or on a network (of any kind). By purchasing or using this book/CD-ROM package(the "Work"), you agree that this license grants permission to use the products contained herein, but does not give you the right of ownership to any of the textual content in the book or ownership to any of the information or products contained on the CD-ROM. Use of third party software contained herein is limited to and subject to licensing terms for the respective products, and permission must be obtained from the publisher or the owner of the software in order to reproduce or network any portion of the textual material or software (in any media) that is contained in the Work.

INFINITY SCIENCE PRESS LLC ("ISP" or "the Publisher") and anyone involved in the creation, writing or production of the accompanying algorithms, code, or computer programs ("the software") or any of the third party software contained on the CD-ROM or any of the textual material in the book, cannot and do not warrant the performance or results that might be obtained by using the software or contents of the book. The authors, developers, and the publisher have used their best efforts to insure the accuracy and functionality of the textual material and programs contained in this package; we, however, make no warranty of any kind, express or implied, regarding the performance of these contents or programs. The Work is sold "as is" without warranty (except for defective materials used in manufacturing the disc or due to faulty workmanship);

The authors, developers, and the publisher of any third party software, and anyone involved in the composition, production, and manufacturing of this work will not be liable for damages of any kind arising out of the use of (or the inability to use) the algorithms, source code, computer programs, or textual material contained in this publication. This includes, but is not limited to, loss of revenue or profit, or other incidental, physical, or consequential damages arising out of the use of this Work.

The sole remedy in the event of a claim of any kind is expressly limited to replacement of the book and/or the CD-ROM, and only at the discretion of the Publisher.

The use of "implied warranty" and certain "exclusions" vary from state to state, and might not apply to the purchaser of this product.

ASTRONOMY AND ASTROPHYSICS

A. B. Bhattacharya
S. Joardar
R. Bhattacharya

Infinity Science Press LLC
Hingham, Massachusetts
New Delhi, India

Publisher: David Pallai

INFINITY SCIENCE PRESS LLC
11 Leavitt Street
Hingham, MA 02043
Tel. 877-266-5796 (toll free)
Fax 781-740-1677
info@infinitysciencepress.com
www.infinitysciencepress.com

This book is printed on acid-free paper.

A. B. Bhattacharya, S. Joardar, R. Bhattacharya. *Astronomy and Astrophysics.*
ISBN: 978-1-934015-05-6

Library of Congress Cataloging-in-Publication Data

Bhattacharya, Asit B.
Astronomy and astrophysics / Asit B. Bhattacharya, Shubhendu Joardar, Rina Bhattacharya.
p. cm.
ISBN 978-1-934015-05-6 (hardcover)
ISBN 978-0-7637-8269-6 (e)
8177
1. Astronomy. 2. Astrophysics. I. Joardar, Shubhendu. II. Bhattacharya, Rina, 1955- III. Title.
QB500.B48 2008
520--dc22

2008001933

Our titles are available for adoption, license or bulk purchase by institutions, corporations, etc. For additional information, please contact the Customer Service Dept. at 877-266-5796 (toll free).

Requests for replacement of a defective CD-ROM must be accompanied by the original disc, your mailing address, telephone number, date of purchase and purchase price. Please state the nature of the problem, and send the information to INFINITY SCIENCE PRESS, 11 Leavitt Street, Hingham, MA 02043. The sole obligation of INFINITY SCIENCE PRESS to the purchaser is to replace the disc, based on defective materials or faulty workmanship, but not based on the operation or functionality of the product.

Table of Contents

PREFACE

The subjects of astronomy and astrophysics have grown tremendously in scope during the last three decades. In fact, over the past few years there has been an explosive growth of these subjects, in particular, about new ideas and concepts as well as a large increase in available information and data in all of the solar technologies. This demands for a comprehensive textbook describing the present state of knowledge, offering the different techniques and instruments developed in this relatively young science. It is hoped that the present book will fulfill this requirement to a large extent.

The subject may be treated traditionally under a single banner; however, at the present level of our understanding it can be well classified into two distinct parts: (*i*) different techniques and instruments employed and the results obtained thereof and (*ii*) purely theoretical approaches. The present volume deals in a comprehensive manner both the categories where suitable texts are highly lacking. This volume has been prepared primarily for post-graduate students of different universities, though many of the sections will prove useful for practicing engineers and scientists actively engaged in advanced research.

To ensure the highest degree of reliability, besides a large number of reference books, we have consulted freely with many specialists who have provided an up-to-date overview of their areas and have given very valuable information and ideas. The authors are grateful for their indirect contribution to the text. We would like to acknowledge Sri Dibyendu Joardar, for drawing many complicated diagrams. Finally, we would like to take this opportunity to express our deepest gratitude to many of the astronomers and other scientists whose works are directly or indirectly included in the text.

— Authors

Chapter 1

ASTRONOMICAL UNIVERSE AND BASIC TECHNIQUES OF RADIO ASTRONOMY

1.1 INTRODUCTION

The recorded history of astronomy reveals that in the beginning there was almost no difference between astronomy and astrology. The first references are found in the Indian astronomy-astrology, in the *Rig Veda* (2000 B.C[1]). Around 1800 BC, Yajnavalkya advanced a 95-year cycle to synchronize the motions of the Sun and the Moon. Lagadha wrote a text on astronomy in 1350 BC. Horoscopes contained the positions of Sun, planets and nodes of the Moon (Rahu and Ketu)[2]. The stars were recognized as the Sun, with the Sun being the center of the Universe. Aryabhatta in the 5th century propounded the heliocentric theory of gravitation, later translated as "Aryabhatta's Magnum Opusthus." He wrote 1,582,237,500 rotations of the Earth equals 57,753,336 lunar orbits. It is perhaps the oldest accurate fundamental astronomical ratio. The 7th century BC book by

[1] The Vedas existed much before 2000 BC but not in the form of written text. In fact they were handed down by teaching and memorizing from generation to generation.

[2] The two points formed by the intersection of the ecliptic's great circle with the same of the Moon's orbit as seen from the Earth (eclipse points).

Bhramhagupta, 'Brahmasphutasiddhanta' describes the circumference of the Earth as 5000 Yojana[3] (36,000 km). In the 6th century BC, Varahamihira defined a term 'Gurutwakarshan' (attraction towards the master) as an explanation to the falling bodies. Another astronomical text 'Surya-Siddhanta' appeared around 400 AD. The Egyptian astronomer-priests divided the year into 12 months of 30 days, followed by a five day feast period for annual festivals. They maintained a 365.25 day separate calendar for agriculture. In 1100 BC, Amenhope created a catalog listing 5 constellations. The Babylonian astronomy dates back to 1800 BC. They divided the sky into zones and estimated the apparent paths of the Sun, Moon and planets. By 350 BC, the length of the solar year was known to the Chinese as 365.25 days. In the 6th century BC, the Greeks Anaximander, Anaximenes, Pythagoras, and Anaxagoras propounded early theories about the Sun and planets. In the 3rd - 4th centuries BC, Eudoxus and Heraclides propagated a model with Earth as the center. Using philosophical arguments, Aristotle proved the Earth as the center. Later Aristarchus brought forth a heliocentric model. He, Eratosthenes, and Hipparchus through observations, cataloged the positions of some stars. Around 125 AD, Ptolemy summarized the Greek astronomy, called 'Almagest' by the Arabs, who in the next 800 years developed their astronomical tradition. Thus over many years, astronomy remained attached with astrology, religion and philosophy.

Indeed, a major divergence of astronomy and astrology came with the work of Nicholas Copernicus in the sixteenth century. Copernicus explained the complicated motion of the planets in terms of a stationary Sun about which revolved Mercury, Venus, Earth, Mars, Jupiter, and Saturn. The Moon would still have to go around the Earth, but the looping motions of the other planets became simple to explain if the Earth itself moved about the Sun. Copernicus forced a major change in the prevailing philosophy of a human-centered Universe, and the Copernican Revolution is considered one of the major turning points of science. Galileo and Kepler were quick to adopt the Copernicus system and they provided much astronomical evidence in support of it. Later, Galileo was forced by the Church to recant his espousal of a moving Earth. The Copernican viewpoint received its ultimate triumph in the seventeenth century with the work of Isaac Newton.

Newton established the laws of planetary motion in a Sun-centered system derived from Newton's laws of mechanics and gravitation. He also showed how his theory of gravitation could explain the tides raised on Earth by the Sun or Moon. The idea was that the Sun or Moon pulled hardest on the side of the oceans facing toward it. With the maturity of astronomy, scientists attempted to give mechanistic explanations for natural phenomena like seasons, tides, and planetary motions on the basis of laws formulated and tested in laboratories.

---

[3] 1 Yojan = 7.2 miles.

Mercury, Venus, Earth, Mars, Jupiter, and Saturn are only six of the nine planets that revolve around the Sun; the other three are Uranus, Neptune, and Pluto, though Pluto is controversial as a planet. However, planets and their satellites are not the major constituents of the Universe. The Sun is a star, the closest one to Earth, and it contains about 99.9 percent of the mass of our solar system. Our Galaxy contains more than 10^{11} stars and it is only one of myriads in the observable Universe. The observable Universe contains about 10^{10} galaxies. Thus, in the observable Universe there are about 10^{10} galaxies $\times$ 10^{11} stars/galaxy = 10^{21} stars.

1.2 THE BIG BANG NUCLEOSYNTHESIS

Today, we know that the Universe is expanding and the major constituents of the matter are hydrogen and helium with a smidgen of heavier elements like oxygen, carbon, and iron. If we extrapolate back in time by 10 to 20 million years, we find that the density of the Universe was almost infinite. The entire space was filled with thermal radiation at 3 K. In 1965, Penzias and Wilson discovered the cosmic microwave background radiation. These facts agree with the possibility that the Universe started with a sudden decompression, or a *Big Bang*. It happened simultaneously everywhere in space. The standard model of the Big Bang Theory is explained in the following eight steps.

1. **Before the Universe:** Before space and time existed, the entity of the Universe was compressed into the confines of an atomic nucleus also known as a singularity. The temperature was extremely high of the order of trillions of degrees or more and the density was almost infinite.
2. **The Big Bang:** The creation process of the Universe began with an expansion and not an explosion as the name suggests.
3. **The Inflationary Epoch (immediately after the Big Bang):** There is no compelling physical model for the first 10^{-35} seconds. However, we give a picture for some possible happening during this period. The four forces of nature: strong nuclear, weak nuclear, electromagnetic, and gravity were combined as a single super force called *Wald*. Elementary particles known as quarks begin to bond in trios forming photons, leptons, antileptons, gluons, photons, neutrinos, and antineutrinos. There were smaller amounts of neutrons and protons, approximately 1 in every one billion photons, electrons, or neutrinos. At this stage, the density of the Universe was assumed to be nearly 1094 g/cm^3 with the majority of this being radiation. The Universe was expanding much faster than the speed of light during this creation and annihilation of these particles. The rate of inflation of the Universe was assumed to be doubling every 10^{-5} seconds.

4. **Nanoseconds after the Big Bang:** At this point the Universe was in a ionized plasma state with radiation and matter inseparable. There were equal amounts of particles and antiparticles. As the temperature fell below 10^{14} K, the quarks, antiquarks, and gluons annihilated and transformed into less massive particles. The quarks were slightly more in number than the antiquarks and as a result some quarks were left behind to form protons and neutrons. The leptons and antileptons were annihilated with the fall in temperature.
5. **Between a millisecond to a second after the Big Bang:** As the temperature fell around 10^{10}, the density of the Universe drastically reduced to neutrinos to effectively interact with the matter. These non-interacting decoupled neutrinos formed a universal gas within the Universe.[4] Immediately after this, the annihilation of electron-positron pairs removed all positrons and most electrons.
6. **100 seconds after the Big Bang:** As the temperature fell to 10^{9} K, protons and neutrons slowed down enough to allow nucleosynthesis to take place. Atomic nuclei of helium were produced as two protons and neutrons each bonded. For every helium nucleus formed, there were about ten protons left over allowing for 25 percent of the Universe to be comprised of helium. The light nuclei were formed by the combination of neutrons and protons.
7. **Thirty minutes after the Big Bang:** At this stage an important phase of the expansion occurred when the creation of photons increased through the annihilation of electron-positron pairs. It is fortunate that the Universe began with slightly more electrons than positrons which insured that our Universe was able to form the way it has.
8. **Three-hundred thousand years after the Big Bang:** The Universe continued to expand and cool until it reached a temperature of 10,000 K. It allowed for helium nuclei to absorb free-floating electrons and to form helium atoms. Also, hydrogen atoms were bonding together and forming lithium. As the temperature reduced to 4000 K, hydrogen and helium nuclei combined with electrons to form neutral hydrogen and helium atoms. It was here that the density of the Universe had reduced so much that light could be perceived. Until this point photons continued to be trapped within matter. Thus, the Universe became transparent to electromagnetic radiation.[5]

---

[4] Because of the expansion of space this gas has cooled to a temperature of 2 K as of today. However, until today it has not been possible to detect this universal background of neutrinos.

[5] The radiation freed from interacting with matter has now cooled to a temperature of 3 K due to the expansion of space, which is the cosmic microwave background radiation.

1.3 ROUGH SCALES OF THE UNIVERSE

We now consider the scales of some astronomical objects. Listed in Table 1.1 are the very rough sizes, masses, and ages of the Sun, the Galaxy, and the observable Universe. For comparison, the quantities are listed here with that of a small child.

Table 1.1: Rough scales of the Universe.

Object	Size	Mass	Age
Child	10^2 cm	10^4 gm	10^8 sec
Sun	10^{11} cm	10^{33} gm	10^{17} sec
Galaxy	10^{23} cm	10^{45} gm	$\geq 10^{17}$ sec
Observable Universe	10^{28} cm	10^{55} gm	$\geq 10^{17}$ sec

Table 1.1 gives the age of the Universe as somewhat greater than 10^{17} sec. Astronomers believe that the time since creation is between 10 and 20 billion years. Table 1.1 notes that these estimates apply to the observable Universe. This distinction between the Universe and the observable Universe is required because astronomers have not yet settled the question of whether the Universe is open or closed. If the Universe is open, space-time (and the number of stars in the Universe) is infinite. As time had a beginning, but no end, the amount of space and the number of stars in it are actually infinite.

1.4 CONTENTS OF THE UNIVERSE

The inhabitants of the Universe are mainly stars of which the Sun is only one example. Besides suns of all colors there are the tremendously distended red giants and the mysteriously tiny white dwarfs. The Sun itself is an average star, neither very massive nor very light.

Varieties of stars live in our Galaxy either singly or in pairs. Sometimes, hundreds or thousands of stars can be found in loose groups called open clusters. The Pleiades is a fairly young open cluster, and the reflection nebulosity surrounds the *seven sisters* of this cluster. The oldest stars in our Galaxy are found in tighter groups called globular clusters, which may contain more than a million members.

The space between the stars is not completely empty. The diffuse matter between the stars is called the interstellar medium. Clouds of gas and dust, as well

as energetic cosmic-ray particles gyrating in magnetic fields, reside in the space between stars. Other objects like the crab nebulas are known to be the remnants expelled by stars which died in titan tic supernova explosions. The cores left behind neutron stars and black holes and represent some of the most intriguing denizens of the astronomical kingdom. Other stars die less violently, ejecting the outer shell as a planetary nebula which exposes the hot central core.

Stars and the material between them are called galaxies. Our own galaxy, the Milkyway System, happens to be one of the largest systems in the *Local Group* of two dozen or so galaxies. The other is the Andromeda galaxy. It stretches more than 100,000 light years from one end to the other, and it is situated about two million light years from us. Apart from small groups (like our own Local Group) galaxies are found in great clusters, containing thousands of members. Most cluster galaxies are elliptical rather than spirals found in general field clusters.

1.5 ASTRONOMICAL INSTRUMENTS AND MEASUREMENTS

Undoubtedly, the human eye is a poor photon detection and storage instrument. All types of electronic equipment help the astronomer (*i*) measure the brightness of astronomical objects (photometry), (*ii*) calculate the wavelength distribution and spectral features (spectrophotometry and spectroscopy), (*iii*) assess the amount of polarization (polarimetry), and (*iv*) fix the position of the object in the sky (positional astronomy). The one basic tool of modern astronomy is definitely the computer. This can be used to run many instruments more or less automatically, guide the telescope, reduce data, and help construct theories. The basic astronomical measurements are listed in Table 1.2.

Table 1.2: Some basic astronomical measurements.

Parallax	The apparent angular shift of nearby objects against the positions of distant objects. This shift yields the distance to the nearby object by triangulation, as the Earth-Sun distance is known.
Proper motion	The real angular shift of nearby objects against the positions of distant objects. This shift yields the component of velocity perpendicular to the line of sight provided the distance to the object is known.

Doppler shift	The shift in wavelength of a spectral line in comparison to its actual value. These are usually associated with motion along the line of sight.
Apparent brightness	The total brightness of an astronomical body. If the distance to the object is known, a measurement of its apparent brightness allows calculating luminosity.
Color	The difference in apparent brightness when measured through two filters that pass light of two different wavelength bands. This technique gives a quantitative measure of the wavelength distribution of the continuum light and the surface temperatures of stars can be estimated.

1.6 RADIO ASTRONOMY

Radio Astronomy is the exploration of the Universe through naturally produced radio waves received from certain objects of extraterrestrial origin. Although it is a relatively young science, it has given rise to many great discoveries of remarkable objects and phenomena including radio galaxies, quasars, pulsars, cosmic masers, 3K microwave background radiation, and a large number of molecules in interstellar space such as water vapor, ammonia, alcohol, etc. Many of these and other related phenomena are being actively studied in many parts of the world.

The National Center for Radio Astrophysics (NCRA) has developed the Giant Meterwave Radio Telescope (GMRT), a most powerful radio telescope used at meter wavelengths, to investigate a variety of outstanding astrophysical problems in modern astronomy and cosmology. It is, in fact, one of the most challenging projects in basic science.

1.6.1 Radio Astronomy and the Cosmos

Radio Astronomy is the investigation of the Universe through naturally produced radio waves reaching us from a variety of celestial objects. This new science was born in 1932 when Karl Jansky at the Bell Labs in the United States discovered radio noise coming from our Milkyway Galaxy. Radio waves, like light waves, are also electromagnetic waves but with much longer wavelengths. Celestial objects usually radiate in different regions of the electromagnetic spectrum—such as light-rays, ultra-violet, optical light, infra-red, or radio waves—depending on their physical conditions, such as temperature, magnetic field, etc.

1.6.2 Sources of Cosmic Radio Waves

With the use of powerful radio telescopes, astronomers have discovered a series of remarkable objects and phenomena, such as radio galaxies, quasars, pulsars, supernova remnants, interstellar molecules, cosmic microwave background (CMB) radiation, etc. The strongest radio waves generally arise from highly energetic relativistic electrons in the presence of magnetic fields created in the aftermath of violent and catastrophic events occurring in stellar explosions and in the ultra-strong gravitational fields of the central regions of galaxies. The nuclei of some galaxies are assumed to contain massive *black holes* of millions of solar masses concentrated in a volume not much larger than that of our solar system. Black holes are objects so highly condensed that nothing (not even light) can escape out of their strong gravitational pull. In modern astronomy their study is one of the most fascinating and challenging problems without doubt.

Another very important development in Radio Astronomy was the discovery of 3K microwave background radiation in 1965 by Penzias and Wilson. This radiation was observed to be coming isotropically from all directions in the sky and is assumed to be the relic radiation from the early Big Bang phase of the Universe. This fundamental discovery was a great impetus to the study of the origin and evolution of the Universe.

1.6.3 Spectral Lines

Besides the continuum radio waves covering a wide range of radio frequencies, atomic and molecular processes also give rise to line radiation confined to a relatively narrow frequency range. The electromagnetic spectrum is shown in Fig. 1.1. The atmosphere of our Earth is transparent only to optical and radio waves. From the ionosphere, radio waves of wavelengths longer than about 15 m may be reflected. The most important of these lines occurs at 1420 MHz (i.e., a wavelength of 21 cm due to a hyperfine transition in the atoms of neutral hydrogen). The study of this line radiation has not only provided a wealth of information on the structures of our own as well as other nearby galaxies, but also has the potential to reveal the state of the early Universe before the formation of galaxies. The detection of radio lines from many complex organic molecules indicates that conditions favorable to the development of life may exist on many places in our Galaxy.

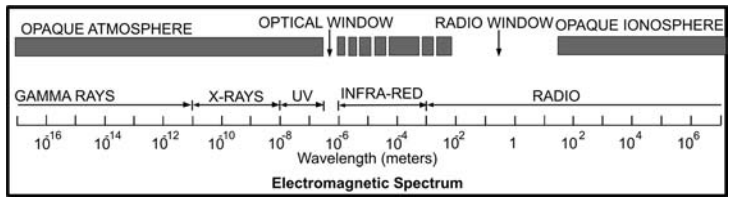


Figure 1.1: Availability of extra-terrestrial electromagnetic spectrum on Earth.

1.7 TECHNIQUES OF RADIO ASTRONOMY

The basic tool for exploring the radio Universe is a radio telescope. Essentially, it consists of a reflector antenna, a feed, and a receiver system. The reflector focuses the radio waves from a celestial object on to the feed. The feed is usually a dipole or a horn that collects the radio signals and feeds these weak signals to a sensitive low-noise amplifier where it is further processed by a sophisticated electronic receiver system. The two fundamental characteristics of a telescope are (*i*) high sensitivity to detect the weak radio waves from distant celestial objects and (*ii*) high angular resolution to distinguish features close together in the sky. The larger the collecting area or size of a telescope the greater its sensitivity as well as its resolving power.

1.8 INTERFEROMETRY AND APERTURE SYNTHESIS

Owing to the very long wavelengths of radio waves, ranging between about 1 cm and 10 m, the resolution of even a large, single radio telescope is relatively much poorer. At a wavelength of 1 m, for instance, a parabolic dish with a diameter of about 1 km would be required to obtain even the modest resolution of the human eye. Radio astronomers have overcome this limitation by development of an *interferometry* technique in which the voltage outputs of a pair of modestly sized radio telescopes are brought together to a common point and multiplied together. By using an array of telescopes spread over tens of kilometers such interferometric techniques can be utilized to synthesize much larger telescopes with effective diameters equivalent to the maximum separations or *baselines* between the most distant antennas in array. This process of *aperture synthesis* is greatly aided by the Earth's rotation, which constantly changes the orientation and effective lengths of different interferometric baselines as viewed from the radio source.

Radio astronomers can now obtain resolutions as high as a thousandth of an arc second through telescopes by the technique of Very Long Baseline Interferometry (VLBI). In VLBI radio signals are independently recorded by telescopes separated by thousands of kilometers and subsequently correlated in a computer. The resolution is, of course, much higher than possible with optical telescopes.

1.9 METER-WAVE ASTRONOMY

A major part of the pioneering work in Radio Astronomy after World War II was carried out at meter wavelengths. However, the development of interferometric

techniques and low-noise receivers found a steady shift in emphasis toward shorter wavelengths, where the Galactic background noise is smaller and higher angular resolution can be attained for a given antenna size or interferometer baseline. Most powerful radio-astronomical facilities in the world today operate in the centimeter and decimeter part of the radio spectrum even though there are many exciting astrophysical problems that are best studied at meter wavelengths. The importance of this part of the radio spectrum has been highlighted by a variety of astronomical research carried out using instruments built for meter-wave operation. In general, these instruments are limited by sensitivity and resolution and limited frequency coverage.

The largest radio telescopes constructed over the past two decades have been designed to operate mainly at centimeter and decimeter wavelengths. There are several reasons for this: (*i*) a radio telescope of a given aperture provides higher resolution at shorter wavelengths; (*ii*) a larger collecting area is necessary to reach similar sensitivities at larger wavelengths due to a steep increase of the galactic background temperature with wavelength; (*iii*) ionospheric irregularities introduce increasingly serious distortion in the wave-front from a distant radio source at wavelengths longer than about 20 cm; (*iv*) there is a great deal of man-made radio interference in the allocated radio astronomy bands at longer wavelengths that at shorter wavelengths, particularly in the western countries. Meter-wave astronomy, on the other hand, offers certain distinct advantages. It is sufficient to achieve only a moderate resolution at meter wavelengths in order to study the diffuse emission arising from the oldest relativistic electrons in galactic and extra-galactic radio sources. The radiation from these becomes dominant at meter wavelengths and is a useful probe of the evolution of the parent objects. Moreover, large antennas required for meter-wave operation can be constructed quite economically as their reflecting surface does not require high precision. Harmful effects of ionospheric irregularities at longer wavelengths can be largely overcome through the application of methods like self-calibration that have been developed for use with synthesis radio telescopes over the last decade.

1.10 SYSTEMS OF COORDINATES IN ASTRONOMY

To locate stars and other objects in the sky we consider all the objects to lie on the surface of a sphere with ourselves at the center of it. This imaginary sphere is known as the *celestial sphere*. For simplicity, we fix the radius of the sphere as unity. The direction of an object in the sky is then given in terms of its position on the celestial sphere. The point and the direction of the object can then be assigned two angle coordinates. As the principle underlying the astronomical coordinate systems is the same as the terrestrial system, we consider it first.

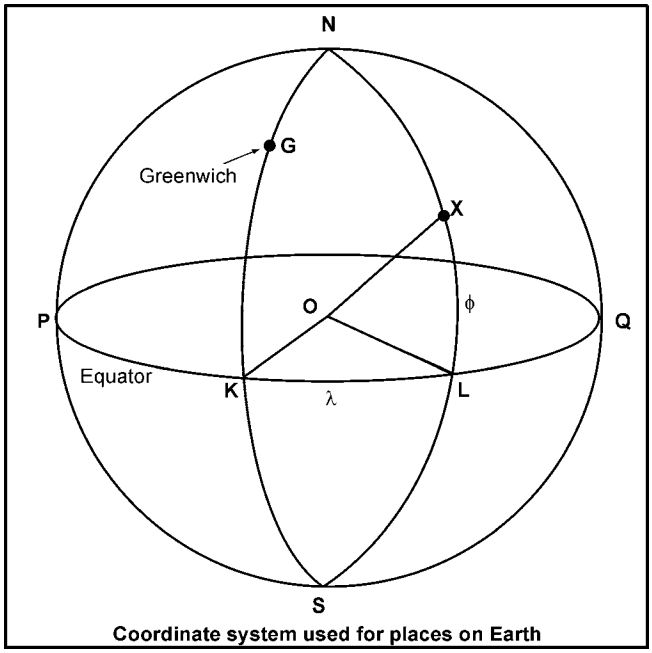


Figure 1.2: The coordinate system used for astronomical observations from Earth.

In Fig. 1.2, let NQSP be the globe where N represents the North Pole and S the South Pole. The plane of the circle QLKP is perpendicular to the line joining N and S and passes through the center of the sphere. This plane is the equator. The circles whose planes pass through the center of the sphere are known as great circles while if the plane of the circle does not pass through the center, it is called a small circle. Any half-circle through N and S is a meridian and the meridian through Greenwich is the principal or prime meridian. The point of intersection of the prime meridian with the equator is the reference point to measure arc lengths along the equator. The meridian through the fixed point X cuts the equator at L. The longitude (λ) of X is then called the spherical angle GNX or the angle KOL in the plane of the equator or the arc length KL. The other coordinate of X is called the latitude (ϕ). It is the arc length LX or the angle LOX. Latitude ranges from 0 to 90 degrees North or South, accordingly, as the plane is in the North or the South hemisphere. Likewise, the longitudes range from 0 to 180 degrees East or West depending on whether the plane is to the East or to the West of Greenwich. The 360 degrees of longitude is equal to a complete rotation of the Earth on its axis, which takes 24 hours. Hence, 15 degrees of longitude becomes equal to one hour of time (i.e., 1 degree of longitude becomes equal to 4 minutes of time).

The longitude-latitude system provides the essential requirements of an astronomical coordinate system, which are (*a*) a fundamental great circle (e.g., the

equator), and (*b*) a reference point or origin (e.g., the point K). Depending on the fundamental circle and reference point, the astronomical systems of coordinates are: (*i*) the Horizon System, (*ii*) the Equatorial System, (*iii*) the Ecliptic System, and (*iv*) the Galactic System.

Fig. 1.3 (*a*), (*b*), (*c*), and (*d*) represent, respectively, the above four coordinate systems used in astronomy. Fig. 1.3 (*a*) illustrates the Horizontal System of the coordinates for astronomical objects. In it, SZNZ is the celestial sphere while point L is the intersection of the great circle through X and the horizon. The azimuth A is to be measured from the North point N. Fig. 1.3 (*b*) illustrates the equatorial system of coordinates, where Y is the first point of Aries. In it, PXQ is a great circle through X. The horizon system and the spherical triangle ZXP are shown. This gives the relation between the Equatorial and Horizontal systems of coordinates. Fig. 1.3 (*c*) explains the Ecliptic System of coordinates. The spherical triangle MXP establishes the relation between the Ecliptic and the Equatorial systems of coordinates. Fig. 1.3 (*d*), on the other hand, shows the Galactic system of coordinates. In it, point A indicates the direction of the center of the Galaxy.

1.11 DISTANCE MEASURING UNITS IN ASTRONOMY

Different units are used for measuring the distances between astronomical objects depending on their positions. For measurements within our solar system, or other solar systems, the common unit used is the *astronomical unit*. 1 astronomical unit = the average distance between the Sun and the Earth (≈ 150,000,000 km). This is very small when measurements are to be made for galactic objects like the stars where another unit called *parsec* (pc) is used. 1 pc = 3.26156378 light years (≈ 206265 AU), which is nearly equal to the distance of the nearest star. The term parsec is the short form of *parallax of one arc-second*. Note that if an object has a parallax of 0.5 arc-second, it will be twice the distance of an object with 1 arc-second of parallax, or about 2 parsecs away (6.52 light years). Conversely, an object with 2 arc-seconds of parallax will be twice as close as an object with 1 arc-second of parallax, or about half a parsec away (1.63 light years). For measuring distances within our galaxy or beyond (other galaxies) kilo-parsec (kpc) is used. 1 kpc = 1000 pc. For distances between galaxies and the cosmology Mega-parsecs (Mpc) are used. 1 Mpc = 1,000,000 pc. It may be noted that though a light-year is a large distance, it is generally not used in present day astronomy.

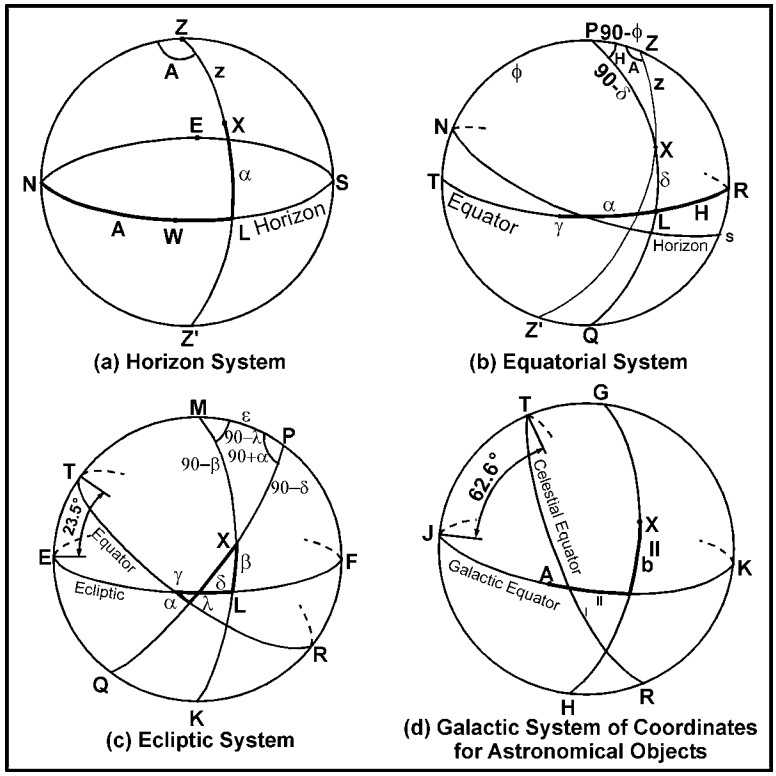


Figure 1.3: The various coordinate systems. (a) Horizon System, (b) Equatorial System, (c) Ecliptic System, and (d) Galactic System of coordinates for astronomical objects.

EXERCISES

1. What was the explanation of Copernicus for the complicated motion of the planets in terms of the stationary Sun? How did Galileo and Kepler adopt the Copernicus system?
2. How did Newton establish the laws of planetary motion in a Sun-centered system?
3. How many stars are there in our Galaxy and in the observable universe?
4. What is Big Bang?
5. Explain the standard model of the Big Bang theory in eight different steps.
6. Who first discovered the cosmic microwave background?
7. Provide in a tabular form: very rough sizes, masses and ages of the Sun, the Galaxy and the observable universe. You may compare the quantities with respect to a small child.
8. Which are the oldest stars in our Galaxy?

9. What is meant by interstellar medium? What objects are called galaxies?
10. What is the Milky Way system?
11. What is the basic difference between 'proper motion' and 'Doppler shift' in astronomical measurements?
12. What is radio astronomy? What are the remarkable objects and phenomenon it has successfully attempted to discover?
13. Who first discovered the radio noise coming from our Milky Way Galaxy?
14. What is the difference between radio waves and light waves?
15. What are the main sources of cosmic radio waves?
16. What are black holes?
17. Draw the spectral lines of the electromagnetic spectrum and discuss.
18. Mention the two fundamental characteristics of a radio telescope.
19. What is the utility process of 'aperture synthesis'?
20. Commet on the usefulness of the technique of very long baseline interferometry?
21. What is the importance of the centimeter and decimeter parts of the electromagnetic spectrum?
22. Explain clearly with a suitable diagram the systems of coordinates used to locate the stars and other objects in the sky (from Earth).
23. What are the essential requirements of an astronomical coordinate system?
24. Give the names of four astronomical systems of coordinates depending on the fundamental circle and reference point. Illustrate with a suitable diagram.
25. Mention the different units used for measuring the distances between astronomical objects (based on their positions).
26. What is 1 Astronomical unit? What is pc?

Chapter 2

THE SOLAR SYSTEM

2.1 INTRODUCTION

As we approach the solar system the first bodies we encounter are frozen balls of gas and dust that exist in the billions. At least one of those frozen balls has wandered too close to the Sun. It has taken the familiar form of a comet. Pluto[6] is the most distant planet from the Sun, but its eccentric orbit has actually carried it inside the orbit of Neptune. Also, inside Neptune's orbit lies Uranus, the seventh planet of the solar system. Pluto, Neptune, and Uranus are all poorly known.

The sixth planet, Saturn, is bright enough to be visible. The next planet, Jupiter, is more massive. Its face is covered by an intricate system of zones and belts. Jupiter is surrounded by a retinue of satellites including 15 known moons.

Inside the orbit of Jupiter lies a vast wasteland, the asteroids. Closer to the Sun lies the fourth planet, Mars. The third planet is the Earth, the second planet Venus, and closest to the Sun is Mercury.

[6] Currently, Pluto is under controversy in the astrophysical world. It is not accepted as a planet anymore, though we have given its old heritage in various places in this book.

At the center of the solar system lies the mighty Sun. Thus, as we swing around the Sun, we pass Mercury and Venus, and approach the third planet, the Earth, which is nothing but an insignificant speck of dust in this immense Universe.

2.2 PLANETS AND THEIR SATELLITES

Table 2.1 gives the properties of the nine planets that circle around the Sun. It shows that the planets divide into two categories: (*i*) the inner or terrestrial planets, which are small and have a mean density of 4 to 5 gm/cm^3; and (*ii*) the outer or Jovian planets, which are large (except Pluto) and have a mean density of 1 to 2 gm/cm^3.

Each of the Jovian planets, i.e., Jupiter, Saturn, Uranus, and Neptune, has more than one moon. In fact, Jupiter has at least 15 moons. Saturn has at least 22 moons. Uranus has five known moons; and Neptune, two. In addition to satellite systems, the Jovian planets may also all possess rings (e.g., the rings of Uranus are shown in Fig. 2.1).

Table 2.1: Properties of the planets revolving around the Sun.

Planet	Equatorial diameter (Earth = 1)	Gravitational mass (Earth = 1)	Mean density (gm/cm^3)	Mean distance from Sun (Earth = 1)	Sidereal rotation period (day)
Mercury	0.38	0.055	5.4	0.39	58
Venus	0.95	0.81	5.1	0.72	-243
Earth	1.0	1.0	5.5	1.0	1.0
Mars	0.53	0.11	3.9	105	1.0
Jupiter	11.0	318	1.3	5.2	0.41
Saturn	9.4	95.0	0.70	9.5	0.43
Uranus	4.4	15.0	1.0	19	-0.51
Neptune	3.9	17.0	1.7	30	0.66

Celestial mechanics experiments plus ultraviolet photometry by Pioneer-II revealed that the total mass of Saturn's rings cannot exceed 3 millionths of the mass of the planet. This mass consists mostly of small icy specks. Analysis of radio waves transmitted through the rings during the Voyager-1 flight past Saturn revealed many boulders with diameters of several meters. In contrast, the rings of Jupiter and Uranus are made of much smaller and less reflective particulate material.

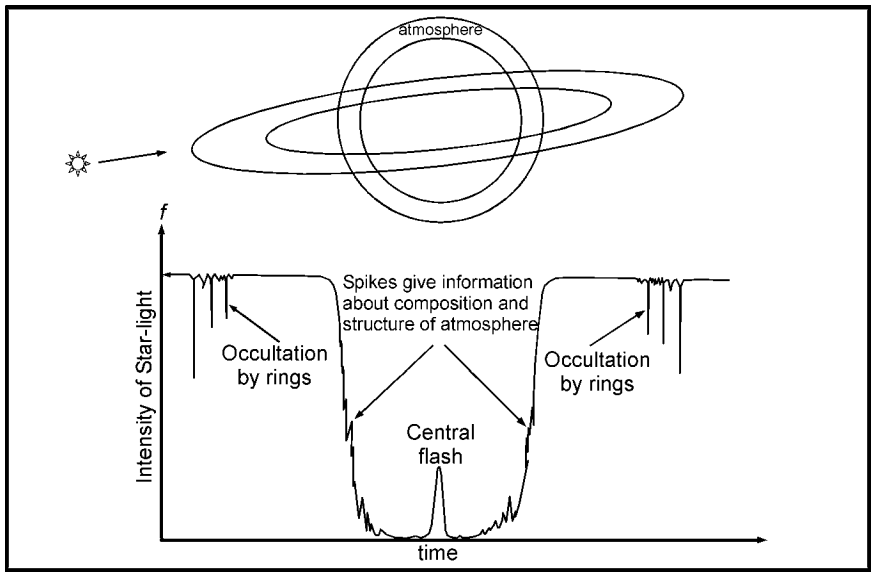


Figure 2.1: The rings of Uranus were revealed by studies of the occultation of a distant star. The lower light curve shows schematically the observed light intensity of the star as its position relative to a ringed planet changes. Such light curves can reveal not only the presence of occulting ring material but also much information about the composition and structure of the planets atmospheric layers. A central flash occurs when the star is directly behind the center of the planet and is caused by refraction and extinction in the planets atmosphere acting like a lens in this configuration.

The terrestrial planets, i.e., Mercury, Venus, Earth, and Mars, have no rings and very few moons. Mercury and Venus have no moon, Earth has one (Luna or the Moon), and Mars has two (Phobos and Deimos). Furthermore, in comparison with the large moons of the Jovian planets, the moons of the terrestrial planets, except for Luna, are quite small (Table 2.2). As the two small moons of Mars are captured asteroids, the Earth-Moon system is unique among the terrestrial planets.

Table 2.2: Largest moons of terrestrial and Jovian planets.

Planet	Largest satellites	Mean diameter (Earth = 1)	Gravitational mass (Earth = 1)	Mean density (gm/cm^3)	Distance from planet (1t-sec)
Mercury	none	-	-	-	-
Venus	none	-	-	-	-
Earth	Moon	0.27	0.012	3.3	1.3
Mars	Phobos	0.0016	1.7×10^{-9}	1.9$^{\oplus}$	0.03
Jupiter	Ganymede	0.41	0.025	1.9	3.3
Saturn	Titan	0.40$^{\otimes}$	0.023	1.9	4.1
Uranus	Titania	0.14	0.0007	1.4	1.5
Neptune	Triton	0.40$^{\otimes}$	0.023	1.9	1.2

2.3 MINOR PLANETS OR ASTEROIDS; METEOROIDS, METEORS, AND METEORITES

In addition to the nine major planets and their satellites, there are between 10^4 and 10^6 minor planets or asteroids. Their orbits lie between Mars and Jupiter at distances of 2.2 to 3.3 astronomical units from the Sun. There may also be minor planets like Chiron whose orbit lies mostly between Saturn and Uranus. The total mass of all the asteroids is less than 10^{-3} of the mass of Earth. Though most of the asteroids lie in a *belt* between Mars and Jupiter, there are meteoroids which have Earth-crossing orbits. Fig. 2.2 shows schematically the theory about the production of meteoroids from asteroids. The idea is that asteroids collide with one another, which can shatter the bodies, and some of the fragments may be thrown into resonant interactions with Jupiter that ultimately give them Earth-crossing orbits. A meteoroid, which intersects the Earth and enters the Earth's atmosphere will

---

$^{\oplus}$ A gravitational mass estimate for Phobos is available from Viking Orbiter and suggests that Phobos has chondritic material.

$^{\otimes}$ Voyager-1 measurements showed that Titan is slightly smaller than Ganymede contrary to previous belief. It remains for spacecrafts to settle whether Triton or Ganymede is the solar system's biggest moon.

heat up from the friction generated in the passage. It will then appear as a fiery *shooting star* called a meteor. If the mass of the meteor is less than 10^{-10} gm, it may slow down so fast that it survives the flight; if so, it is called a micrometeorite. On the other hand, if the mass of the meteor is greater than 10^{3} gm, the object is called a meteorite. Meteorites come in three types depending on their chemical composition: *stones*, *stony irons*, and *irons*. Stones resemble rocks; stony irons have some metal-rich inclusions, while irons contain mostly metals like iron and nickel.

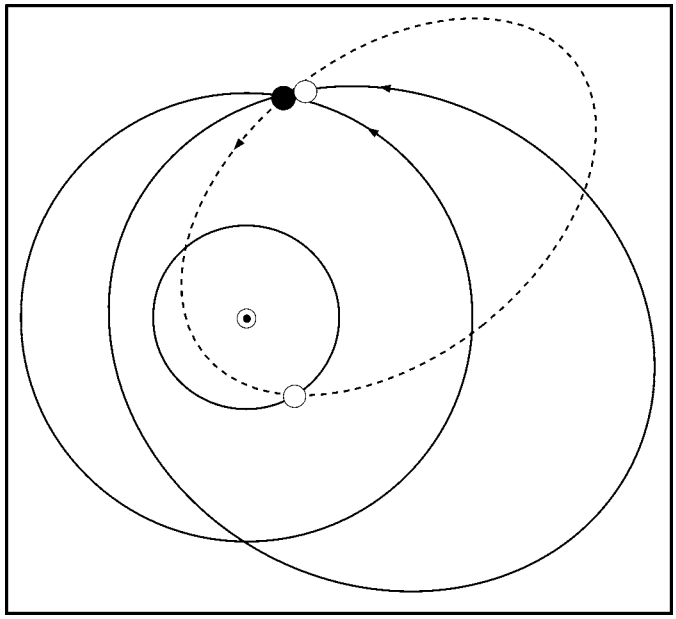


Figure 2.2: The collision of two asteroids shatters them into many pieces. Some of these may suffer resonant interactions with Jupiter and within 10^{4} to 10^{5} years, enter Earth-crossing meteoroids.

2.4 COMETS

The word comet is derived from the Latin word *cometa*, which means *long-haired*. Comets are astronomy's most spectacular displays. Fig. 2.3 shows schematically some features found in comets; a head consisting of two parts: a nucleus and a coma. In addition it has one or two tails. The nucleus of the comet is the essential part, as it is the ultimate source of all the mass and is generally less than 50 km across. The coma could be larger than the Sun and the ion tails have been observed to extend 150 million km (1 astronomical unit) or more. Whipple proposed that the

nucleus is composed of chunks of dust and frozen ices of compounds like methane (CH_4), ammonia (NH_3), water (H_2O), and carbon dioxide (CO_2). Comets are often called *dirty snow balls* based on their composition.

Before our solar system came into existence, the interstellar cloud out of which the comets were formed had uncountable trillions of these bodies mixed in with a cloud of various gases of carbon monoxide, hydrogen, formaldehyde, ammonia, and so on. The process by which this diffused gas clumped into comets still unknown. These bodies probably measured from centimeters to tens of kilometers. Today, all the remnants of the original interstellar cloud around the solar system are these comets. Sometimes comets visit the inner solar system while moving in their orbit. We shall continue to see these bodies even after most of the asteroids have vanished. These objects can also be a threat to our existence as they could collide with the Earth like the comet Shoemaker-Levy, which collided with Jupiter.

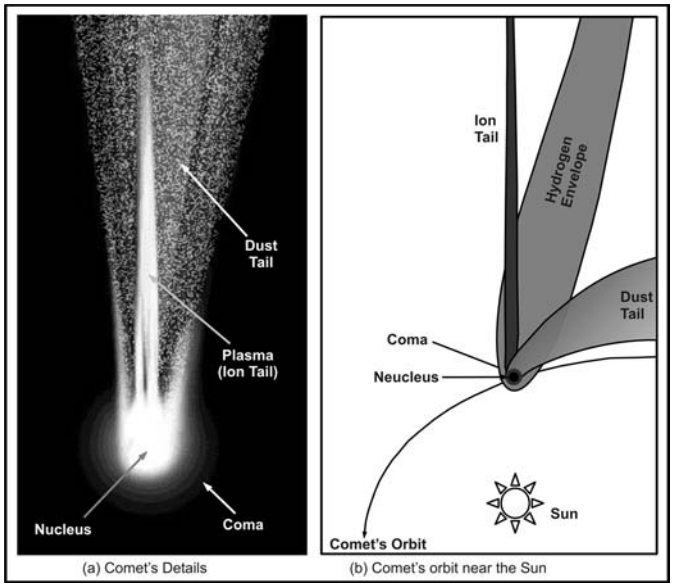


Figure 2.3: The structure of a comet. (a) Details of a comet. (b) The orbit of a comet near the Sun.

The orbits of comets are constantly changing since their origin is outside the solar system. Some comets move into Sun-grazing orbits and get destroyed when they near the Sun, while others get thrown out of the solar system permanently. Comets are identified from their periods, i.e., short-period comets or long-period comets. The former have orbits of less than 200 years, while the latter have longer orbits (thousands of years) but remain gravitationally bound to the Sun. These are

also called periodic comets since their orbits are elliptical. Due to their low masses and elliptical orbit, the comets are pulled close to the giant planets and their orbits are often perturbed. Non-periodic or single apparition comets have parabolic or hyperbolic orbits, which make them permanently exit the solar system after a single pass by the Sun.

As the comet approaches near the Sun, both the coma and tail gets illuminated as they reflect sunlight and could be visible from the Earth. In general, most comets are extremely faint and can be seen only through telescopes. It is surprising that comet nuclei are among the darkest objects known to exist in the solar system since they reflect only 4 percent of the light falling on them. In 1996 it was also found that comets emit x-rays. These are thought to be generated by the interaction between comets and the solar wind when highly charged ions penetrate the comet's atmosphere and collide with the atoms and molecules. The collision process results in the capture of one or more electrons by the ions leading to emission of x-rays and far ultraviolet photons.

Some well-known comets are Comet Borelly, Comet Enche, Comet Hale-Bopp, Comet Halley (also known as Halley's comet), Comet Hymason, Comet Ikeya-Seki, Comet Mrkos, and Comet Shoemaker-Levy 9.

EXERCISES

1. Where is the position of the mighty Sun in the solar system?
2. Present in a tabular form all the planets that circle the Sun. Hence from that table show how the planets are divided into two categories.
3. What are the numbers of moons of the different Jovian planets?
4. Show the rings of Uranus (as revealed by studies of occultation of a different star) and discuss.
5. Justify the statement - "The Earth-Moon system is unique among the terrestrial planets."
6. Compare the equatorial diameter, gravitational mass, mean density, mean distance from the Sun, and sidereal rotation period of the planets Mercury, Venus, and Earth.
7. Give the names of the largest moons of the terrestrial and Jovian planets.
8. Discuss asteroids.
9. Discuss the theory of meteoroids from asteroids.
10. What are the three different types of meteorites (depending on their chemical composition)?
11. What is the meaning of the word comet as derived from the Latin word 'cometa'?
12. Point out some features found in comets. Why are comets termed as 'dirty snow-balls'?
13. Draw the orbit of a comet near the Sun and discuss in details.

14. What happens when a comet approaches the Sun?
15. What is the explanation for the emission of x-rays from comets?
16. Give names of some well-known comets.

Chapter 3

THE ATMOSPHERES OF THE PLANETS

3.1 INTRODUCTION

We know that the atmosphere of a terrestrial planet is the mass of gas that lies above the surface of the planet. But a Jovian planet does not have a solid surface. Also, a star like the Sun has no solid surface. But we have no difficulty distinguishing between its interior and its atmosphere. By the atmosphere of an astronomical body we mean the layer where the photons make a transition from *walking* to *flying*. This definition is fairly unambiguous for a star as most of the ambient photons have similar wavelengths. For a planet, sunlight provides a source of incoming radiation, which has photons in the visual and near-ultraviolet regime. This radiation field affects the atmosphere qualitatively. Radio, infrared, optical, and ultraviolet wavelengths also affect the atmosphere of a planet.

3.2 THERMAL STRUCTURE OF TERRESTRIAL PLANETS

Mars has only a very thin atmosphere while Mercury and the Moon have no atmosphere to speak of. The temperature structure of the Martian atmosphere is

affected by the direct absorption of sunlight by dust particles swept into the air (Fig. 3.1). The terrestrial planets with gaseous atmospheres are Venus and Earth.

Fig. 3.2 shows a schematic plot of the temperature of the Earth's atmosphere as a function of height above the ground. The thermal structure of the Earth's atmosphere has four distinct regions:

(*i*) The troposphere is the region that adjoins the ground. It is the region with which we associate weather. Vertical heat transfer in the layer consists mostly of the upward diffusion of infrared photons and fluid transport by convection. The temperature is highest near the ground and steadily decreases to the tropopause, the top of the troposphere.

(*ii*) The stratosphere is heated by the ultraviolet component of sunlight in forming and destroying the ozone O_3. This direct absorption leads to an inversion of the temperature above the tropopause, so that the stratosphere is actually warmer than the troposphere. This increase in temperature with height continues until the stratopause is reached.

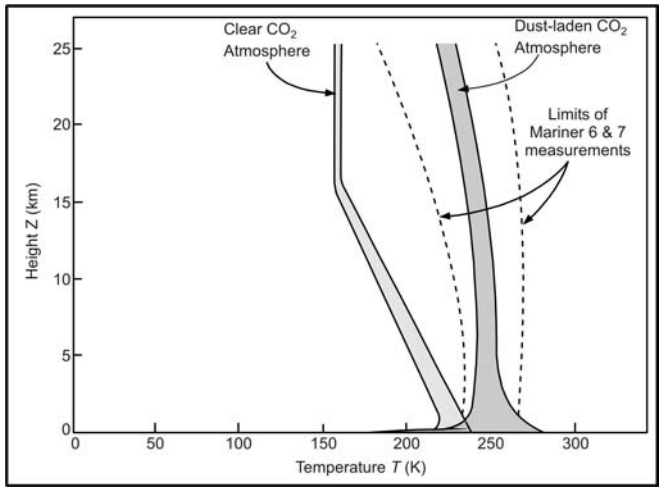


Figure 3.1: Thermal structure of the atmosphere of Mars is affected by the amount of dust swept into the air by the winds of that planet. The shaded region between the curves demarcates the theoretical range of temperatures observed in calculations using typical daytime conditions for a dust-laden and clear Martian atmosphere. The dashed curves indicate the limits found by aircraft observations of Mars.

(*iii*) In the mesosphere, there is a decrease in ozone production and an increased rate of infrared cooling by carbon dioxide (CO_2). The temperature decreases with increasing height up to the mesopause.

(*iv*) In the thermosphere, heat input from sunlight occurs by the photo-dissociation and photo-ionization of molecular oxygen (O_2). The balance

of this heat input leads to a rise of the temperature with increasing height in the thermosphere until temperatures of about 1,000 K are reached. The primary energy transfer in the lower thermosphere is a downward conduction of heat into the mesosphere.

As the atmosphere is transparent to optical photons, sunlight strikes the ground, with little interaction with the troposphere except for scattering by molecules, dust, and reflection from cloud tops. The troposphere can be regarded as a gaseous atmosphere. It is heated from below by the emanation of infrared radiation from the ground. The troposphere is partially opaque to this emergent infrared radiation. In planetary astronomy, this is known as the *greenhouse effect*, but the basic principle is the same as that which governs the radiative transfer inside a star.

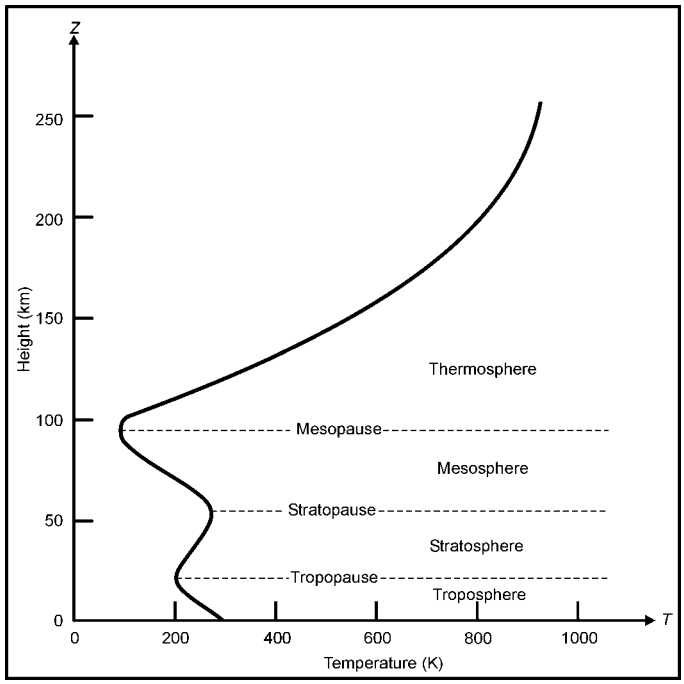


Figure 3.2: The thermal structure of the Earth's atmosphere. The various regions are presumably duplicated qualitatively in the other planets. The detailed mechanisms of ultraviolet heating in the upper atmospheres of the other planets may, however, differ from those of the Earth due to the different chemical compositions involved.

As the Sun does not shine equally on the Earth's equator and poles, we expect some variation of this temperature with latitude. Measurements show that the tropopause at the tropics and polar caps have different temperatures. The tropopause is warmer at the poles (about 225 K) than at the tropics (about 195 K).

This situation is opposite to that which prevails at sea level. As an explanation Victor Starr proposed that the refrigeration of the tropopause at the tropics results from the planetary circulation pattern of the Earth's atmosphere.

The upper levels of the troposphere are convectively stable while the lower levels are not. As the radiative equilibrium solutions require the temperature in the lower troposphere to drop quickly with increasing height, these solutions may become unstable and allow convection to develop. The purely radiative solutions actually require the ground heated by both the trapped infrared radiation and the incident optical sunlights to be warmer than the air just above it (Fig. 3.3). This situation produces convection currents at least at ground level. The humidity depends, in part, on the weather patterns that develop as a result of the convection and the large-scale circulation pattern. The combination of all the complications on the Earth produces an average surface temperature of about 290 K, close to the freezing point of water.

Let us now enquire what the surface temperature is on Earth's sister planet, Venus. Venus is enshrouded with clouds, so that its surface cannot be examined by ground-based telescopes in either the optical or the infrared. Radio measurements can reach the surface. Such measurements indicate that the surface temperature of Venus might be in the neighborhood of 750 K, a temperature high enough to melt lead.

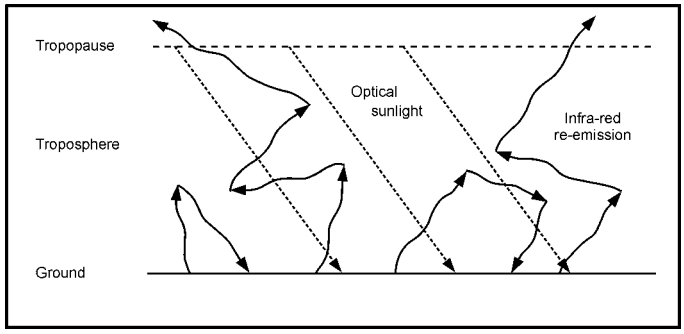


Figure 3.3: The greenhouse effect in planetary atmospheres involves the additional warming of the surface layers due to the partial trapping of the remitted infra-red radiation. Once it was assumed that the name for this effect might be a misnomer as greenhouses may not be kept warm by the same mechanism. However, a reanalysis by Silverstein reveals that the greenhouse effect may be nontrivial even for greenhouses. In any case, the effect is the same as that which keeps the interiors of stars warmer in their interiors than at their surfaces.

Venus is 1.38 times closer to the Sun, but it reflects about 72 percent of the incident sunlight while the Earth reflects 39 percent. Thus, the average flux of sunlight absorbed by Venus is $(1.38)^2(0.28/0.61) = 0.87$ of that absorbed by Earth (i.e., the effective temperature of Venus should only be $(0.87)^{1/4}$ that of Earth,

or about 238 K). Venus has whopping greenhouse effects and has a whopping thick atmosphere. Surface pressure is about 90 times that at sea level on Earth. This atmosphere is composed of 96 percent of carbon dioxide (CO_2), which is an efficient absorber in the infra-red. Venus has a much thicker atmosphere than the Earth with a composition heavily weighted toward CO_2.

3.3 RUNAWAY GREENHOUSE EFFECT

The crucial factor to be accounted in this connection is the existence of liquid water on the surface of Earth and not on Venus. Sedimentary rocks have a capacity to absorb carbon dioxide if there is liquid water present. The Earth has an estimated equivalent of 70 atmospheres of CO_2 locked up in rocks like lime-stone ($CaCO_3$). The corresponding amount of CO_2 in Venus is actually in the atmosphere.

Simpson showed that as long as there is liquid water on the surface of a planet, the top of the convection zone in the troposphere will hold as much water vapor as it physically can. This is because more water can be held in warm air than in cold air. The layers above the top of the convection zone are convectively stable. So there is no descent of dry air into the top of such a convection zone. This means that the air at the top tends to be exactly saturated, with the excess water going into the formation of clouds of small droplets of liquid water or particles of ice.

Water vapor has a saturation temperature T, which is a function of the partial pressure P_{H_2O}. This can be expressed as,

$$T = T_S(P_{H_2O}). \tag{3.1}$$

The temperature above the top of the convection zone, on the contrary, follows the radiative equilibrium law

$$T = T_e\left[\frac{3}{4}\left(\tau + \frac{2}{3}\right)\right]^{1/4}. \tag{3.2}$$

Here, T_e is the planetary effective temperature. If the radiative equilibrium keeps the top of the convection zone too warm to allow water vapor to saturate at the top of the convection zone, more water will be drawn from the surface by the ascending convection currents. The dwindling supply of liquid water to form sedimentary rocks would release more CO_2 into the atmosphere. The increase of both CO_2 and H_2O concentrations in the atmosphere would increase the greenhouse effect. The runaway would not be halted until all the available surface water is released into the atmosphere and the surface of the planet is left in an arid and sweltering state.

Scientists have speculated that the opposite of a runaway greenhouse effect may have occurred on Mars. Spectroscopic and photometric studies suggest a

surface pressure of about 0.6 to 0.7 percent that of the Earth. A surface temperature varies greatly with latitude and between day and night, but it never exceeds 300 K and may fall below 150 K (i.e., the freezing point of carbon dioxide when it forms the *dry ice* at the surface pressure of Mars). Mars' atmosphere consists of 95 percent carbon dioxide (CO_2), 2.7 percent nitrogen molecules (N_2), 1.6 percent argon (Ar), and small amounts of free oxygen (O_2) and water vapor (H_2O). The average surface pressure and temperature on Mars are too low to allow the presence of liquid water. However, it is likely that Mars once had a much denser atmosphere, and much higher surface temperatures. Most of the water and carbon dioxide now seems to be locked up either in the polar caps or in carbonate rocks.

An earlier and denser atmosphere on Mars did allow the presence of liquid water on the planet. However, the gradual locking up of the atmospheric carbon dioxide into sedimentary rocks led to a thinning of the atmosphere, with a consequent lowering of the temperature. This caused more water vapor to precipitate out of the atmosphere resulting in even more carbon dioxide being locked up.

Fig. 3.4 shows a schematic diagram of the greenhouse effect for Venus, Earth, and Mars. The solid curves represent the surface temperature of the planet, assuming an opacity associated with various values of the vapor pressure of water at ground level. The dashed curves show the loci of thermodynamic equilibrium possible for the three phases of water. Water opacity cannot raise the surface temperature on Mars higher than a point where its vapor becomes saturated with respect to ice, and on Earth, with respect to liquid water. For Venus, a runaway greenhouse effect occurs because the solid curve is incompatible with the existence of liquid water on the surface of the planet.

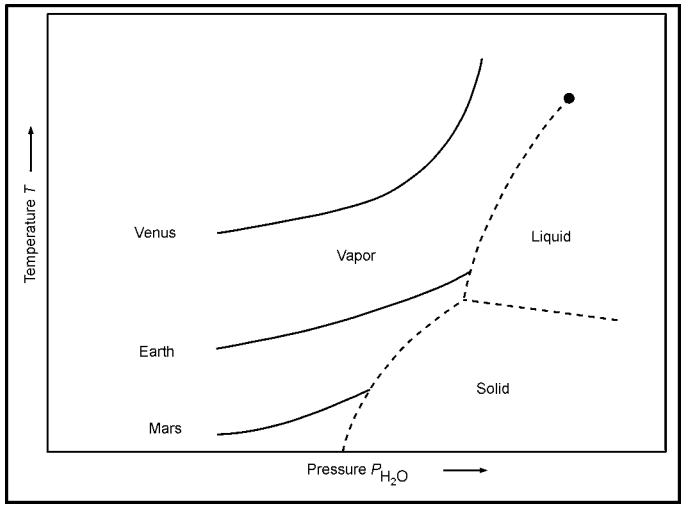


Figure 3.4: A schematic diagram of the greenhouse effect for Venus, Earth, and Mars.

Venus' atmosphere is 96 percent CO_2, about 3.5 percent of the remainder is nitrogen molecules, N_2, and there is only a trace amount of H_2O.

3.4 THE EXOSPHERE

Molecules cannot escape from the lower levels of a planetary atmosphere any more than a photon in the deep interior of the Sun can fly out in a straight line (Fig. 3.5). A fast molecule would collide with another molecule before it had made any substantial progress toward escape. The path followed by any molecule in these lower regions is a tortuous *random walk*. Only in the rarefied outer portions of a planetary atmosphere can a molecule fly freely and have a chance to escape the gravitational clutch of the planet. It is in these regions where a newly formed hydrogen atom can escape before recombining with another hydrogen atom. The region where fast molecules change from *walking* to *flying* is known as the exosphere. Exospheres are to molecules on planets what photospheres are to photons on stars. However, all photons can escape, as they travel at the speed of light. With molecules, only that fraction in the exosphere can actually escape whose velocity exceeds the escape speed the planet.

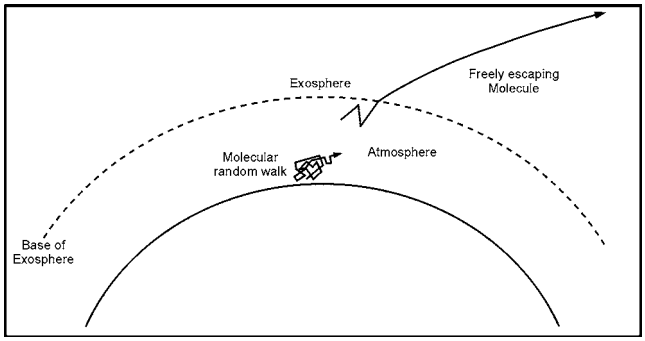


Figure 3.5: Deep in a planetary atmosphere, the mean free path for molecule collisions is so small that even a very fast molecule has no chance to escape from the planet. Above the base of the exosphere, the density of molecules has dropped low enough that the probability for escape becomes high for fast molecules.

We consider the case where an exosphere is located at a radius r_e above the center of a spherical planet of mass M and where the exosphere is at a uniform temperature T, containing only one molecular species of mass m. Let the collisional cross-sectional area of molecules m with each other be A. The rate of loss of the total number of molecules m may be written as

$$N_m = (2\pi)^{1/2} \frac{2}{A} r_e v_T f(x_e) . \tag{3.3}$$

Here, $v_T = \left(\frac{kT}{m}\right)^{1/2}$ is the thermal speed. The dimensionless function $f(x_e)$ is expressed as

$$f(x_e) = x_e(1 + x_e) \exp(-x_e) \text{ with } x_e = \frac{GMm}{r_e kT}. \tag{3.4}$$

The function $f(x_e)$ is plotted in Fig. 3.6 for the relevant regime x_e appreciably greater than unity. In this regime, $f(x_e)$ is a steeply declining function of increasing x_e. With everything being equal, the Jeans escape mechanisms drastically favor the escape of light molecules, since v_T and $f(x_e)$ are both larger for smaller values of m in Eq. (3.4). This gives the number of molecules lost per second. Hydrogen atoms are particularly susceptible to evaporation from a planet.

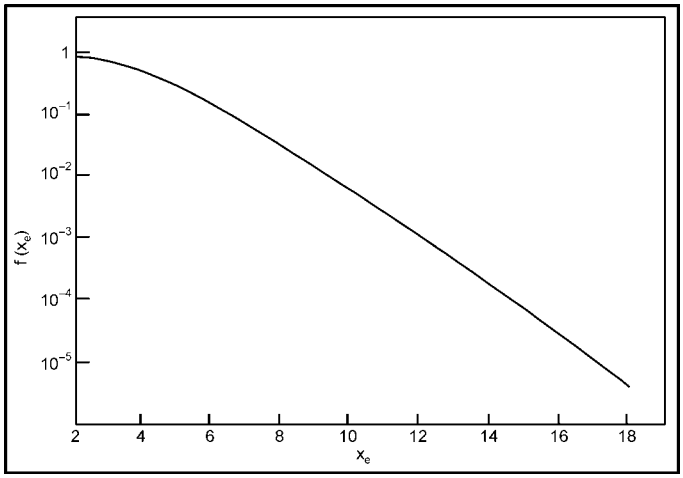


Figure 3.6: The function $f(x_e)$ is a rapidly declining function of its argument for x_e considerably greater than unity.

The substitution of realistic numbers into Eq. (3.4) reveals that the Jeans mechanism is inadequate and to see this, we consider the Earth. Let us adopt for T the temperature 4,000 K, which results in $r_e = 7.0 \times 10^8$ cm (about 600 km above sea level). The collisional cross-section of hydrogen atoms by hydrogen atoms is about 10^{15} cm^2. To take into account collisions with other molecules, we adopt a collisional cross-section per hydrogen atom equal to $A = 6 \times 10^{-14}$ cm^2. With these values we obtain from Eq. (3.3) a total loss rate of hydrogen atoms by Jeans escape from the Earth of $N_m = 1 \times 10^{27}$ sec^{-1}.

This value is comparable to observations of Lyman alpha emission from the Earth's upper atmosphere. During the 4.6 billion years of its existence, the Earth would have lost at the above rate, about 2×10^{20} gm of hydrogen, roughly

three orders of magnitude less hydrogen than is contained in the form of H_2O in the oceans. The hydrogen loss rate measured for Venus is comparable to the above rate. The Jeans mechanism thus seems incapable of giving a much larger rate in the past. If Venus ever had as much liquid water as the Earth, it probably could have lost only 10^{-3} of it by the Jeans mechanism. Hunten has shown that any increase of the exospheric evaporation rate ultimately is limited by the ability of the lower layers to re-supply the lost particles by diffusion. From the regions where the material is well-mixed, the hydrogen or hydrogen-carrying compounds has a limited capacity to rise diffusively relative to the other constituents in the atmosphere.

EXERCISES

1. What is meant by the atmosphere of an astronomical body?
2. Draw the structure of the atmosphere of Mars and discuss.
3. Show a schematic plot of the temperature of the Earth's atmosphere as a function of height above the ground. Hence explain the thermal structure of the Earth's atmosphere considering four distant regions.
4. Explain the greenhouse effect in planetary atmospheres.
5. Explain why the structure of Venus cannot be explained by optical or infra-red telescope?
6. Explain what is meant by a runaway greenhouse effect with special emphasis on Simpson's observation.
7. How is the saturation temperature of water vapor related to the partial pressure?
8. Write down the mathematical form of radiative equilibrium law.
9. "Opposite of a runaway greenhouse effect may have occurred on Mars." How is it viewed by scientists?
10. Draw a schematic diagram of the greenhouse effect for Venus, Earth and Mars and then explain.
11. How is the region exosphere defined from consideration of fast molecular change? Write down an expression for the rate of loss of the total number of molecules in terms of thermal speed and radians.
12. Discuss the interesting features of the exosphere, pointing out the relevant mathematical relations.
13. Use the standard numerical values and verify the values quoted for 'v_m'. Also find the total mass of hydrogen lost in 4.6×10^9 year.

Chapter 4

Origin of the Sun, the Earth, and the Solar System

4.1 INTRODUCTION

The issue of the origin of the solar system is not the same as that of the origin of the Universe. The Universe is sufficiently larger than the solar system. For the issue of origins, geology supports that the solar system is appreciably younger than the oldest objects in the Universe. The solar system thus has formed against the backdrop of an already mature Universe. There are two popular theories about formation of the solar system. The first is popularly known as the *solar nebula theory*, originally proposed by Immanuel Kant during 1775 and later in 1796 by Pierre-Simon Laplace. The second was originally proposed by Buffon, taken up by Chamberlian and Moulton and developed by Jeans and Jeffreys and by 1930 it became the theory of the origin of the solar system. This is known as the *planetsimal theory*. However, these theories do not give proper evidence about the age of the solar system or Earth. More reliable methods are to find the ages of the solar system rocks by the study of the long-lived radioactive elements which they contain.

4.1.1 The Solar Nebula Theory

The interstellar gas and dust particles are assumed to be the ingredient matter of the solar system. The cloud is disturbed and starts collapsing under its own gravity and in the process it starts rotating and gets flattened like a disk. The disturbance could be the shock waves from a nearby exploding star. The compression causes heat to build up for the dust to vaporize. This initial collapse is estimated to be less than 100,000 years. The center compresses enough to become a proto-star while the rest of the gas encircles around it. Gradually, most of the gas flows inward thereby adding mass to the forming star, but they also keep rotating. The centrifugal force prevents some amount of gases from reaching the forming star and in turn forms an *accretion disk* around it. The disk radiates and cools down. The orbiting gas around the proto-star might be unstable and start to compress under its own gravity, which may produce a double star. Otherwise, the gas cools down enough to condense into rocks and tiny particles of dust. The metals condense around the time of formation of the accretion disk,[7] while the rocks condense a bit later.[8] The dust particles keep colliding within themselves and form larger particles until they take the size of boulders or small asteroids. The larger particles pull the smaller ones by their gravity and grow further until a solid matter is formed close to their own orbit. The final sizes depend on their distance from the star and the density of the proto-planetary nebula. It takes about 100 million years to reach the stage of the solar system that we live in today.

4.1.2 The Planetsimal Theory

This theory assumes that the Sun was already formed without the planets and was developed by T. C. Chamberlin and F. R. Moulton in the early part of the twentieth century. It started by describing an encounter or collision of a star passing by the Sun, but finally refined to a theory where the star grazed passed the Sun. This resulted in huge tidal waves on the surface of the Sun. Some of this erupted matter was detached by the cross-pull from the star, which was thrust into elliptical orbits around the Sun. The smaller masses cooled faster and solidified, which are called *planetsimals*. While traveling within the orbits, the higher mass objects picked up the smaller ones by gravitational pull while they crossed closely. These masses grew and formed planets. Later, during 1918, James Jeans and Harold Jeffreys suggested that a huge tidal wave raised on the Sun by a passing star was drawn into long filament and got detached from its principal mass. This stream of gaseous matter condensed gradually and also separated itself into masses of various sizes, which by further condensation took the shape of planets. By 1930s it became the

[7] Between 4.55 to 4.56 billion years ago based upon the isotope measurements of certain meteors

[8] Between 4.4 to 4.55 billion years ago.

theory of the origin of the solar system, although serious objections against the theory still remain today.

4.2 AGE OF THE EARTH

The geologists of the eighteenth and nineteenth centuries showed how the sequence of rock strata and fossils could be arranged in a qualitative and semi-quantitative sequence. However, absolute estimates of ages were stymied by the uncertain and variable rates of weathering, volcanism, erosion, and sedimentation. Estimates of the age of the Earth from theoretical considerations of its rate of cooling suffered from lack of knowledge about the composition of the Earth's interior, the origin of its interior heat, and the role of solar heating for keeping the surface warm. By the nineteenth century, it was clear that if the rate of geological processes in the past was not different from that in the present, then the Earth must be much older than roughly 6,000 years. Geologists insisted that the Earth had to be even older than the 20 or 30 million years that Kelvin and Helmholtz deduced from their contraction theory for the power of the Sun. Rutherford applied the definitive method for determining the age of the Earth by radioactive dating.

4.2.1 Radioactive Dating

Radioactive dating is a sophisticated experimental technique. We assume that we have a rock which contains ^{238}U. Since ^{238}U decays ultimately into ^{206}Pb, the amount of ^{238}U must have been larger in the past, and the amount of ^{206}Pb must have been smaller. However, the sum of the number of uranium-238 and lead-206 atoms must be a constant:

$$N(^{238}\mathrm{U}) + N(^{206}\mathrm{Pb}) = \text{constant}. \tag{4.1}$$

We may get the value of the constant by measuring the amounts of ^{238}U and ^{206}Pb currently present in the ore. If we assume that all the lead-206 comes from the radioactive decay of uranium-238, we can determine the time in the past when $N(^{206}\mathrm{Pb})$ was zero (Fig. 4.1). This time is the age of the ore when the rock solidified.

If we cannot assume that all the lead-206 came from the radioactive decay of uranium-238, a slightly more complicated line of reasoning is required. We then deduce how much of the lead-206 was not the product of the radioactive decay of uranium-238. To do this we have to examine the amount of lead-204 (^{204}Pb). It is another stable isotope that could be present in the radioactive ore, but it is not the product of a radioactive decay. Lead-204 and lead-206 have identical chemical properties. So the radioactive ore will incorporate the monradioactively derived

portions of these two isotopes in the same proportion as in any non-radioactive rock. This proportion can be measured in non-radioactive rocks, and the primitive ratio can then be used to infer the amount of ^{206}Pb, which does not result from the radioactive decay of ^{238}U. We multiply the primitive ratio [^{206}Pb]/[^{204}Pb] by the amount of ^{204}Pb now in the radioactive ore. Fig. 4.2 shows the correction procedure that subtracts out the primitive ^{204}Pb. In fact, Fig. 4.1 is a special case of Fig. 4.2 if $N(^{204}\text{Pb})$ is zero (i.e., if the ore contained no intrinsic lead at the time of its formation). In practice, the procedure has some difficulties because lead-204 and lead-206 are isotopes of the same chemical element and no chemical procedure can measure their abundances separately. Physical methods based on the small difference in atomic weight could distinguish between ^{204}Pb and ^{206}Pb, but the effort is time consuming. Finely tuned lasers like those now used to separate ^{235}U and ^{238}U might be more practical, but such techniques have become available only recently. Geochemists now prefer to use other radioactive species that produce easily identified decay products. One such radioactive species is potassium-40 (^{40}K) whose decay product, argon-40, (^{40}Ar), is a noble gas, which is easily distinguished by mass spectrographs from other isotopes of argon. The half-life of ^{40}K is 1.3×10^9 years, which is a convenient magnitude for geological and astrophysical studies.

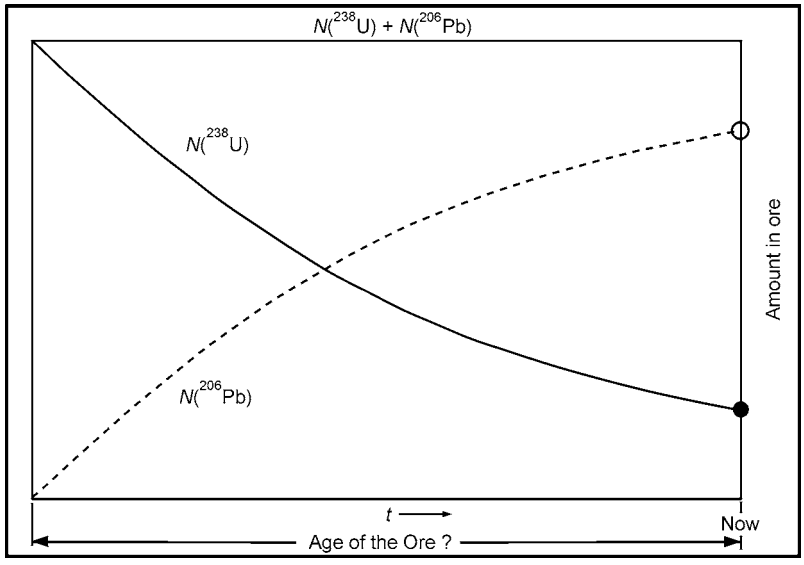


Figure 4.1: An example of radioactive dating of uranium ores. The number of uranium-238 atoms $N(^{238}\text{U})$ has declined; its present value is given by the solid circle, by a known law of exponential decay. Its ultimate decay product lead-206 has a number, $N(^{206}\text{Pb})$, whose present value is given by the open circle, can also be measured. The sum $N(^{238}\text{U}) + N(^{206}\text{Pb})$ must remain a constant when extrapolated back in time. Therefore, the age of the ore cannot be older than a time when this sum consists of a contribution only from ^{238}U and nothing from ^{206}Pb.

By techniques described above, geochemists have found the ages of the oldest rocks on Earth to be about 3.8 billion years. It is, however, believed that the Earth itself is older than this. Rocks older than 3.8 billion years have apparently not survived to this day, or else there were none that solidified in an early phase of the evolution of the Earth. The radioactive dating of lunar rocks shows that the oldest lunar rocks are about 4.2 billion years old. Scientists speculate that still older rocks may lie on the other side of the Moon, where less destruction of lunar rocks by meteoroid bombardment may have occurred. In contrast with the spread of ages found for Earth and Moon rocks, the radioactive dating of meteorites show them to be all the same age, to within the errors of measurement, between 4.5 and 4.7 billion years old. The interpretation has grown that the first objects solidified in the solar system about 4.6 billion years ago. This is the probable age of the entire solar system.

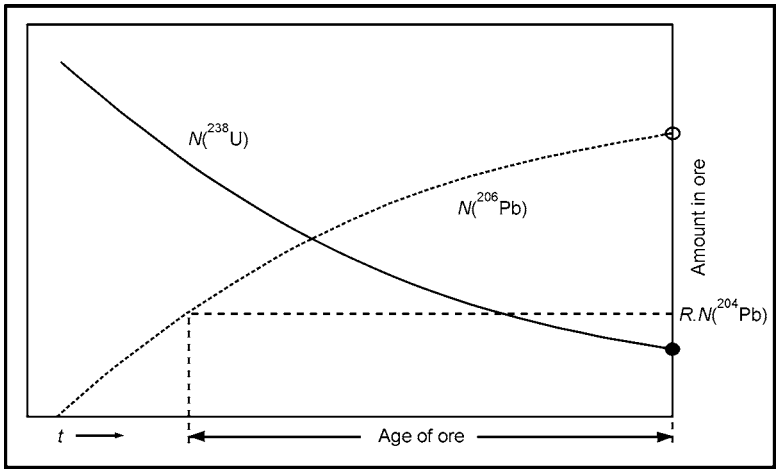


Figure 4.2: A sophisticated method for dating rocks that contain uranium. As in Fig. 4.1 we begin by measuring the present amounts of ^{238}U and ^{206}Pb, given by the solid and open circles, respectively. We also measure the amount of lead-204 (^{204}Pb), which is not a product of radioactive decay, and multiply that amount by R, the ratio $[^{206}Pb]/[^{204}Pb]$ of lead-206 to lead-204 found in non-radioactive rocks. The product $R.N.$ (^{204}Pb) is then assumed to be the amount of primitive lead-206 incorporated into the original rock at its solidification.

4.3 EXPOSURE AGES OF METEOROIDS

The value of 4.6 billion years refers to the time since the asteroids solidified. However, meteoroids are fragments of asteroids. There is another age for meteoroids; the

elapsed time since the shattering that produced the meteoroid. This latter age is possible to obtain as the solar system suffers a continuous bombardment by high-energy, cosmic-ray particles. These high-energy particles can smash the heavy atomic nuclei that exist in an iron meteorite, and produce lighter nuclei, some of which may be radioactive. Before the asteroid was shattered, the particular piece of meteorite, which eventually fell to Earth, may have been deep inside the main body, and therefore shielded from the external cosmic-ray bombardment. Sophisticated techniques exist to measure the exposure ages of meteoroids, based on the idea that longer exposures lead to a greater concentration of long-lived isotopes relative to short-lived ones, as the latter attains an equilibrium between production and decay. Such measurements yield exposure ages of hundreds of millions of years for the iron meteorites and tens of millions of years for the stones. Discrepancies arise from the use of radioactive species with different half-lives. Such apparent discrepancies originate in a variable flux of galactic cosmic rays encountered by the solar system.

4.4 AGE OF THE RADIOACTIVE ELEMENTS

Radioactive dating permits geochemists and astrophysicists to measure the age of the radioactive elements. Since radioactive elements like uranium exist naturally, they must have been created by natural processes some finite time ago; otherwise, uranium would all have decayed into lead by now. Two isotopes of uranium are ^{238}U and ^{235}U, which have half-lives of 4.6×10^9 yr and 7.1×10^8 yr, respectively. In uranium ore, we can measure the current proportion of ^{235}U to ^{238}U. If we knew the primordial ratio, we could establish the age of the elements themselves from the different rates of decay of the two isotopes of uranium.

Heavy nuclei were formed inside massive stars and subsequently expelled in supernova explosions. The relative proportion in which the two isotopes of uranium would be formed are ^{235}U and ^{238}U via neutron capture under the *r*-process with fairly good certainty. However, the radioactive elements in the solar system were not produced in a single supernova event. The use of just ^{235}U and ^{238}U does not permit one to test the validity of this statement. The use of other radioactive isotopes shows that the radioactive elements incorporated in the material that made up the solar system were produced by a continual series of nucleo-synthesis and ejection processes that took place before the actual formation of the solar system. Gerald has studied this problem in detail, using as many radioactive species as possible to avoid too many theoretical assumptions. Different interpretations are possible for the several anomalies that arise from such analyses. The safest

conclusion that we can draw is that the age of the radioactive elements is probably somewhere between 7 and 15 billion years. In order of magnitude, this delineation is in good agreement with astronomers' estimates of the age of globular cluster stars and the age of the Universe. The three independent methods of establishing the ages of ancient objects produce the same number within a factor of two or so, which is very reassuring (Fig. 4.3). We can state with considerable confidence that the Sun and the solar system are probably less than half the age of the oldest objects in the Universe.

Fig. 4.3 shows estimates of the ages of three ancient objects. The horizontal bars indicate the spread of estimates of ages of the Universe, globular clusters, and radioactive elements that most astronomers would consider still to be permissible given the uncertainties of the measurements. Notice that there is significant overlap in the 9 to 15 billion year range. Thus, the age of the solar system, 4.6 billion years, is likely to be between one-half and one-quarter of the age of the oldest objects in the Universe.

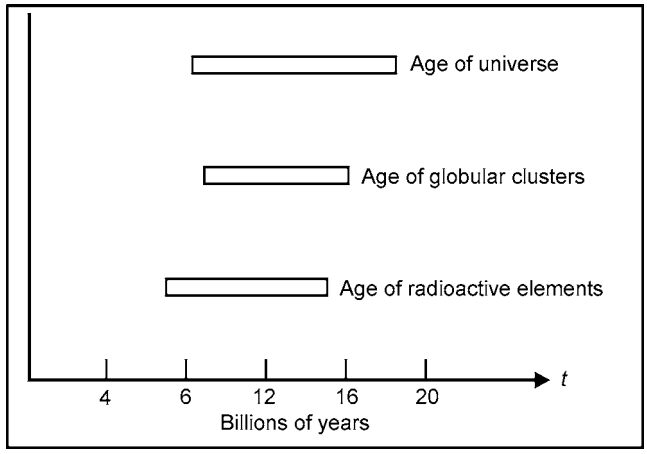


Figure 4.3: An estimate of the ages of three ancient objects.

4.5 MOTIONS OF THE PLANETS

Since Kepler's work in the early part of the seventeenth century, it has been known that the planets follow his three general laws of motion, which are listed in Table 4.1.

Table 4.1: Kepler's three laws governing planetary motions.

1.	The orbit of a planet forms an ellipse with the Sun at one focus.
2.	The Sun-planet radius vector sweeps out equal areas in equal times.
3.	The square of the period of revolution of a planet is proportional to the cube of the semi-major axis of its elliptical orbit.

The geometric meaning of Kepler's three laws can be explained easily. The first law states that the orbits of all planets with long orbital periods and short orbital periods take the form of ellipses with the Sun at one focus. The second law states that equal areas as indicated by the shaded regions in the bottom diagram (Fig. 4.4) are swept out in equal times. The third law states that the square of the period is proportional to the cube of the semi-major axis: $\tau^2 \propto a^3$ with a constant of proportionality, which is the same for all planets. After Kepler, it was Newton who coupled his laws of mechanics with his theory of gravity and explained the Kepler's empirical discoveries in terms of dynamics.

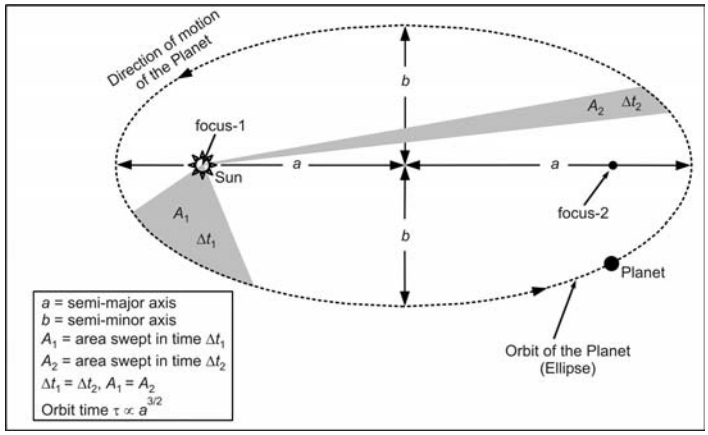


Figure 4.4: Kepler's laws for planetary motions around the Sun.

4.6 EMERGENCE OF LIFE ON EARTH

This topic is, an unfinished story, since humans have a great potential to change the fact of the Earth. One may continue here to the relatively early history of the Earth, where the changes that occurred were in the atmosphere of the Earth or, more generally, what is called the biosphere.

4.6.1 The Evolution of the Earth's Atmosphere

The primitive atmosphere was rich in hydrogen-bearing (CH_4) compounds like methane and ammonia (NH_3). Such an atmosphere is said to be reducing, due to the chemical reactions that would result if ordinary substances were exposed to such gases. Fig. 4.5 gives an educated guess at how geography will be taught 50 million years into the future (e.g., Australia is about to become part of Asia).

Chemists use the word *reduction* when a substance gains valence electrons. In the traditional view, the primitive atmosphere was highly reducing and contrasted sharply with the present atmosphere. The present atmosphere of the Earth is mostly composed of molecular nitrogen (N_2) and molecular oxygen (O_2). When exposed to an atmosphere with free oxygen a substance tends to lose valence electrons to the oxygen atoms it combines with and the substance is said to be *oxidized*, with oxygen being the *oxidizing agent*. Since *oxidation* refers to the process of losing electrons, modern usage permits it to take place even when oxygen itself is not involved in the chemical reaction. The present atmosphere is thus highly oxidizing. Water is present in both the primitive and the present day atmosphere. Except when attacked by alkali metals or halogens, water vapor is not reactive in oxidation or reduction. In H_2O, hydrogen atoms are stably oxidized and the oxygen stably reduced. Liquid water provides an excellent solvent for many chemical reactions. For example, the weathering of rocks is accelerated by the presence of liquid water, which can exert a weak attraction on the solid surface, and eventually dissolve it. Oxidation reduction is an important aspect of chemical reactions.

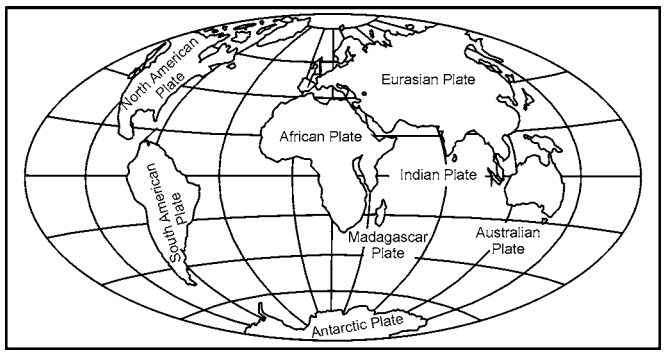


Figure 4.5: World geography as it may be taught 50 million years from now.

It is now believed that the change of the Earth's atmosphere — from a reducing to an oxidizing one — was a two-stage process. The first step is inorganic and depends on the production of ozone, while the second step is organic and depends on the appearance of life. Water plays a vital role in both steps.

4.6.2 The Formation of the Ozone Layer

In the absence of free oxygen, UV rays from the Sun penetrate all the way to the ground and to sea level (Fig. 4.6). Energetic ultraviolet photons dissociate water vapor (H_2O) into hydrogen atoms (H) and free oxygen atoms (O). The Earth cannot retain light hydrogen atoms, and the hydrogen can be expected to escape into space. In contrast, the free oxygen atoms could combine with themselves to form molecular oxygen (O_2) or they might combine with ammonia, NH_3 and methane, CH_4 to form molecular nitrogen, (N_2) and carbon-dioxide (CO_2). Some of the oxygen-carrying compounds would eventually reach the upper atmosphere where photo-dissociation would again produce free oxygen atoms. These free oxygen atoms (O) combine with the molecular oxygen, (O_2), which has not been dissociated to form ozone (O_3). Ozone is a prolific absorber of UV radiation. A slight accumulation of ozone in the upper atmosphere suffices to prevent any more UV radiation from penetrating to lower levels where there is an abundant source of water vapor. With the development of the ozone layer, equilibrium is established between the destruction of ozone by ultraviolet photons and the formation of ozone by the chemical combination of oxygen atoms and molecules. At this point, only visible light would reach sea level. The shielding by the ozone layer prevents further dissociation of water vapor. The amount of molecular oxygen (O_2) present in the atmosphere would then stabilize at some small percentage. The water in the oceans would also be saved from further damage.

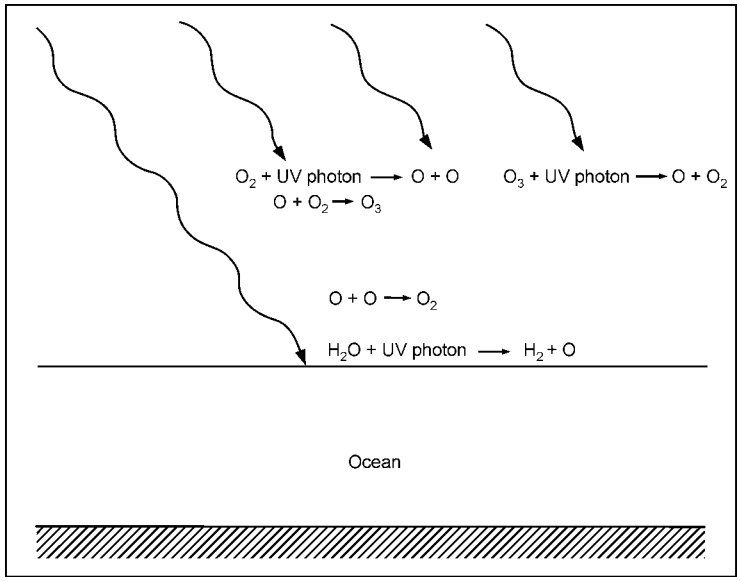


Figure 4.6: Schematic formation of the ozone layer. UV light, which strikes water molecules in either the primitive oceans or the atmosphere dissociates them into hydrogen and oxygen.

The role of the ozone layer in filtering out the UV rays of the Sun cannot be overemphasized. UV rays are very harmful to many lifeforms, inducing skin cancer in humans. This explains the concern over our release into the atmosphere of fluorocarbons, which have an enormous potential for destroying ozone. In geological times, the dissociation of water by ultraviolet light was halted by the appearance of the ozone shield. The full conversion of the Earth's atmosphere from a reducing one to an oxidizing one had to wait for the development of life.

4.7 THE ROLE OF LIFE IN CHANGING THE EARTH'S ATMOSPHERE

Life played a major role in the evolution of the Earth's atmosphere. Walcott was the first to draw attention to thinly layered deposits of limestone rock called stromatolites and interpreted these structures as fossilized reefs of angle growth. The angles are now gone, but they have left behind 'beds of stone.' Paleontologists remained skeptical until 1954, when Tyler and Barghoorn reported the finding of a large number of fossil microorganisms in similar structures. Some of the stromatolite deposits in Africa and Australia are known to date back to 3.5 billion years ago. This suggests that life developed on Earth soon after the crust re-solidified after the melting of the Earth. In 1980, Awramik and Schopf discovered microfossils of long filamental chains of bacterial cells in rocks from the region of the Australian stromatolites. Around 3.5 billion years ago, unicellular organisms existed on Earth that could perform photosynthesis, a process by which organisms that contain chlorophyll convert carbon dioxide and water into carbohydrates. Carbohydrates are the basic source of chemical energy for all organisms. Schematically, for example, the production of glucose, $C_6H_{12}O_6$, by photosynthesis can be expressed as

$$6CO_2 + 6H_2O + \text{light energy} \xrightarrow{\text{Chlorophyll}} C_6H_{12}O_6 + 6O_2.$$

The carbon dioxide on the left-hand side could have come from geological activity when carbonate rocks are heated or from biological activity via respiration. Respiration involves the burning of carbohydrates with oxygen to release chemical energy, e.g.,

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O.$$

Respiration is, therefore, the opposite of photosynthesis. As long as there was more photosynthesis activity than respiration, there would have been a slow build-up of free molecular oxygen in the atmosphere. It is conjectured that photosynthesis by blue-green algae began to release free oxygen into the Earth's atmosphere around 3.5 billion years ago. There is evidence that the amount of

free oxygen was not substantial until two billion years ago. In rocks more than two billion years old, geologists find reduced minerals, which implies that these rocks were not exposed to much free oxygen when they formed, while in rocks less than two billion years old, geologists find oxidized minerals. This means that in the intervening period, within a few hundred million years, there was a major shift of the Earth's atmosphere from reducing to oxidizing. The shift is responsible for all the iron in the oceans to rust and sink (i.e., to precipitate to the ocean floor). When the iron buffer in the oceans had been swept clean, the oxygen released could begin to rise to modern atmospheric concentrations (Fig. 4.7). The shift from a reducing to an oxidizing atmosphere allowed a wild proliferation in varieties of microorganisms. The large number of microfossils turned out to be amazingly diverse. The organisms have shapes that include long filaments, umbrellas, spheroids, and stars. Fossil stromatolites became quite common after about 2.3 billion years.

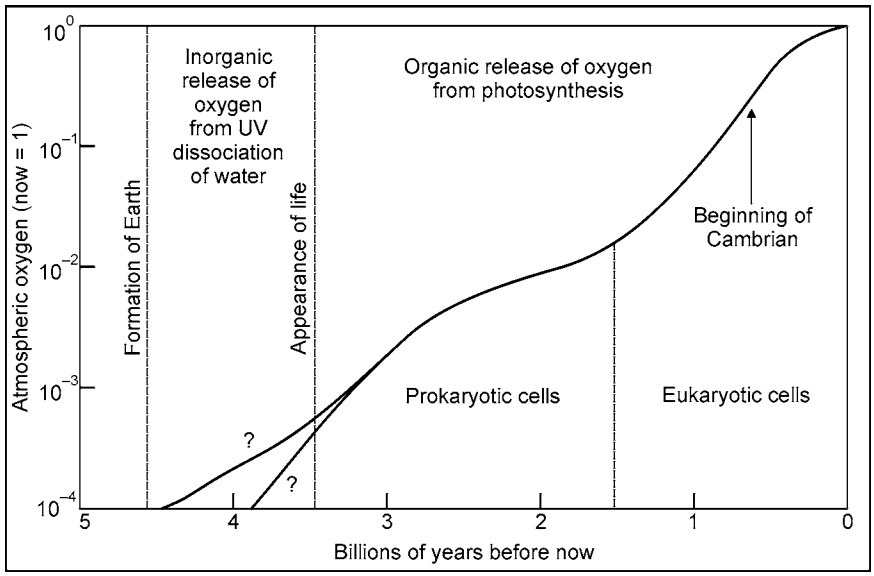


Figure 4.7: Schematic plot of free oxygen in the Earth's atmosphere in the past as a fraction of the present supply.

The shift from a reducing to an oxidizing atmosphere is caused by blue-green algae or similar simple organisms. Only in rocks less than 1.5 billion years old geologists find microfossils of cellular organisms, which are more complicated than blue-green algae. Blue-green algae exemplify prokaryotic cells, which do not contain nuclei. Only in rocks less than 1.5 billion years old do we find examples of eukaryotic cells, cells with true nuclei (Fig. 4.8). Cells with nuclei offer more

versatility than cells without nuclei. For one thing, two cells with the same nuclei structure can serve entirely different specialized biological functions.

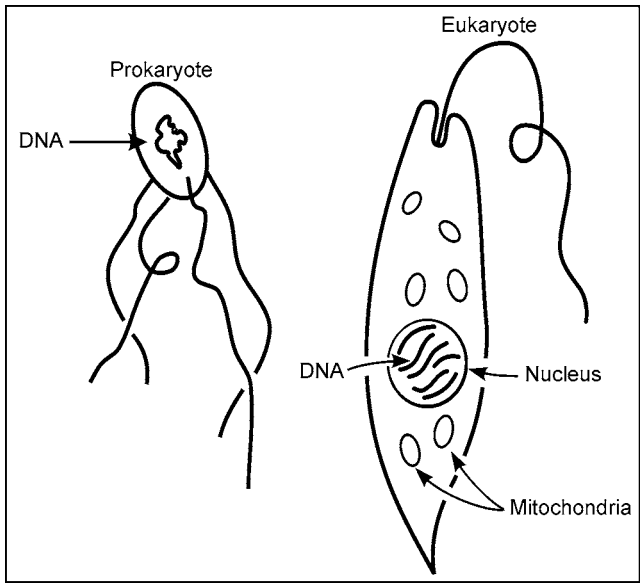


Figure 4.8: Examples of eukaryotic cells in rocks aged less than 1.5 billion years.

EXERCISES

1. What are the two popular theories about the formation of the solar system?
2. Explain clearly the 'Solar Nebula Theory' and the 'Planetsimal Theory'.
3. How was the age of the Earth determined by geologists of the eighteenth and nineteenth centuries?
4. How was the age of the Earth determined (by Rutherford) using radioactive dating?
5. Consider the radioactive dating of Uranium ores to its ultimate decay product of lead, and then explain the radioactive dating technique for determining the age of the Earth.
6. What is the age of the oldest rocks found on Earth by geochemists by applying a sophisticated method of dating rocks?
7. What are meteoroids? What is the idea on which the exposure ages of meteoroids are based?
8. Discuss how radioactive dating permits astrophysicists to measure the ages of radioactive elements.
9. The age of the solar system is likely to be between one-half and one-quarter of the age of the oldest objects in the universe. Please explain.

10. Write down the three laws of Kepler for planetary motion and then explain them briefly.
11. For $^{238}_{92}U$ and $^{206}_{82}Pb$ how many total α and β decays need to occur for a transformation from ^{238}U to ^{206}Pb?
12. Does fluorescence arise from the excitation of the nucleus of an atom?
13. Explain why radium is used in watch dials.
14. Given the ratio of ^{235}U to ^{238}U in uranium ore is 0.007. Assume the formation by a single event and calculate the elapsed time since the creation of uranium. The half lives of ^{235}U and ^{238}U are $t_s = 7.1 \times 10^8$ year and $t_s = 4.6 \times 10^9$ respectively.

Chapter 5

THE STARS IN THE SKY

5.1 INTRODUCTION

A star can be considered a huge sphere made of very hot gas mainly hydrogen and helium, which generates its energy through nuclear fusion. The self-gravitational pull of the star is so high that it keeps its atmosphere of hot gases intact. A fraction of the energy generated in the star is radiated out. The stars are categorized depending on size or mass (i.e., white dwarfs, main sequence, giants, and super-giants).

5.2 LIFE-CYCLE OF THE STARS

It is believed that the Universe was created about 10 to 20 billion years ago with the Big Bang (see Chapter 1). The cosmic expansion resulted in the spreading of helium and hydrogen elements, which are the major constituents of a star. The distribution of the gas varied with time and the expansion of space. In the regions where the gas concentration was large, the mutual gravitational attractions of these gas molecules led to the growth of the first generation of stars.[9] These swirling

[9] A prominent star-forming region is within the Orion Nebula and is called Orion KL. Most of the star systems are separated by many light years.

gaseous clouds gradually became massive, and began to collapse under self-gravity, thereby forming a dense central region called a 'protostar,' which warmed up due to compression. As more and more gases are pulled in the process, the internal pressure mounts up, thereby igniting nuclear reactions in its core. Thus, a star is born. The protostars generally eject gaseous streams in opposite directions. Some of them generate highly focused or collimated outflows, while the rest produce more oval-shaped winds. It is not clear what produces these outflows, but astronomers believe that the magnetic fields and the growing accretion disk around the newborn star play crucial roles. As more and more material is attracted, the protostar spins faster and faster. The increasing rotational energy might direct the stellar winds along the axis of rotation. The powerful magnetic fields might further constrain this outflow thus resulting in bipolar jets extending over several light years. The bipolar outflows disrupt the surrounding gases to a large extent. The motion of these gases may be studied under radio spectrum to understand their effect on the protostar. The heat released by the protostars in the form of ultraviolet radiation breaks apart the surrounding molecular gases into atoms, which are ejected in bipolar outflows. These outflowing gases cool down and recombine into molecules, which emit millimeter waves in radio. These emissions could be mapped and more information on the activity around the protostellar region could be obtained.

It must be noted that the clouds of interstellar gas are very large and their collapse does not originate just one star, but a set of stars called *a star cluster* after having fragmented into smaller clouds. Each fragment may give rise to a single star or a set of stars that orbit around a common barycenter. For example, in our Milky Way Galaxy, single stars are nearly 50% of the total. The majority among the remaining stars are gathered into binary systems, and in some cases, into multiple systems consisting of even six stars.

Until the nuclear furnace in a star remains alive (i.e., until the hydrogen fuel exists in it), the star keeps shining and is said to be a *main sequence* star. After the fuel ends, the outer layers of the star can no longer resist the central force of gravity, which is pulling the star's outer matter inward toward its core. The length of a star's life also depends on its mass. While some of the stars may explode violently through a supernova and rapidly create even heavier elements and spew much of the stellar material into space, in other cases, the process could be slower, and instead of an explosion the elements from the star's interior zones rises to the surface and get lost in space when the outer layers blow off. In either case the end results are similar (i.e., the space between the stars gets enriched with heavy elements, many of which condense to form small solid grains). These processes of the birth and death of stars occur over and over again, with each successive generation of stars starting off with a greater quantity of heavy elements than the previous generation.

5.3 THE BLACK BODY MODEL OF A STAR

Depending on the temperature of an object, it thermally radiates a unique spectral flux. A black body is an object that absorbs all the electromagnetic energies incident on it and does not reflect anything. On the other hand, if the temperature of the black body is non-zero it radiates energy governed by Planck's law in Eq. (5.1).

$$I(\nu, T) = \frac{2h\nu^3}{c^2}\left(\frac{1}{e^{\frac{h\nu}{kT}} - 1}\right) \tag{5.1.1}$$

$$I(\nu, T) = \frac{2hc^2}{\lambda^5}\left(\frac{1}{e^{\frac{h\nu}{\lambda kT}} - 1}\right) \tag{5.1.2}$$

The quantity I is called the *spectral radiance* or *spectral intensity*, which is the energy per unit time per unit surface area per unit solid angle per unit frequency and is measured in $\mathrm{W\,m^{-2}\,sr^{-2}\,Hz^{-1}}$, ν is the frequency of radiation measured in hertz, T is the temperature of the black body in K, h is the Planck's constant ($6.6260693 \times 10^{-34}$ Joules/Hz), c is the velocity of light (2.998×10^8 m/sec), e is the base of natural logarithm (2.718282), and k is the Boltzmann constant ($1.3806505 \times 10^{-23}$ joule/kelvin). If the black body is in thermal equilibrium with the surroundings, then Kirchoff's law states that the emissivity of the body is equal to its absorptivity. The radiated energy increases rapidly with increases in temperature and can be expressed by Steffan-Boltzmann law as

$$F = \sigma T^4 \ \mathrm{W/m^2}. \tag{5.2}$$

Here, F is called the energy flux or power per unit radiated from an object and $\sigma = 5.7 \times 10^{-8}$ W m^{-2} K^{-4} is the Steffan-Boltzmann constant for all wavelengths. The total energy radiated from an object per unit time is defined as the *luminosity*. The surface area of a star can be calculated from its radius R as $4\pi R^2$ and thus the luminosity L of the star can be expressed as

$$L = 4\,\pi\,R^2\,\sigma\,T^4 \ \mathrm{W}. \tag{5.3}$$

Eq. (5.3) clearly indicates that the luminosity of a star is directly proportional to its surface area and to the square of the square of its temperature. The wavelength at which the emission spectra of black body is maximum, i.e., λ_{peak} is also dependent on its temperature and is governed by Wien's displacement law as expressed in Eq. (5.4).

$$\lambda_{peak} = \frac{2.9 \times 10^{-3}}{T}\mathrm{m}. \tag{5.4}$$

When viewed from a distance like from Earth, the observed flux of an object, also known as its apparent brightness, is the power that we actually receive from it. This quantity is generally measured in Wm^{-2} and is dependent on the distance of the object. If r is the distance between the object and point of measurement, then apparent brightness or observed flux f can be expressed as

$$f = \frac{L}{4\pi r^2}. \tag{5.5}$$

The brightness of the star is sometimes expressed logarithmically and called *magnitude*. These could be (*i*) apparent or (*ii*) absolute, respectively, denoted by m and M. The apparent magnitude m of a star is a measure of its apparent brightness as seen by an observer on Earth. Since the amount of light received actually depends on the thickness of the Earth's atmosphere coming in the line of sight to the object, the apparent magnitudes are normalized to the value it would have outside the atmosphere. The apparent magnitude m_x in the spectral band x is defined as

$$m_x = -2.5 \log_{10} (f_x) + C \tag{5.6}$$

Here, f_x is the flux at the spectral band x and C is a constant, which depends on the units of the flux and the band. Thus, the dimmer an object appears, the higher its apparent magnitude. On the other hand, absolute magnitude refers to a measurement at a fixed distance of 10 parsecs. That is, if the luminous object is placed at a fixed distance of 10 parsecs from Earth, the magnitude measured would be its absolute magnitude. If the actual distance of the object is known, then the absolute magnitude M is expressed as

$$M_x = m_x - 5(\log_{10} D_L - 1). \tag{5.7}$$

Here, D_L is the luminosity distance in parsecs. For the nearby objects like the stars in our galaxy, the luminosity distance D_L is more or less identical to the real distance, since the space-time is almost Euclidean. For much more distant objects, this approximation is invalid and the general relativity must be applied for calculating. Table 5.1 lists the apparent and absolute magnitudes of some stars.

Although in a true sense a star is not strictly a black body, it may be approximated with it within some range of frequency. The reason being is that the star contains various elements, which produce absorption lines at different wavelengths depending on the elemental composition of its photosphere and chromospheres. Thus, if the luminosity of a star is plotted as a function of wavelengths, absorbing type glitches would be found on the curve. Moreover, the star also produces sudden active emissions such as radio bursts, etc., whose intensity are much higher than when the star is quiet. If the glitches in the spectrum produced out of absorption lines in a star are neglected, then within a small range of spectrum the quiet stars may be approximated to a black body as shown in Fig. 5.1.

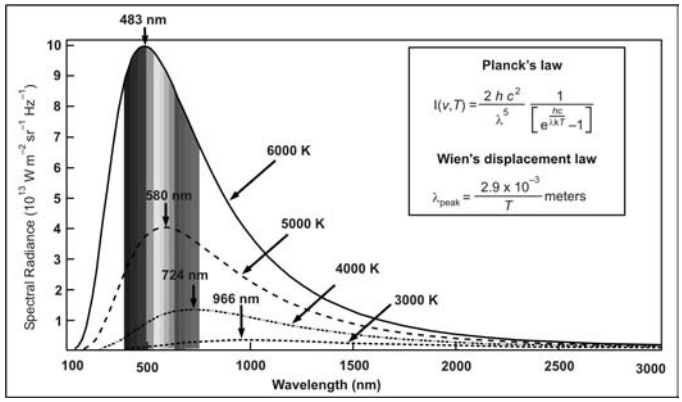


Figure 5.1: Approximate black body models of stars at different temperatures using Planck's law. Note that the wavelength at peak emission is highly temperature dependent and is governed by Wien's displacement law.

A star like our Sun may be approximated to a black body when the glitches due to spectral absorption are neglected and when it is quiet for within a range of 10^{-7} to 10^{-4} m wavelengths as shown in Fig. 5.2. The Sun produces an extra amount of electromagnetic radiations when certain activities in it are alive. These

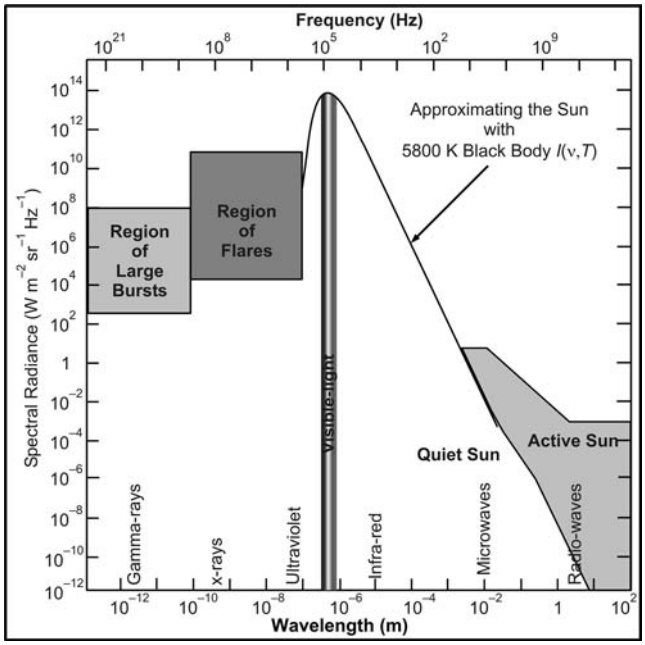


Figure 5.2: The Sun is approximated with a black body at 5800 K. This approximation is reasonable within the range of the wavelength between 10^{-7} to 10^{-4} meters.

confer to the large bursts in the gamma-rays and flares in upper ultraviolet and x-rays, microwaves, and radio waves. The shaded regions in the diagram show the most probable ranges of intensity fluctuation of these special radiations, which vary with time and are unpredictable. However, at certain times, the amplitude range could exceed.

Table 5.1: Apparent and absolute magnitudes of some stars.

S.N.	Name of the Star	Constellation	Apparent Magnitude	Absolute Magnitude	Distance (pc)
1.	Sun	--	–26.7	+4.77	0
2.	Sirius	Canis Major	–1.46	+1.45	2.63
3.	Canopus	Carina	–0.72	–5.63	29.14
4.	Arcturus	Bootes	–0.04	–0.30	11.34
5.	Rigil Kent A	Centaurus	–0.01	+4.34	1.35
6.	Vega	Lyra	+0.03	+0.58	7.66
7.	Capella	Auriga	+0.08	–0.48	12.88
8.	Rigel	Orion	+0.12	–6.75	236.08
9.	Procyon	Canis Minor	+0.38	+2.68	3.37
10.	Achernar	Eridanus	+0.46	–2.76	42.92
11.	Betelgeuse	Orion	+0.50	–5.09	131.84
12.	Hadar	Centaurus	+0.61	–5.42	162.50
13.	Altair	Aquila	+0.77	+2.21	5.21
14.	Aldebaran	Taurus	+0.85	–0.65	19.93
15.	Antares	Scorpius	+0.96	–5.38	183.96
16.	Spica	Virgo	+0.98	-3.55	79.71
17.	Pollux	Gemini	+1.14	+1.07	10.42
18.	Fomalhaut	Piscis Austrinus	+1.16	+1.73	7.67
19.	Mimosa	Crux	+1.25	–3.92	107.31

20.	Deneb	Cygnus	+1.25	–8.73	981.12
21.	Acrux	Crux	+1.33	–3.63	98.11
22.	Rigil Kent B	Centaurus	+1.33	+5.70	41.35
23.	Regulus	Leo	+1.35	–0.53	23.91

5.4 MORGAN-KEENAN SPECTRAL CLASSIFICATION OF STARS

Depending on the temperature of the photosphere of stars, they can be classified under different groups represented by alphabets using Wein's displacement law. However, for distant stars practical difficulties arise in the measurement of absolute luminosity or temperature. In such cases, the stars are identified based on their spectral absorption, which provides information about temperature in a different way. This spectral classification scheme is called *Morgan-Keenan* and was devised between late nineteenth century and the beginning of the twentieth century. The initial criteria used to define the sequence of the stars were primarily based upon the hydrogen absorption lines along with other features. As stated in earlier discussions, the light coming from a star contains missing spectral lines at the absorbing frequencies of the various elements present in the photosphere and chromosphere. Particular absorption lines can be seen only over a certain range of temperatures since it is only in that range that the involved atomic energy levels are populated. The standard classes of the stars are listed in Table 5.2. It is easy to remember the sequence of the class of stars using a popular mnemonic, *O Be A Fine Girl, Kiss Me*.

Table 5.2: Morgan-Keenan classification of stars based upon temperature and Spectra.

Class	Temp × 10^3 K	Color of the Star	Mass × M_Θ	Radius × R_Θ	Lumino-sity × L_Θ	H lines	Other spectra
O	30 – 60	bluish	60	15	1.4×10^6	weak	ionized helium
B	11 – 30	bluish-white	18	7	20,000	medium	neutral helium
A	7.5 – 11	white with bluish tinge	3.2	2.5	80	strong	weak ionized Ca

F	6 – 7.5	yellow-white	1.7	1.3	6	medium	weak ionized Ca
G	5 – 6	light yellow	1.1	1.1	1.2	weak	medium ionized Ca
K	3.5 – 5	light orange	0.8	0.9	0.4	very weak	strong ionized Ca
M	2 – 3.5	reddish orange	0.3	0.4	0.04	very weak	strong titanium dioxide

Before proceeding further, we shall discuss a bit more about the features of stars based on their class. The strength of the hydrogen line is related to the surface temperature of the star. The spectrum of stars are highly size dependent. The super-giant stars often swing between O or B (blue) and K or M (red). But they do not stay in the G classification which is an extremely unstable place for a super-giant. The classes are further subdivided between 0 and 9. For example, G0 is the hottest and G9 is the coolest within the class-G. The Sun is a G2 star.

(1) ***Class-O*:** The O stars are bluish and very hot and luminous with most of their emission in the ultraviolet range. They constitute only 1 out of 32,000 main sequence stars and shine a million times brighter than our Sun. They produce prominent helium lines (ionized and neutral) and weak hydrogen lines. Due to their huge size, their fuel gets consumed quickly thereby leaving the main sequence stars. Examples of O stars are Zeta Puppis and Epsilon Orionis.

(2) ***Class-B*:** The B stars are blue and extremely luminous consisting of neutral helium and moderate hydrogen lines in their spectra. The B stars also live as a main sequence only for a short time like the O stars due to their huge size and large fuel consumption rate. They constitute about 1 out of 770 stars (0.13%) in the main sequence. Owing to their short life span they do not have sufficient time to stray far from the region of their births. Thus, they tend to form clusters[10] known as OB1 associations. These are associated with giant molecular clouds. Our Milky Way Galaxy has one of

[10] These are gravitationally bound groups of stars. Two types of clusters known as globular and open. The globular clusters constitute hundreds of thousands of very old stars. The open clusters generally constitute less than a few hundred stars and are generally very young.

its entire spiral arm as OB1 association known as Orion OB1 association. Examples of B stars are Rigel and Spica.

(3) ***Class-A*:** These are white or bluish-white stars commonly seen with the naked eye. Their spectra consists of strong hydrogen lines and ionized metals. They constitute 1 out of 158 stars (0.63%) in the main sequence. Examples of A stars are Vega and Sirius.

(4) ***Class-F*:** These are white stars with a slight tinge of yellow. Their spectra consists of weak hydrogen lines and ionized metals. These constitute 1 out of 32 stars (3.1%) in the main sequence. Examples of F stars are Canopus and Procyon.

(5) ***Class-G*:** These are yellow or light-yellow stars. Their spectra contain weak hydrogen lines compared to the F stars along with metals (ionized and neutral). These constitute 1 out of 12 or 13 stars (8%) in the main sequence. Examples of G stars are Sun and Capella.

(6) ***Class-K*:** These are orangish or light orangish stars. Their temperatures are slightly less than our Sun. Their spectra consists of extremely weak hydrogen lines or nil along with metals (mostly neutral). These constitute of roughly 1 out of 8 stars (13%) in the main sequence. Examples of K stars are Arcturus and Aldebaran. Some of these could be giants or super-giants and may not be a main sequence star.

(7) ***Class-M*:** These are reddish or reddish orange stars. Their spectra consists of lines belonging to molecules and neutral metals, but the hydrogen is usually absent. More than 78% of the stars in the main sequence are red dwarfs. Examples of M stars are Betelgeuse and Barnard's star. Some of the stars could be super-giants such as Antares and Betelgeuse.

Additionally, a number of new classes have been identified (also called extended classes) from the newly discovered ones which are not described here.

5.5 HERTZSPRUNG-RUSSEL DIAGRAM

Based on the above classifications, a diagram was first prepared in 1910 by Ejnar Hertzsprung and Henry Norris Russel for easy identification known as Hertzsprung-Russel diagram (H-R diagram, HRD) as shown in Fig. 5.3. This diagram is also known color magnitude diagram (CM diagram). It illustrates the relationship between absolute magnitude/luminosity with surface temperature for any stellar classification. The descriptions of the stellar colors are traditional in astronomy and they actually describe the light after it has been scattered by the atmosphere. The Sun is actually a white star without any trace of yellow.

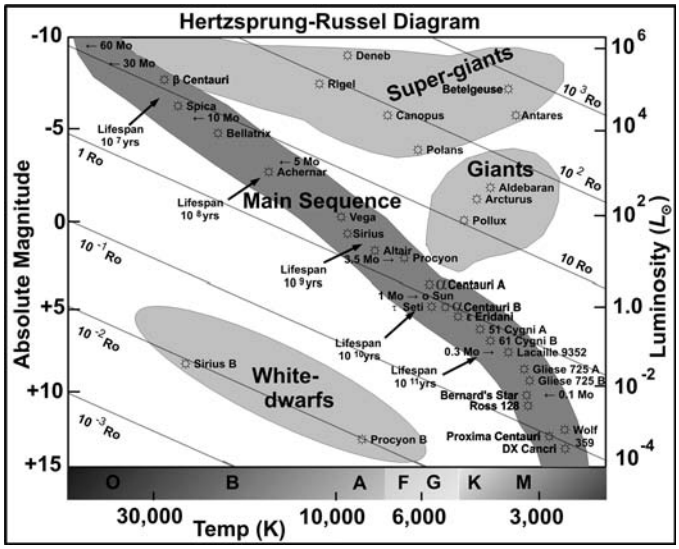


Figure 5.3: The Hertzsprung-Russel diagram showing the spectal classification of the stars. Only a few stars are highlighted by their names including our Sun. Stars falling on the angled lines take the radii indicated on the line. In the main sequence the lifespan of the stars are marked.

An examination of the HRD reveals that the stars tend to fall within certain regions of the diagram. The most predominant is the diagonal belt called the main sequence. The upper left regions hold the stars with very high temperature in the blue region. Note that the masses are very high. In the central region lies our Sun. The bottom right region of the main sequence consist of smaller stars with mass less than our Sun and are relatively much cooler.

5.6 MASS LUMINOSITY RELATIONSHIP OF A STAR IN MAIN SEQUENCE

The luminosity of a star is dependent on its mass and its current state of existence. The star passes through different states of existence like protostar, main sequence, red-dwarfs, super-giants, white-dwarf etc. These are very much dependent on their mass. The luminosity also varies accordingly. For example, a star like our Sun while it was forming, probably it was 100 to 200 times brighter than now. In its first billion years in the main sequence it was probably 2/3 times brighter than present. At the end of the main sequence its brightness would probably reduce to 50% of today following which it would probably swell up to become a Red Giant with its diameter exceeding the orbit of Earth and its brightness increasing about 1000 times of today's brightness. After this it would probably shrink to become a

white-dwarf 1000 to 10,000 times fainter than today. In general stars have a huge variation of luminosities, from red-dwarfs which are 50,000 times fainter than the Sun to super-giants which are 1,000,000 times brighter than the Sun. But if we restrict our study of brightness for the main sequence alone, we would probably be able to quantify the mass luminosity relationship.

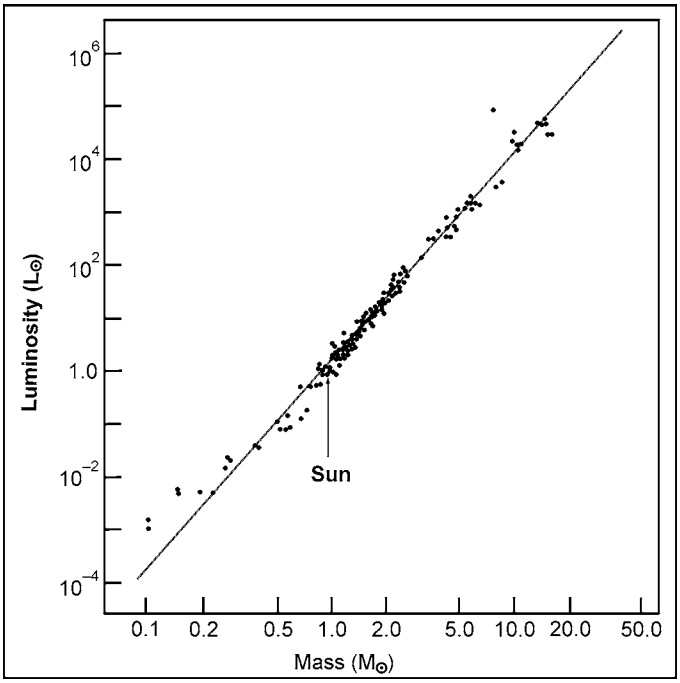


Figure 5.4: The mass luminosity relationship data for various stars in the main sequence.

Fig. 5.4 shows the luminosities of main sequence stars possessing different masses. It may be observed that luminosity steadily increases with mass in a non-uniform manner. In other words, the luminosity is proportional to some power of mass. Further the variations are different for mass above and below 0.43 M_Θ. Based on these the luminosities are related to the masses as follows

$$L = L_\Theta \left(\frac{M}{M_\Theta} \right)^{3.5} \quad \text{for } M \geq 0.43\, M_\Theta \tag{5.8}$$

$$L = L_\Theta \left(\frac{M}{M_\Theta} \right)^{2.3} \quad \text{for } M < 0.43\, M_\Theta \tag{5.9}$$

5.7 THE SIZE LIMITS OF A STAR

In all the above we have seen that the luminosity of the star is dependent on its temperature which in turn is dependent on the mass of the star. Thus there should be a minimum mass below which the temperature of the star's core[11] will not be at sufficient temperature to conduct the nuclear reactions. This is known to be about 0.08 M_Θ or 80 times than the mass of Jupiter. Similarly the limits on the maximum mass of a star based on theoretical modeling is thought to be somewhere between 150 and 200 M_Θ. These very much short lived and rarely found.

EXERCISES

1. How are the stars categorized?
2. Discuss briefly the life-cycle of the stars.
3. What is meant by 'a star cluster'?
4. Give the characteristic features of a 'main sequence' star.
5. What is a black body? Explain the black body model of a star, considering relevant mathematical relations.
6. Explain the apparent and absolute magnitude of different stars.
7. Using Planck's law, discuss the approximate black body models of stars at different temperatures.
8. Plot the spectral radiance against wavelength to show the region of large bursts and regions of flares (due to the Sun).
9. Depending on the temperature of the photosphere of the stars, classify them under different groups and discuss.
10. What is the Morgan-Keenan spectral classification of stars?
11. What is the popular mnemonic to remember the sequence of the class of the stars?
12. Discuss the different features of the stars based on each class.
13. Illustrate the Hertzsprung-Russell diagram to show the spectral classification of the stars.
14. Discuss the mass luminosity relationship of the stars in main sequence and possessing different masses.
15. On what factor does the luminosity of a star depend?
16. Establish that the total amount of energy per unit time crossing an imaginary sphere (centered on the Sun) with radius r is equal to $4\pi r^2 f$, where f is the energy falling per unit time per unit area perpendicularly outside the Earth's atmosphere. Hence from the principle of conservation of energy justify that the quantity must be equal to the luminosity of the Sun.

[11] It is the innermost region of a star where the energy is generated through nuclear fusion.

17. Calculate the numerical value of luminosity of the Sun from the following data: $r = 1.5 \times 10^{13}$ cm and $f = 1.36 \times 10^{6}$ erg sec^{-1} cm^{-2}.
18. Assume the luminosity of the Sun is 3.9×10^{33} erg sec^{-1}; find the linear size of the Sun and its effective temperature. Given: $r = 1.5 \times 10^{13}$ cm and the angular diameter of the Sun $\theta = 32$ minutes of arc.
19. Prove that the total energy of the Earth's orbital motion is equal to one-half of its potential energy.
20. Establish that the numerical value for the Sun's central temperature is equal to 1.5×10^{7} K.
21. Calculate the ratio of plasma-energy density to the radiation energy density at the center of the Sun.

Chapter 6

THE SUN AS A STAR

6.1 INTRODUCTION

It is the high quality of sunshine, which maintains life on Earth. The quantity or sunshine is equally important for maintaining the comfortable average temperature of about 290 K on the surface of the Earth. A happy combination of quality and quantity of sunshine makes Earth a thriving garden.

To get an idea about how high-grade sunshine is, let us start with directly measurable quantities and deduce the luminosity and effective temperature of the Sun. From direct measurements, the energy which falls per unit time per unit area perpendicularly outside the Earth's atmosphere (i.e., the solar constant) is obtained as

$$f = 1.36 \times 10^6 \text{ erg sec}^{-1} \text{ cm}^{-2}. \tag{6.1}$$

From a variety or different measurements, the Earth is known to be located approximately at a distance from the Sun of

$$r = 1.5 \times 10^{13} \text{ cm}. \tag{6.2}$$

Thus, light takes 500 seconds to get from the Sun to the Earth. From the knowledge of f and r and applying the *$1/r^2$ law of radiation* from Eq. (5.5), the luminosity L_Θ of the Sun must be

$$L_{\Theta} = 3.90 \times 10^{33} \text{ erg sec}^{-1}. \tag{6.3}$$

The value of L_{Θ} represents an enormous total power. The Earth intercepts less than one billionth of this power, and even this is mostly wasted by immediate re-radiation in the infra-red back into space.

6.2 ATMOSPHERE OF THE SUN

By comparing the observed spectral-energy distribution of sunlight with the spectral-energy distribution of a true black body of temperature, 5,800 K, we may check the effective temperature of the Sun. This comparison is shown in Fig. 6.1. We see that the agreement is good from the ultraviolet to the infra-red areas of the spectrum. Discrepancies arise in the x-ray, far ultraviolet, and radio portions of the electromagnetic spectrum. These discrepancies are noticeable during flare and burst activity in the Sun. Even gamma rays are detectable in a

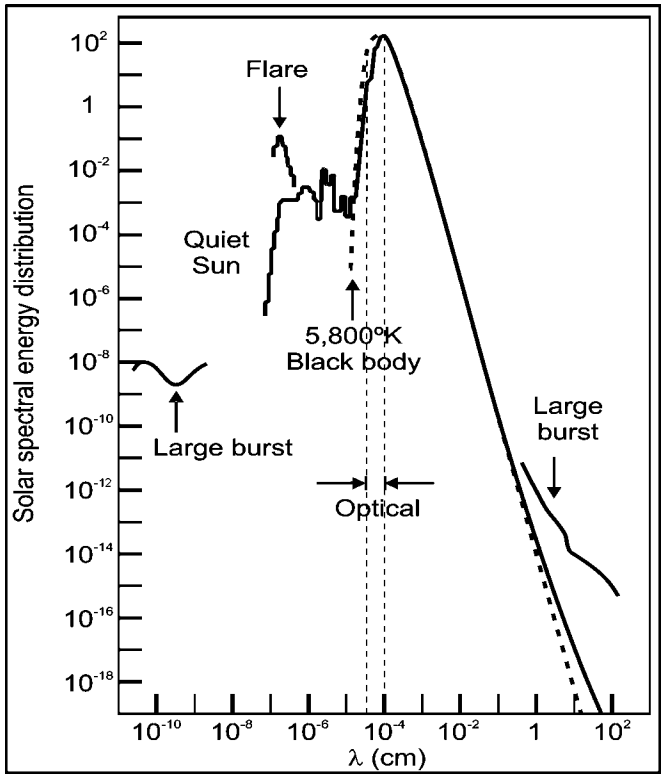


Figure 6.1: Solar spectral energy distribution.

large burst on the Sun. The total energy contained at these wavelengths is small, in the quiet Sun. Near the peak of the Sun's emission a 5,800 K, black body gives a representation of most of the output of the Sun. Even at optical wavelengths, Fig. 6.1 shows detectable departures in the spectral-energy distribution of the continuum radiation, at higher dispersions dark absorption lines appear in the solar spectrum. These clues suggest that the Sun is not at a uniform temperature, and that its surface layers are not in perfect thermo-dynamic equilibrium with its radiation field. The temperature 5,800 K characterizes only part of the surface layers.

Since the zeroth law of thermodynamics states that heat flows from hot to cold, we may infer that the Sun is hotter on the inside. This can be checked observationally by the phenomenon known as limb darkening. The edge of the Sun, called the *limb* in astronomical nomenclature, appears darker than its center. The explanation is given in Fig. 6.2, which shows two lines of sight for an observer; one directed toward the limb of the Sun and the other toward its center. For the same optical path lengths traversed through the outer layers of the Sun, the eye looks closer to point C along a line of sight that passes through the Sun's center than along the Sun's limb. As the deeper layers are hotter, and intrinsically brighter, the limb seems dark relative to the center. This is the origin of *limb darkening*. The evidence of absorption lines can be interpreted to give the same conclusion. We assume that the atmosphere of the Sun is characterized by cooler rarefied regions, which overlie hotter dense regions (Fig. 6.3). Thermal photons flow from the deeper layers with a continuous distribution of wavelengths. Some of them with the proper wavelengths will be absorbed by atoms situated in the upper layers (Fig. 6.3 *b*). The excited atom may then de-excite in two different ways as shown in (Fig. 6.3 *c*).

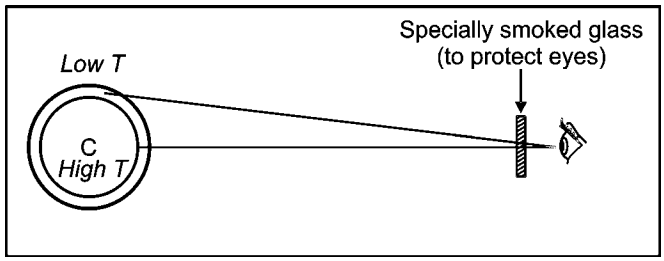


Figure 6.2: The explanation for limb darkening in the Sun. For the same optical path lengths traversed through the outer layers of the Sun (the short solid lines), the eye sees deeper and hotter layers along a line of sight aimed through the Sun's center than along one aimed through the Sun's limb.

(*i*) It may de-excite collisionally with the excess energy carried off as extra kinetic energy of the two colliding partners. Since the temperature of the gas in the upper layers is lower than the temperature of the radiation

field emerging from the lower layers, statistically there are more radiative excitations followed by collisional de-excitations than there are collisional excitations followed by radiative de-excitations. So, there is more absorption than emission. The net effect is the removal of some line photons from the underlying continuum radiation field. This mechanism for producing the intensity at the wavelength of the line which is darker than the intensity at the wavelengths of the neighboring continuum is known as *true absorption*.

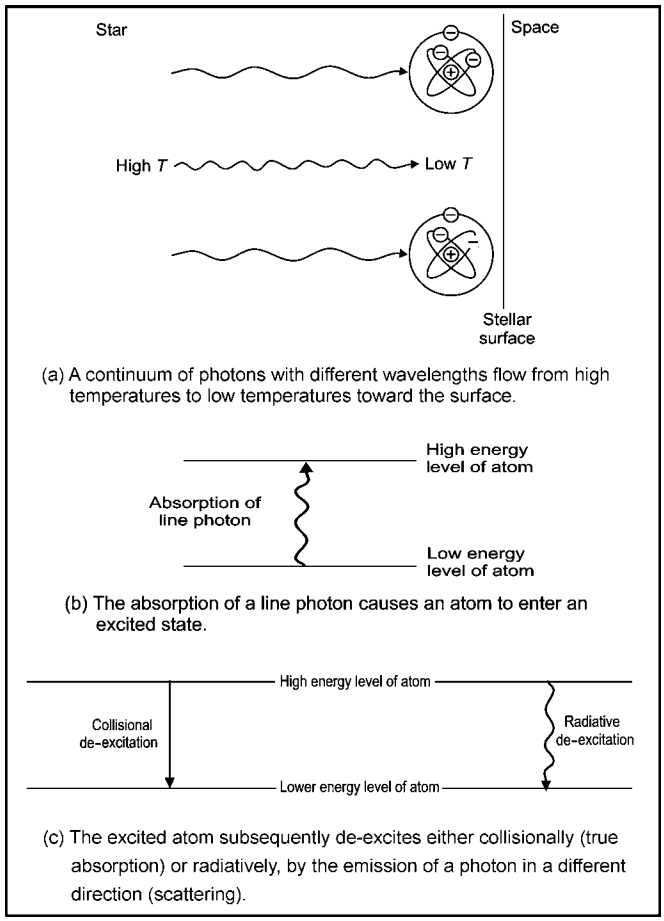


Figure 6.3: Formation of a dark line in a stellar photosphere. (a) A continuum of photons with different wavelengths flow from high temperatures to low temperatures toward the surface. (b) The absorption of a line photon causes an atom to enter an excited state. (c) The excited atom subsequently de-excites either collisionally (true absorption) or radiatively by the emission of a photon in a different direction (scattering).

(*ii*) The radiatively excited atom may also de-excite radiatively, with the emission of a line photon traveling in a different direction than the original photon. This process of *resonant scattering* does not constitute *true absorption*. It can also lead to the appearance of a dark line. The statistical condition of *detailed balance* is equivalent to the condition of true thermodynamic equilibrium. However, in a stellar atmosphere, the presence of a boundary prevents *detailed balance*. There exists a basic asymmetry at the surface of a star. A hot, bright star lies to one side while cold, dark space lies to the other. The net loss of line photons from our line of sight compared to the neighboring continuum photons leads to a dark *scattering* line. Dark lines are produced generally by a combination of *true absorption* and pure scattering. The quantitative study of the formation of the continuum radiation and the pattern of absorption that lies in the outer layers of a star is part of the discipline called *stellar atmosphere*. The layers that can be analyzed in this way only constitute the outer 0.1 percent of the radius of the Sun. These layers, where optical photons can fly to us more-or-less directly, are collectively called the *photosphere of the Sun*.

6.3 INTERIOR OF THE SUN

In order to discover the conditions in the deep interior of the Sun astronomers have to resort to theory as there are no direct observations of the Sun's interior. This theory is complex, though the main qualitative ideas are simple.

For getting the mass of the Sun we know that the Earth orbits the Sun approximately in a circle with radius $r = 1.5 \times 10^{13}$ cm and period $P = 1$ year $= 3.16 \times 10^{7}$ sec. Therefore, the speed of the Earth's motion about the Sun is

$$v = \frac{\text{circumference}}{\text{period}} = \frac{2\pi r}{P} = 2.98 \times 10^{6} \text{ arc sec}^{-1}. \tag{6.4}$$

If we ignore the motion of the Sun, we have $F = ma$ in the form where a is the centripetal acceleration (magnitude = v^2/r) and F is the gravitational force ($F = G M_{\Theta}/r^2$) with m = mass of the Earth. Solving we get the mass of the Sun M_{Θ} as

$$M_{\Theta} = \frac{rv^2}{G} = 2.0 \times 10^{33}. \tag{6.5}$$

The Sun is thus the most massive body in the solar system. It is about 300,000 times more massive than the Earth. The Sun is about 100 times larger than the Earth. So, the mean density of the Sun is not very different from the mean density of the Earth. The mean density of the Sun is

$$\text{mean desnity of the sun} = \frac{\text{mass}}{\text{volume}} = \frac{M_\Theta}{4\pi R_\Theta^3} = 1.4 \text{ gm/cm}^3. \quad (6.6)$$

This is to be compared with the average density of the Earth, 5.5 gm/cm^3.

To estimate the pressure, we consider the observation that the Sun is not shrinking or expanding rapidly. For the Sun to remain mechanically static, all forces must be very nearly in balance, the condition of hydrostatic equilibrium. This hydrostatic equilibrium involves the balance of gravitational force and pressure force at each level in the star (Fig. 6.4). In hydrostatic equilibrium, the pressure at radius *r* equals the weight of the vertical column of matter of unit area above it. In this way, when the forces are balanced at all values of *r*, no net force acts on any element of gas, and the star remains mechanically steady.

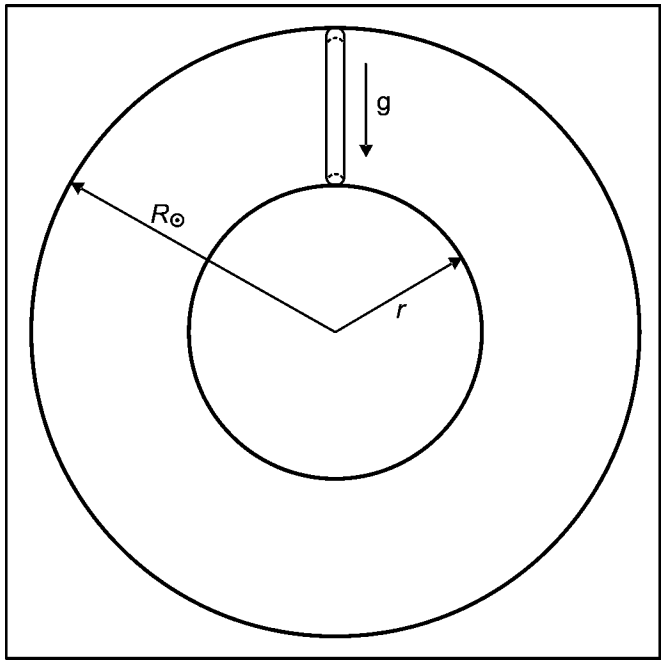


Figure 6.4: Hydrostatic equilibrium involves the balance of gravitational force and pressure force at each level in the star.

The high mass of the Sun indicates that the pressure needed at the center to support the overlying layers against its own weight must be enormous. In such high-pressure regions, neutral atoms and molecules would be pushed so hard against one another that their electronic shells could not possibly survive intact. Thus, the matter in the interior of the Sun must be composed of bare nuclei and free electrons, a state of matter called plasma. Although such a mixture of ions

and electrons is electrically neutral on average, the sizes of the particles are much smaller than atoms or molecules. Consequently, even at the relatively high mean density of the interior of the Sun, there exists plenty of room between the individual particles (Fig. 6.5) so we may say that the plasma in the Sun behaves as a perfect gas.

The pressure of a perfect gas is associated with the random thermal motions of the particles. If the gas has a number density n and a temperature T, the perfect gas law states that the total pressure

$$P = nKT. \tag{6.7}$$

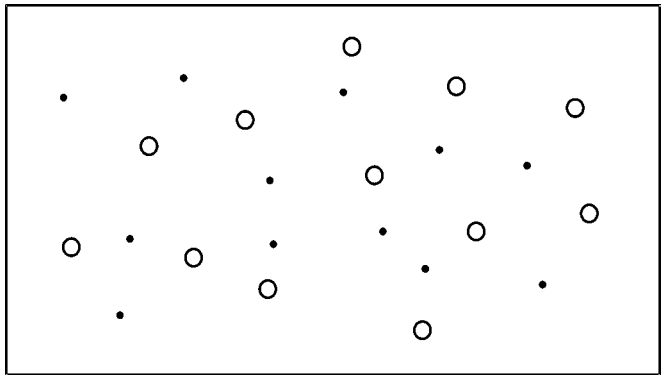


Figure 6.5: The solar plasma contains the bare nuclei of atoms (open circles) plus their stripped orbital electrons (black dots). The mean separation between particles is much larger than their individual sizes and the medium is electrically neutral as a whole. The electric energy associated with neighbors is small in comparison to the thermal energy. Thus, the solar plasma can be considered a perfect gas.

The central mass density ρ_c in the Sun is about 110 times its average value 1.4 gm/cm^3 (i.e., $\rho_c = 150$ gm/cm^3). The mean particle mass m_c at the center is less than the mass m_p of the proton: $m_c = 1.5 \times 10^{-24}$ gm. Thus, the central number density $n_c = \rho_c / m_c = 1.0 \times 10^{26}$ cm^{-3} with the pressure $P = 2.1 \times 10^{17}$ erg cm^{-3}. Eq. (6.7) implies that the Sun's central temperature must equal 15 million degrees Kelvin.

$$T_c = 1.5 \times 10^7 \text{ K}. \tag{6.8}$$

At the densities and temperatures characteristic of the interior of the Sun, the matter is quite opaque. As a consequence, we can expect the matter and the radiation field to be very nearly in thermodynamic equilibrium with each other locally. The thermal radiation field associated with a central temperature of 15 million degrees contains mostly x-rays. X-rays interact quite strongly with matter, and the x-rays in the interior of the Sun fly only about half a centimeter on the

average before they crash into a material particle and are either absorbed or scattered into a completely different direction.

6.3.1 Stability of the Sun

The Sun has a built-in safety valve. It *burns* by *controller thermonuclear fusion*. Fig. 6.8 shows how the Sun's safety valve for thermonuclear fusion works. If the power L_c produced in the core exceeds the luminosity $\boldsymbol{L}$ escaping from the surface, the net input of energy into the Sun would cause it to expand slightly (outward straight arrows in figure). This would lower the central pressure and the central temperature, lowering the rate of energy generation until $L_c = L$. Conversely, if the power L_c should be less than L, the net loss of energy would cause the Sun to contract slightly (inward straight arrows in figure). This would increase the central pressure and the central temperature, raising the rate of energy generation until $L_c = L$.

The goal of modern plasma physics is to find other confinement mechanisms. Self-gravity is impractical without astronomical masses, as gravity is such a weak force intrinsically. The two promising proposals involve magnetic confinement and inertial confinement. The former aims at true *controlled thermonuclear fusion*, and the latter hopes to achieve ignition by bombarding a small pellet of nuclear *fuel* with laser beams or particle beams until it implodes.

6.3.2 Radiative Transfer in the Sun

Eddington first appreciated that there is a net slow effect in the seemingly crazy dance performed by the photons inside stars like the Sun. The dance of the photons is called a *random walk* and is to be distinguished from free flight. By such a random walk, a photon is able to zigzag from the center of the Sun and eventually make its way to the surface (Fig. 6.6). In executing this path, photons suffer many interactions with matter and are transformed each time. The photons are in fact degraded from being x-ray photons characteristic of the central temperature of 1.5×10^7 K to optical photons characteristic of the effective temperature of 5,800 K at the surface. This *degradation* of high-quality x-ray photons to *low-quality* optical photons is to be expected from the Second Law of Thermodynamics.

6.3.3 Source of Energy of the Sun

As the photons slowly leak out of the Sun, new photons are created to take their place as the matter is still hot. Life has existed on Earth for more than 3 billion years and during that interval, at least, the Sun must have been shining more or less stably with luminosity close to its present value. The store of thermal energy present in the Sun can withstand such a steady drain for only 1 percent of 3 billion

years. So there must be some other, much larger supply of energy that explains the Sun's power.

Although thermonuclear energy is a very powerful source, it does not have infinite capacity to offset radiation from the surface of a star. Thus, sooner or later, every star gets hotter and hotter while losing more and more energy to a cold surface. This confrontation provides the central plot to the life stores of the stars.

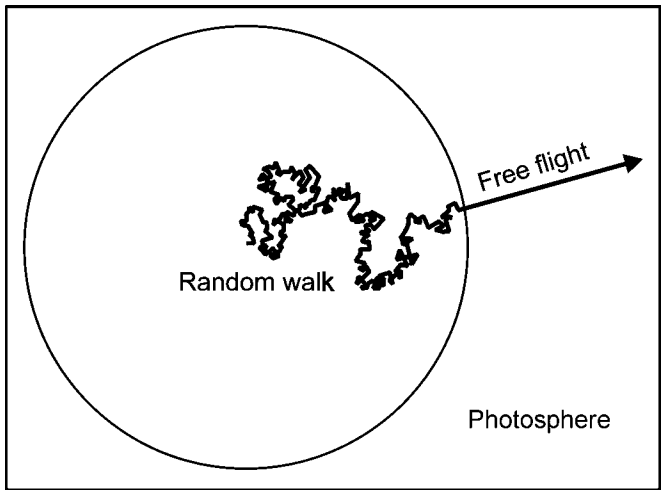


Figure 6.6: Radiative transfer in a star in the stellar interior. Photons suffer many interactions with matter and slowly diffuse outward by a tortuous series of 'random walks,' which eventually bring them to the photosphere. In the process, one x-ray photon is dragged into many optical photons.

One source of energy is gravitational energy. The basic idea is based on the virial theorem. The Sun is a statistically steady object. The total energy E of the Sun is numerically equal to half its gravitational potential energy. In the Sun, it is the thermal energy of the plasma rather than the kinetic energy of orbital motion that produces the resistance to gravitational contraction. By the virial theorem the total energy E, which is the sum of the thermal energy and the gravitational potential energy, is equal to half the gravitational potential energy, i.e.,

$$\left.\begin{aligned} E &= \text{thermal energy} + \text{grav. pot. energy} \\ &= \frac{1}{2}\text{grave. pot. energy} \end{aligned}\right\}. \tag{6.9}$$

Hence, we get,

$$\text{thermal energy} = -\frac{1}{2}\text{grave. pot. energy.} \tag{6.10}$$

The gravitational potential energy is always negative for a self-gravitating body. It would become more negative if a body of fixed mass were to contract. In such a contraction, Eq. (6.9) shows that only half the energy released gravitationally would remain in the star while the other half would presumably be released as radiation. Eq. (6.10) shows that the thermal energy would increase if the gravitational potential energy were to become more negative. This increase is understandable as classical gases usually get hotter when compressed.

A self-gravitating star differs from a cooling ember in a very important respect. A hot ember will radiate into its cooler surroundings, cool down, and arrive eventually at the same temperature as its surroundings. A hot star, without any source of energy other than its internal heat and self-gravitation, will also radiate into its surroundings. But instead of cooling down, it will contract and heat up. So a star made of a perfect classical gas tends to evolve toward states with a greater and greater temperature disparity with the rest of the Universe. This is illustrated in Fig. 6.7.

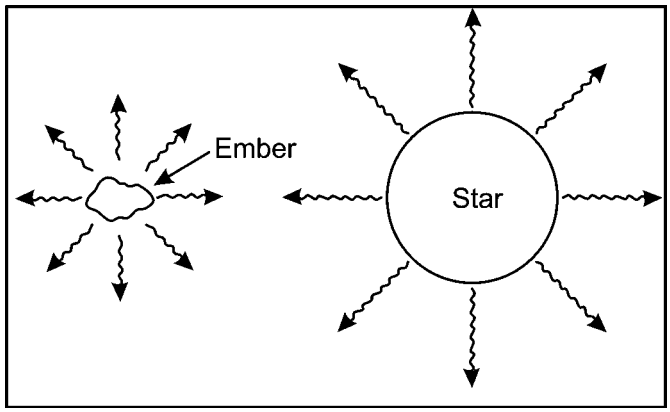


Figure 6.7: A self-gravitating star differs from a cooling ember in one very important respect.

The obstacle preventing the Sun from reaching thermodynamic equilibrium is thermonuclear reactions. Eq. (6.8) implies that only half the released gravitational energy is available for radiation, and this half is numerically equal to the present thermal-energy store. In fact, the virial theorem shows that only the combination of thermal energy and gravitational energy produces a net source for radiation in a quasi-statically contracting self-gravitating body.

At the central temperatures of about 15 million degrees Kelvin in the Sun, only hydrogen reactions can proceed near the core. The requirement that these reactions provide a stable *burning* lifetime of greater than 3 billion years implies that the Sun must be made mostly of hydrogen. The Sun's energy source gives the important clue that the predominant form of matter in the Universe is the simplest form of all possible elements, hydrogen.

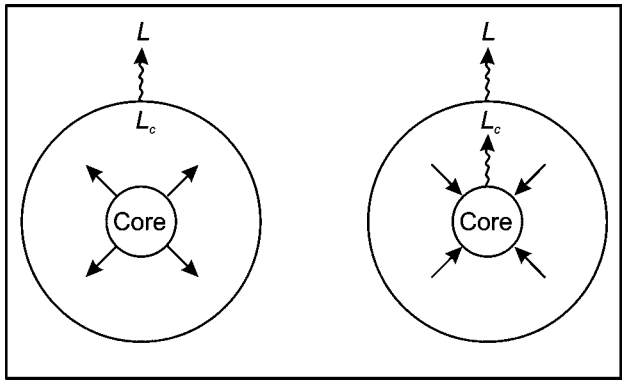


Figure 6.8: The Sun's safety valve for thermonuclear fusion.

6.4 CHROMOSPHERE AND CORONA OF THE SUN

The Sun has a perfectly sharp outer edge, coinciding approximately with the location of its photosphere. The total thickness of the photosphere occupies only 0.1 percent of the radius of the Sun. Therefore, the Sun looks as if it has a sharp edge, which explains the conventional definition of the photosphere of the Sun as its *surface*. There is gas belonging to the Sun, which lies above the photosphere, in the chromosphere and the corona. The chromosphere itself is a transition zone between the properties of the gas in the photosphere and those of the gas in the corona (Fig. 6.9). The gas temperature as we go outward from the photosphere initially decreases, and later it begins to climb dramatically, until temperatures

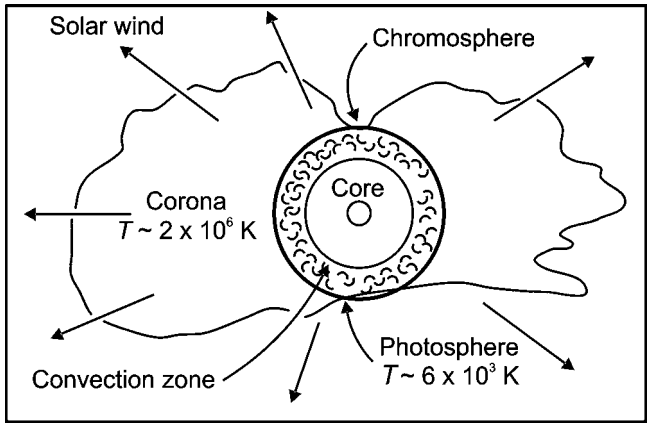


Figure 6.9: Outer layers of the Sun.

around 2 million degrees Kelvin are reached in the corona. This temperature is much higher than the photospheric temperature of around 5,800 K. However, the light output from the photosphere vastly dominates that from the corona. This is because the corona gas is so rarefied that it has little emissive power. The gas in the chromosphere and the corona is *optically thin*. For this reason, it is difficult to see the corona in optical light unless the light from the photosphere is blocked, either naturally by an eclipse of the Sun by the Moon, or artificially by a coronagraph.

When the light from the solar photosphere is blocked by a solar eclipse, light arises from the chromosphere wherein hot gas lies above cool gas. This situation gives rise to emission lines instead of absorption lines. This is the origin of the *flash spectrum* associated with emission lines in the chromosphere observed during solar eclipses. That element is helium. Helium is quite rare on Earth, but it is now known to be the second most abundant element in the Universe. Helium is the second simplest chemical element. This is a second important clue to the creation of the Universe. The 2 million degree temperature of the solar corona means that this gas can emit x-rays. This explains the presence of x-rays in the solar spectrum. X-ray photographs of the Sun look quite different from the usual optical pictures. More information about the structure and dynamics of the solar corona is obtained from satellite x-ray observations of the solar corona.

Astronomers believe that solar corona is heated by acoustic waves and other kinds of mechanical waves generated in the solar convection zone lying below the solar photosphere. These waves propagate upward into the more rarefied corona where they steepen into shock waves and dissipate their energy into heat. This heats the corona, which has very little heat capacity, to such high temperatures that the gravitational field of the Sun finds it hard to hold onto the gas in the outer parts of the solar corona. Theoretically, it was shown that these portions blow continuously from the Sun to form a *solar wind*, which pervades the solar system (Fig. 6.10). It is now possible to collect the particles in this wind by satellites, and to verify that the outer parts of the Sun are made primarily of hydrogen.

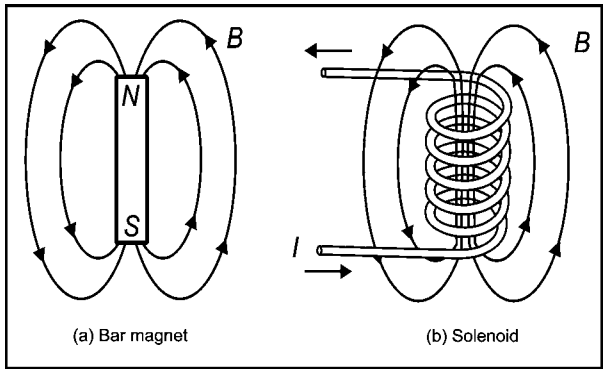


Figure 6.10: Comparison of the magnetic fields of a bar magnet and that produced by an electric current circulating through the coil of a wire.

6.5 MAGNETIC ACTIVITY IN THE SUN

In addition to the solar wind, the solar atmosphere also manifests other violent activity in the form of flares, loops, arches, etc. This activity may provide a substantial input of energy and particles into the expanding solar corona. The active regions are associated with the presence of magnetic fields in the surface layers of the Sun. Magnetic field activity can be studied qualitatively by observing sunspots, or by measuring the Zeeman effect splitting in atomic lines.

Fig. 6.10 makes comparison of the magnetic fields of a bar magnet and that produced by an electric current circulating through the coil of a wire. The magnetic field lines never end on an object. Magnetic field lines could end on or originate from an object only if particles called magnetic monopoles existed. Such particles have not been found in nature.

Fig. 6.11 shows a schematic energy versus a magnetization curve for a bar of ferromagnetic material. At high temperatures, the state of minimum total energy corresponds to zero magnetization. At low temperatures the state of zero magnetization is a local maximum of total energy. There are two minima of total energy, which correspond to spontaneous magnetization with the north pole of the magnetic dipole at the left or right end of the bar. The actual bar magnet will settle on one of these two states as denoted by dots. The phenomenon of spontaneous magnetization is another example of our general rule for the thermodynamic behavior of matter. At high temperatures, the state of least energy corresponds to more freedom for the atomic magnets to twist and turn. At low temperatures, ferromagnetic materials prefer to gain more magnetic binding energy by giving up some freedom for the atomic magnets to twist and turn.

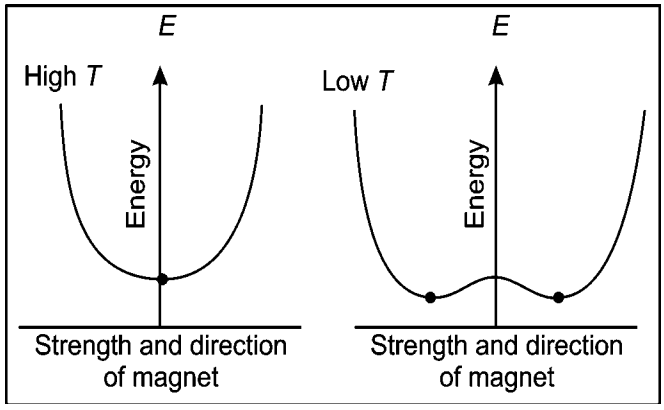


Figure 6.11: Schematic energy versus magnetization curve for a bar of ferromagnetic material.

Fig. 6.12 shows the magnetic-field structure of the Sun, as deduced from ground-based observations. As seen it is usually quite chaotic. The average strength of the magnetic field on the surface of the Sun is comparable to that on the surface of the Earth. In localized spots, however, the solar magnetic field can rise to thousands of times higher than the average value.

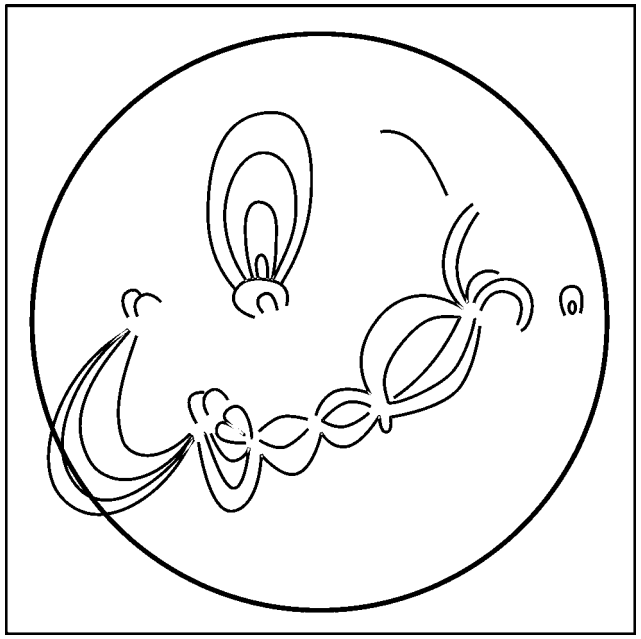


Figure 6.12: Magnetic field structure of the Sun, as deduced from ground-based observations.

6.6 PRINCIPLES OF STELLAR STRUCTURE

The present radius of the Sun has the right value to maintain a central temperature to provide a rate of nuclear energy generation that exactly balances the leakage rate due to the *random walk* of the photons from center to surface. A summary of the important concepts of stellar structure is given in Table 6.1.

Table 6.1: Summary of the principles of stellar structure.

1.	**Hydrostatic Equilibrium.** For a star to be mechanically in equilibrium, the pressure at every level must equal the weight of a column of material of the unit cross-sectional area on top.

2.	**Energy Transfer.** Photons in the interior of the star carry heat outward by random walking from regions of high temperature to regions of low temperature. At the photosphere of the star, the photons make a transition from *walking* to *flying*. If the luminosity required to be carried out is too large for radiative transfer to handle stably, convection results.
3.	**Energy Generation.** For a star to be thermally in equilibrium, the energy carried outwards by radiative diffusion and/or convection past any spherical surface must be balanced by an equal release of nuclear energy in the interior of that sphere. If the nuclear source is inadequate, gravitational contraction must result. In stars like the Sun, the size of the star has adjusted itself so that exactly as much energy is lost from the surface as is released by nuclear reactions in the interior.

6.7 THE CONVECTION ZONE OF THE SUN

Fig. 6.13 shows schematically the structure of the Sun from core to photosphere as understood from detailed modeling. The solar core is defined as that region

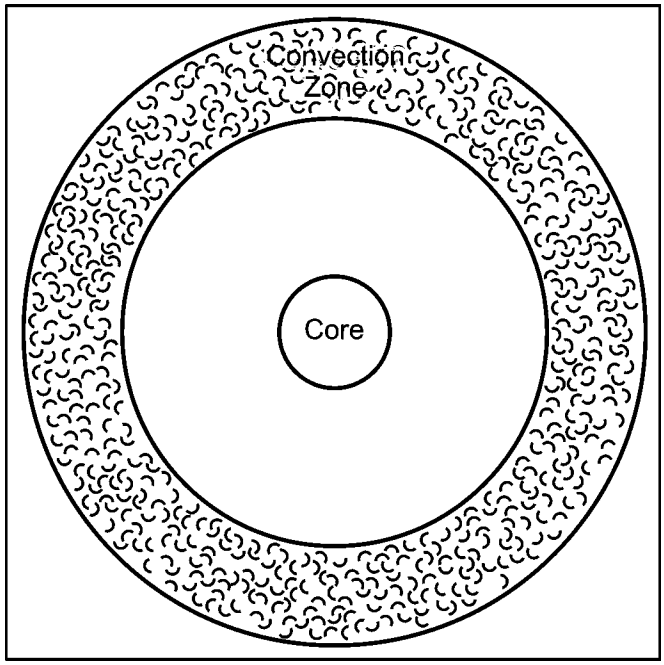


Figure 6.13: The interior structure of the Sun.

where the temperatures are high enough to yield appreciable rates of hydrogen fusion into helium. The solar photosphere is defined as the outer layers where the photons stop *walking* (i.e., *optically thick*) and begin *flying* (i.e., *optically thin*). In the sub-photospheric layers lies the solar convection zone. The Sun generates nuclear energy in its core at a rate equal to $L_\Theta = 3.9 \times 10^{33}$ ergs/sec. This luminosity is ultimately radiated from the photosphere according to the formula

$$L = 4 \pi R^2 \sigma T_e^4. \tag{6.11}$$

Here, L is L_Θ and R is R_Θ for the Sun.

EXERCISES

1. Define the solar constant. What is the value of the solar constant obtained from direct measurement?
2. How can you determine the effective temperature of the Sun?
3. What is the origin of "limb darkening" of the Sun? Explain with a suitable diagram.
4. Explain the formation of a dark line in a stellar photosphere.
5. How were the conditions in the deep interior of the Sun discovered?
6. How are the mass, mean density and pressure inside the Sun estimated?
7. What do you mean by solar plasma? How can the solar plasma be considered as a perfect gas?
8. Explain how the Sun's safety valve for thermonuclear fusion works.
9. "There is a net slow effect performed by photons inside stars like the Sun." Considering the dance of the photons as "random walk," explain the statement.
10. How does a self-gravitating star differ from a cooling ember?
11. Discuss elaborately the two outer layers- 'Chromosphere' and 'Corona' of the Sun.
12. Critically discuss the magnetic field structure of the sun as deduced from ground-based observations.
13. Compare the magnetic fields of a bar magnet and that produced by an electric current circulating through the coil of a wire.
14. Give the summary of the important concepts of "stellar."
15. Show schematically the structure of the Sun from core to photosphere, as understood from detailed modeling and discuss.
16. A typical amateur's telescope is given to you. It has a 1 meter objective focal length and a 40 mm eyepiece. What would be its magnification? Does the magnification apply only to observation of double stars?
17. Find the diffraction limit in sec of an arc of a 5-meter telescope corresponding to a wavelength of 5000 Å. Compare the calculated value to the seeing limit.

Chapter 7

ASTRONOMICAL TELESCOPES

7.1 INTRODUCTION

Astronomical telescopes can be constructed using either reflection or refraction to gather light over a large collecting surface and thereby to focus it to a much smaller area. This property of gathering more light and focusing it, not magnification, is the primary purpose of most astronomical telescopes.

7.2 REFLECTION

If a thin layer of metal is deposited onto a glass surface and is polished, a mirror is formed. The reflecting property of the mirror is a result of the microscopic fact that metals contain many free electrons, which are not bound to the atoms in the lattice of the metal. These free electrons respond readily to the presence of electromagnetic waves. Also, they allow almost no radiation of sufficiently long wavelengths to pass through the metal. A mirror, therefore, reflects almost all the optical light that falls on its surface. The geometric properties of a flat mirror can be deduced from the knowledge of simple wave properties of light. Light reflects from a flat surface in such a way that the angle of reflectance is equal to the angle

of incidence. What fraction of the light is reflected or absorbed or transmitted depends on the kind of metal used in the mirror and also on the wavelength of the incident light (e.g., radio, optical, or x-rays). The incident and reflection angles are always symmetric, as the wave propagation before and after reflection take place in the same medium.

7.3 REFLECTING TELESCOPE

The high reflectance of metallic surfaces permits reflecting telescopes to be used with waves of any length from radio to ultraviolet. All these are constructed on the same general principle. The basic principle of optical and radio telescopes is explained in Fig. 7.1. To get good imaging, the surface of the reflecting element should be paraboloidal (i.e., a parabola of revolution and the surface should not contain any irregularities larger than a fraction of the wavelength of the light that needs to focus). As the visual light is shorter than radio waves, optical telescopes are made by depositing a metal film on accurately figured glass. On the contrary, much coarser techniques of directly shaping a metal dish that has small holes in it to make it lighter can be used to construct a radio telescope.

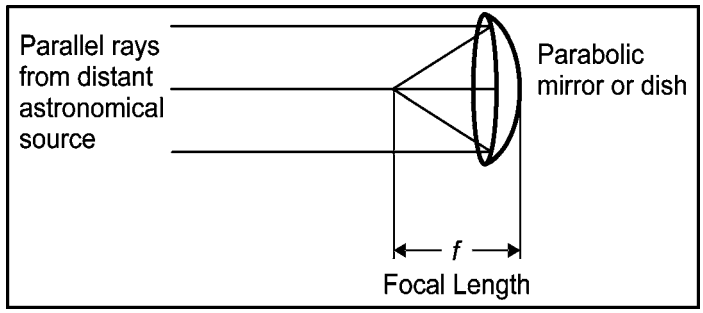


Figure 7.1: Principle behind optical and radio telescopes.

To understand why the mirror or the dish has to be paraboloidal we imagine a mirror in the shape of an ellipse and put a point source of light at one of the foci (Fig. 7.2). Light, which leaves this source, will strike the ellipse somewhere and reflect off it, with the angle of reflectance equal to the angle of incidence. This means that each ray will be brought to the other focus. So to collect all the light that leaves the source we should make our mirror an ellipse of revolution. Such an ellipsoid would represent the ultimate mirror in terms of light-gathering power for the assumed geometry. If only part of the ellipsoidal mirror is constructed as shown in Fig. 7.3 it will collect only part of the light rays that leave a point source.

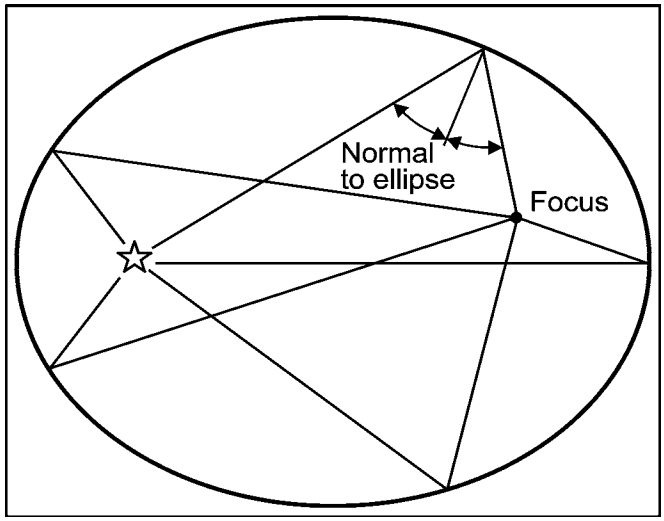


Figure 7.2: An elliptical mirror brings the light from a point source at one focus to the other focus.

Let us now assume that the point source of light is at infinity. If the two foci of an ellipse become infinitely separated the ellipse becomes a parabola. Therefore, to bring the light rays from a very distant point source like a star to a single focus, the objective of a reflecting telescope should be a paraboloidal mirror. To collect as much light as possible, we know that the aperture of the paraboloid should be made large.

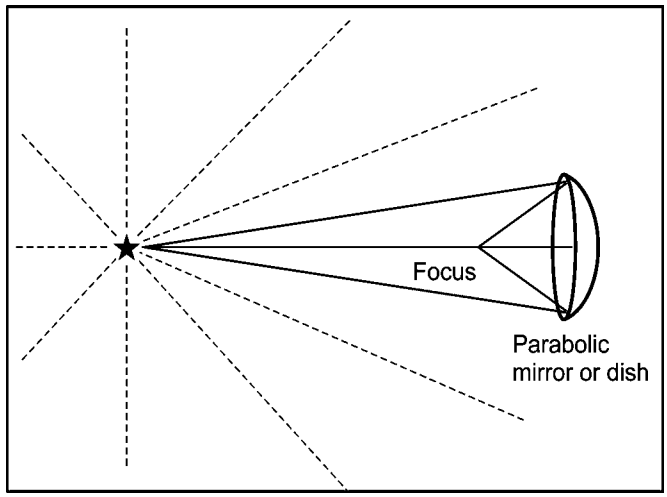


Figure 7.3: Part of a complete ellipsoidal mirror will capture only part of the light that leaves a point source.

X-rays are more penetrating than electromagnetic radiation of longer wavelengths. But x-rays incident almost perpendicularly on a metallic surface will be mostly absorbed rather than reflected. This problem is less severe at grazing incidences, so x-ray telescopes are built to bring the rays to a focus by grazing incidence reflections (Fig. 7.4). This configuration results in less collecting area for a given metallic surface. As the telescope has to place above the Earth's atmosphere, x-ray astronomy is a more expensive proposition than optical or radio astronomy.

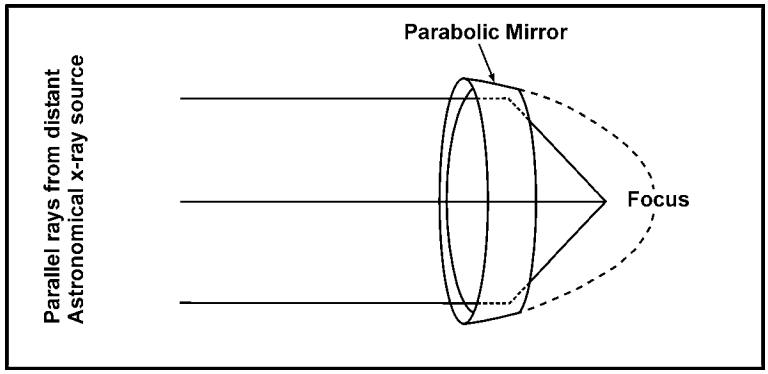


Figure 7.4: An x-ray telescope, which works by grazing incidence reflections. The focus is brought to the back of the mirror arrangement.

One disadvantage of a reflector is that its prime focus is in front of the mirror. Consequently, for an astronomer to observe at the prime focus, an appreciable fraction of the incoming light may be blocked (Fig. 7.5 *a*). This problem can, however, be partially solved in several ways. The first solution was invented by Newton. In that arrangement (Fig. 7.5 *b*) a secondary mirror is used to focus the light at the side of the telescope tube. This arrangement tends to make the system lopsided. A more satisfactory solution was devised by Cassegrain. In the Cassegrain arrangement (Fig. 7.5 *c*) a secondary mirror is used to focus the light behind the telescope tube through a hole in the primary mirror. However, none of these arrangements are very suitable for making precise spectroscopic observations. This is because the bulky and heavy spectrographs would strain on the telescope tube if they were suspended in front, to the side, or in the back of the primary mirror. For such studies the Coude arrangement was devised where a series of secondary mirrors bring the light down the polar axis of the telescope drive to a fixed focus in a room below the telescope (Fig. 7.5 *d*). In the Coude room, large pieces of equipment can be mounted properly in a stationary configuration.

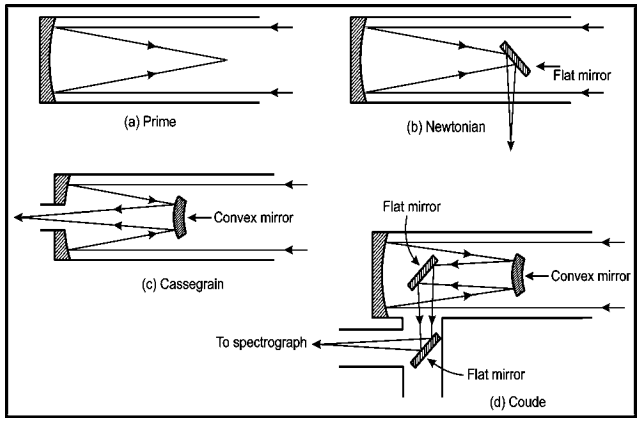


Figure 7.5: Various arrangements for reflecting telescopes: (a) The prime focus in front of the primary mirror. (b) The Newtonian focus at the side of the telescope. (c) The Cassegrain focus at the back of the telescope. (d) The Coude focus underneath the telescope.

7.4 REFRACTION

We know that the speed of light in air v_a is greater than the speed of light in glass v_g. A whole collection of photons are incident as a plane wave on an air-glass interface at an oblique angle (Fig. 7.6). The photons are imagined to come from a distant source so that they arrive in waves whose fronts are connected by the dashed lines. The initial direction of the propagation of the photons and the wave-front is represented by the heavy top arrow. As the photons pass through the glass interior, they are effectively slowed down and travel less distance in a given interval of time. The net effect is to *turn* the direction of the wave-front, so that the propagation of the wave-front in the glass is in a direction closer to the normal. This *turning* is a property of the entire collection of photons. An individual photon cannot be suddenly *turned* in direction upon entering the glass.

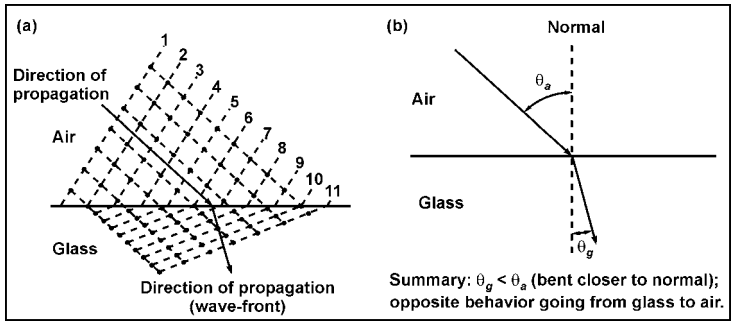


Figure 7.6: Refraction of waves.

In Fig. 7.6 *a* an air-glass interface is drawn as if it were an infinite flat plane, while in Fig. 7.6 *b* we draw only the direction of the propagation of the wave-fronts as being rays of light.

7.4.1 Principle of the Prism

In matter, light of different wavelengths traverses at different speeds. Thus, violet light will generally interact more with matter than blue light, and blue light more than red. This means that refraction in glass will turn violet light more than red light. Fig. 7.7 explains how a prism breaks up white light into its component colors.

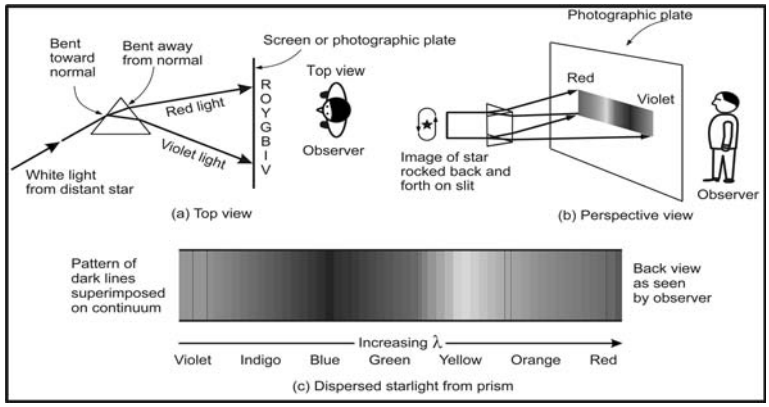


Figure 7.7: Principle of the prism.

To analyze the wavelength composition of starlight, astronomers use the *dispersive* property of prisms. Fig. 7.7 *b* shows a set up of how this is done. To use as much as possible of the photographic plate, the image of a star is rocked back and forth mechanically along the length of an entrance slit. The light that enters the slit is dispersed by a prism. For high accuracy a diffraction grating is used instead of a prism. The dispersed light is then allowed to fall onto a photographic plate. The exposed photographic plate will then show violet light on the left and red light on the right (Fig. 7.7 *c*), a sequence of increasing wavelength readings from left to right. With real starlight we have an absorption-line spectrum. An example is the Fraunhofer spectrum of the Sun. The spectrum reveals *lines* at certain wavelengths, and not *dots*. An extended object like the Sun produces lines without mechanical rocking. The patterns of dark lines are associated with the presence of various chemical elements in the atmosphere of the star. In this way spectral analysis can provide information on the chemical makeup of stars and other astronomical bodies.

7.5 REFRACTING TELESCOPE

In a refracting telescope the main optical element is the objective lens. It collects light over the entire area of the aperture and forms an image of the observed object at the focus. In order to discuss the focusing ability of a lens we imagine a source like a star, which is a single point of light. Since the star is at infinity, the rays from the star arrive at Earth nearly on parallel lines. A perfect lens would bring these parallel rays to a single point at the focus. This point is the correct image for our idealized star. In practice, starlight is made up of light of different colors. Since blue light is bent more by a glass lens than red light, chromatic aberration results (Fig. 7.8). A photographic plate placed at the blue focus will exhibit a blurred red image, and vice versa at the red focus. We ignore this chromatic aberration to discuss

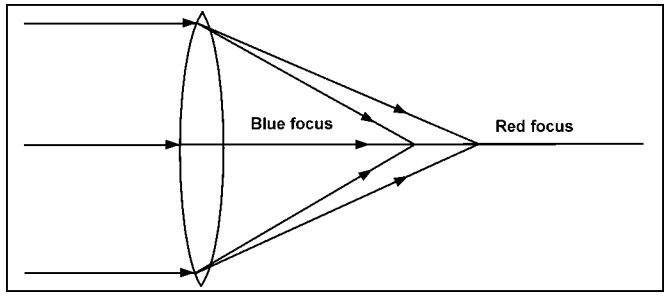


Figure 7.8: Chromatic aberration.

light-gathering power. A lens can be used in *reverse* of the above description; i.e., if an object (or the image of a real object) were placed at the focus of a lens, the rays from it, in passing through the lens, would leave in parallel lines (Fig. 7.9). A human eyeball placed behind this lens would perceive the object or the image as if it were at infinity. Thus, the combination of a large objective lens plus an eyepiece (Fig. 7.10) serves as an astronomical telescope (i.e., refractors) and this enables an astronomer to gather starlight over a much larger area than his or

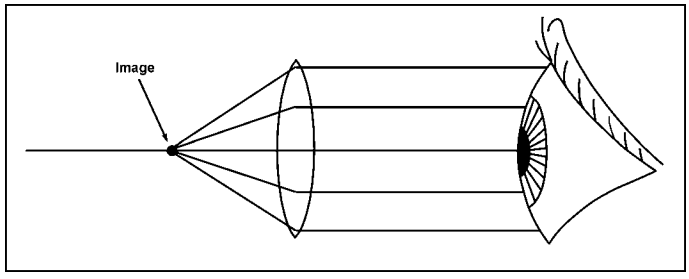


Figure 7.9: Principle of the eyepiece.

her own eyeball could. Electronic devices (i.e., *image intensifiers*) with or without photographic film allow the astronomer to *observe* to much fainter limits than is possible with naked-eye observations, even on telescopes with very large apertures.

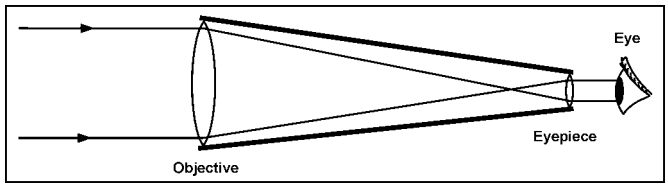


Figure 7.10: Sample refracting telescope.

Different apparent sizes of the final image, for example, the apparent separation of a double star result from using different eyepieces. The light-gathering power is found by the square of the diameter of the objective, not by the eyepiece. For an extended object, spreading the same amount of light over a larger area decreases the apparent surface brightness of the observed object. Thus, high magnification may be useful for solar-system observations, but it is not required for other types of astronomical investigations.

We assume that there is another star lying at an angle θ_o with respect to the first. The rays from both the stars will pass through the center of the objective lens on a straight line. These rays will be brought into focus at a distance equal to its focal length f_o behind the objective. If the eyepiece is placed at its focal length f_e further back, the double-star image will pass through the eyepiece and be collimated into two parallel beams, at an angle θ_e, to each other. Behind the eyepiece an astronomer will perceive an inverted image of two stars separated by an angle θ_e, whereas the actual double star is only separated by the angle θ_o. The angular separation thus has been magnified by the ratio θ_e/θ_o. To calculate this magnification in terms of the properties of the telescope, as indicated in Fig. 7.11, we may consider the relation

$$f_o \tan \theta_o = f_e \tan \theta_e. \tag{7.1}$$

In general, the angles θ_o and θ_e are very small. A typical amateur's telescope may have a one-meter objective focal length and a 40-millimeter eyepiece.

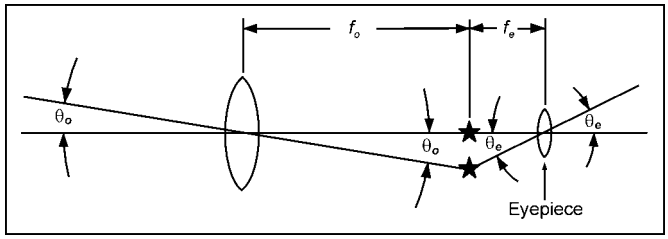


Figure 7.11: Basic property of the telescope.

7.6 ANGULAR RESOLUTION

An important advantage of a telescope is the ability to distinguish fine details in the image. This ability is known as *angular resolution*. Fig. 7.12 illustrates the angular resolution concept. We imagine a telescope pointing directly toward a star on its optical axis. Another star is lying at an angle θ with respect to the first. A perfect focus would result in two point images displayed slightly from one another. However, such a perfect focus is never attainable in practice. Ground-based optical telescopes contend with the blurring produced by turbulence in the Earth's atmosphere. The varying amount of refraction in a turbulent atmosphere makes the focused optical image dance around *seeing* on the photographic plate. If the source is bright, astronomers can use a series of short exposures. For faint objects, this technique is impractical and atmospheric turbulence limits the seeing image of a point source to about one second of arc even at the best sites. Thus, if two stars are separated by less than one second of arc the observer cannot tell from the blurred images that more than one point of light is really present.

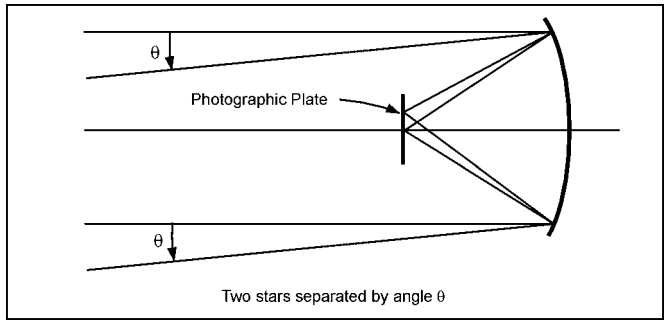


Figure 7.12: Concept of angular resolution applied to a double star.

Radio waves do not suffer from this atmospheric blurring. The angular resolution of modern radio telescopes is limited by the diffraction produced by the wave nature of light. We know that the ray-tracing method to determine the focusing of electromagnetic waves assumes that the wavelength is too small compared with the aperture of the telescope. However, on angular scales comparable to the ratio of the wavelength λ of light to the diameter D of the telescopes collecting surface, diffraction effects produce an unavoidable blurring even of point sources. This is due to the fact that the secondary waves produced by the interaction of light and matter at the edges of the objective interfere to give a pattern of light and dark fringes which, on small angular scales, is different from that predicted by ray tracing. For a circular aperture, the blurring produced by diffraction limits the angular resolution by an amount

$$\theta_{\text{diffraction limit}} = 1.22\,\lambda/D. \tag{7.2}$$

The fully steerable single-dish telescopes of the 100-meter dish in West Germany thus cannot resolve two radio sources closer than about 9 minutes of arc when it operates at a wavelength of 21 cm.

There are two direct remedies to improve the angular resolution of astronomical telescopes. For optical work, improvement is possible by getting above the Earth's atmosphere (i.e., above the *seeing*). This is one principal motivation for the Space Telescope project. For radio work, the computer *recombination* of the signals from more than one radio telescope can increase the effective baseline D of the observations. It thus reduces the ratio of λ/D, which governs the diffraction limit. A technique would be merely to build a huge array of telescopes as that proposed for project Cyclops. A more elegant solution uses the rotation of the Earth. In the simplest case, a line of telescopes may be placed in a manner as shown in Fig. 7.13. During a 12-hour period, the rotation of the Earth would swing the line of radio telescopes through a full circle. It thus allows mimicking the resolving power of a vast single-dish with the corresponding aperture diameter. Such powerful full-synthesis instruments have been built in England and in Holland. A more powerful synthesis instrument is the Very Large Array (VLA) in New Mexico, whose movable telescopes are so arranged that they need not rely on the rotation of the Earth to nearly get the complete aperture coverage.

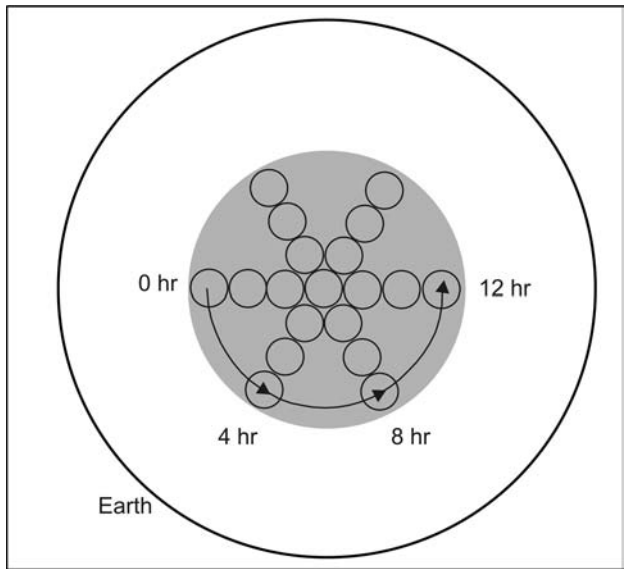


Figure 7.13: Principle of aperture synthesis illustrated by an imaginary line of radio telescopes at the North Pole. During one 12-hour period, the rotation of the Earth causes this line to sweep out in a full circle.

EXERCISES

1. What is the general principle of constructing astronomical telescopes?
2. Give the basic principle of optical and radio telescopes. Why does the mirror on the dish have to be paraboloidal?
3. Which has more penetrating power: x-rays or electromagnetic radiation of longer wavelengths?
4. What are the major advantages of reflectors over refractors? What is the main disadvantage of a reflector?
5. Because the prime focus of a reflector is in front of the mirror, an appreciable fraction of the incoming light may be blocked. How is this problem solved?
6. Draw the optical diagrams of various arrangements for reflecting telescopes.
7. What is the dispersive property of a prism?
8. Explain the principle of a refracting telescope.
9. What is a chromatic aberration? What is an eyepiece?
10. What is meant by the angular resolution of a telescope?
11. How is the concept of angular resolution applied to a double star?
12. How is the angular resolution of astronomical telescopes improved?
13. Illustrate the principle of aperture synthesis to an imaginary line of radio telescopes.

Chapter 8

How to Observe the Sun

8.1 INTRODUCTION

There are some primary reasons for observing the Sun (e.g., the Sun is a star closest to us, it is a variable star, and we are more or less dependent on its behavior). This chapter provides the important facts about the behavior of the Sun and describes how interesting observations may be made. This is covered in two sub-sections: (*i*) observation of the Sun in integrated (white) light and (*ii*) observation in monochromatic light (primarily H_α). The discussion includes certain general features such as sunspots, groups of spots, and photospheric activity, flares, prominences, and chromospheric activity. Specific observational techniques suitable for individual or group programmers are described, which may be undertaken for the observer's personal pleasure, or with scientific purposes in mind.

The Sun is a star that varies over the course of time. It is the closest star, and the light we receive from it was emitted only 8 minutes earlier. The largest modern telescopes can detect light from distant celestial objects, which was emitted thousands of million of years ago. The fact that the Sun is so near provides us with a considerable amount of light, enabling us to make detailed measurements with short exposures without having to use large apertures. It is particularly favorable for amateur observation. In fact, it is useful to have a long focal length because

the diameter of the primary solar image is about 1/100th of an objective's focal length.

For the Sun, it is the quality of the observing site that ultimately determines the largest aperture to be used. We have learned that solar observations may be carried out with small diameter telescopes and that 150 to 200 mm is adequate. The main problem is to reject the excess light, and thereby to reduce the heating of the interior of the instrument and to protect the eyes of the observer.

8.2 OBSERVATION OF THE SUN AND ITS DANGERS

Even a small amateur telescope may be used to examine the surface of our daytime star and to study some of its features. But one should know the dangers posed by direct observation of the Sun. Let us summarize the precautions to be taken to avoid any possibility of damage to the eyes.

Permanent changes to the retina due to the unprotected observation of the Sun, particularly during eclipses, have been analyzed by Quintel for amateur astronomers. When an observation is taken using any magnifying optical system, the amount of light entering the eye is proportional to the surface area of the instrument's objective. For a luminous body like the Sun, the flux received on the retina is considerably increased by the optical system, but the eye can protect itself particularly by reducing the diameter of the iris to a minimum, which is about 1.5 mm for most subjects. If the light is not too strong, one is only dazzled, and no lesion occurs. As soon as the bright light is removed, the eye regains its sense of vision after a few minutes. On the other hand, if the light is intense, the amount of heat absorbed becomes more than the cooling capacity of the blood supply to the retina can handle. There will be an increase of temperature of the cells, leading to definite damage, but the person will not feel any sensation of pain that might serve as a warning. The severity of the burns to the retina depends on the flux received and the length of the exposure. It is probable that a weak exposure without sufficient protection may cause the death of some cells in the center of the retina. It is therefore essential to ensure that the eyes are properly protected even if it seems unlikely that one will be dazzled. The fact of being dazzled, which leads to a decrease in visual response, may, paradoxically, give the impression that the observation is comfortable.

Therefore, one should never put his or her eye to the eyepiece before having checked to ensure that a suitable filter or other device has been fitted to the instrument. The second is to cover the objective of the finder to avoid the risk of burns. In addition one should be wary of the solar filters provided with most commercial telescopes, which are located at the eyepiece end, and absorb the light

at the point where it is closest to the focus of the objective. Such a filter should always be used behind an objective filter. Otherwise, it may shatter in a few seconds through the intense heating to which it is subject. The whole of the luminous flux may suddenly reach the retina and cause serious damage before the observer has time to move away from the eyepiece. Moreover, it is to be ensured that the filter absorbs over a sufficiently wide spectral band. Some filters pass a considerable amount of infra-red or near-ultraviolet radiation, which is invisible but may still cause burning of the retina.

8.3 METHODS OF OBSERVING THE SUN

Here, we describe three methods of observing the Sun: direct visual observation, observation by projection, and photographic observation, which are explained below.

8.3.1 Direct Visual Observation

An amateur often observes the Sun through an eyepiece, after having fitted suitable devices such as filters for reducing the amount of light. Or by using a Herschel wedge to get a reflection from a non-aluminized surface, etc. This is perhaps the best way to get a spectacular view of the surface of the Sun. However, the method is not recommended for two reasons.

(*i*) Unless observers have taken all the precautions mentioned above they do run the risks.

(*ii*) The scientific importance of the observations is reduced. A drawing of a group of spots made directly at the eyepiece will not have the same accuracy as a photograph or even as a sketch made by projection.

8.3.2 Observation by Projection

In this method the telescope is directed toward the Sun by watching the shadow cast by the body of the instrument. When the shadow is reduced to a circle, the image of the Sun should be within the held view of the telescope. It is then required to hold a cardboard at about 20 cm from the eyepiece. By changing the focus of the eyepiece gradually a sharp image of the solar disk is obtained on the screen. Instrument makers usually provide a collar and a set of rods that can carry a screen for observation by projection. If the telescope has no drive the image drifts rapidly. In that case frequent adjustment of the instrument is necessary during observation. With a little practice an observer should be able to draw spots in short time. There is an alternative way, but it requires a larger budget. This consists of photographing

the solar image formed on a projection screen. The speed of this process, when the setting is carried out in advance, means that drift may be eliminated and a drive is no longer indispensable. This method demands considerable care in aligning and setting the telescope properly to get a truly circular image. The method is extremely sensitive to atmospheric turbulence.

8.3.3 Photographic Observation

The best results are no doubt obtained by photographic techniques. Direct photography of the solar image at the primary focus of the instrument is not wise because of possible damage to the camera's shutter. To get a proper photograph and to ensure that observation is both comfortable and safe, it is necessary to devise an arrangement that only passes a small fraction of the incident light. The first idea is to use a filter. For both refractors and reflectors, two positions are accessible in front of the objective and in front of the focal plane.

The first solution requires a neutral-density filter with plane-parallel races that is at least as large as the objective. It has a higher degree of flatness and truly parallel faces and is very homogeneous. Reduction by reflection is always preferable to reduction by absorption. This is because it avoids heating the filter and the surrounding layers of air. Observers also use interchangeable filters for alignment and focusing and then for taking the photograph. Such filters consist of a metal-coated Mylar film in a rigid holder that fits over the objective. A reduction factor of 1000 is generally used for photographic work in combination with a shutter speed of $1/500^{th}$ to $1/1000^{th}$ of a second. For visual observation, the reduction should be of the order of 1,000,000 if one wishes to avoid the dangers of being dazzled. The second solution (i.e., a filter near the focal plane) permits all the luminous flux to enter the instrument. Most of this energy is rejected by a smaller surface, lying in the converging cone of light at a point close to the focus. A commercial filter is available that only transmits 3% of the incident light (density 1.5). This consists of an aluminum film deposited on a sheet of silica glass. A diameter of only 30 mm suffices for such a filter. Additional reduction is essential either by using a neutral-density filter or by a welding glass. A Herschel wedge may also be used with one glass-air reflection, or else a penta-prism with two glass-air reflections, with either a neutral filter or one with a color chosen especially for photography.

High contrast is not the most important factor in solar photography. A too contrast negative is unsuitable for certain scientific purposes. The solar observer likes good production of the mid-tones and detail in both the flight and the 'dark' areas. Both the spots and the solar limb should be clearly visible.

In direct photography of the surface of the Sun over-exposure is a problem. After finding a suitable method of reducing the amount of light and using the fastest shutter setting ($1/500^{th}$ or $1/1000^{th}$ of a second), the observer needs to experiment to get the most suitable combination of colored filter and photographic emulsion. Choose a slow-speed film, which has fine grain and high contrast. Ground and polished glass color-filters are preferable to gelatin or plastic ones. These soon degrade with the heat in the beam of sunlight. In an arrangement where most of the light is rejected just in front of the focus, these filters must be placed after the device so that they reject most of the light. As a result, they will not be subject to great heat and will retain their optical properties. In order to choose a filter an idea of the spectral-sensitivity curve of the emulsion is required. Current fine-grain emulsions are all panchromatic but their sensitivity generally drops beyond about 650 nm. By using a red filter with a wavelength longer than 600 nm, it is possible to create a system with a pass-band about 45 nm wide, centered on a wavelength of about 625 nm. Fig. 8.1 gives an example of a combination achieved by using Kodak Panatomic-X 5060 35-mm film and a Schott RG 610 filter (3 mm thick). There is an additional benefit because the Earth's atmosphere is less refractive in the red and SO atmospheric effects are correspondingly reduced. This improves the quality and the resolution of the images.

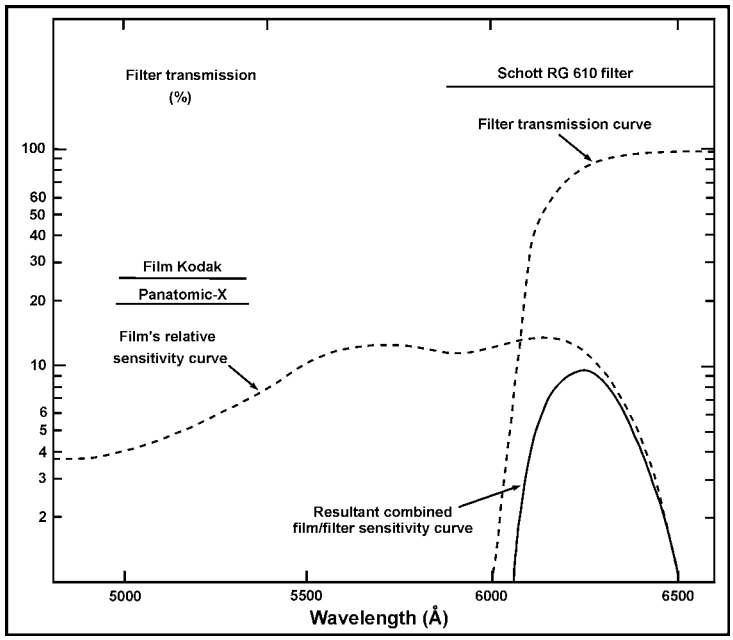


Figure 8.1: An example of a narrow bandwidth obtained by combining the transmission curves of the filter (a Schott RG 610) and the spectral sensitivity of the film (Kodak Panatomic-X).

8.4 THE ORIENTATION OF DRAWINGS AND PHOTOGRAPHS

8.4.1 The Determination of Geographical North and South

We assume that the observer is using the photographic method. An initial exposure is made with a duration that gives a correctly exposed image. This is marked 1 in Fig. 8.2. The position of the instrument is clamped and a second exposure is made 90 seconds later, on the same frame. Due to the diurnal rotation, the second image of the solar disk will have been displaced by about 1.5 times its radius from the first one. To indicate the direction of motion, the second exposure is made with a duration of 1/2 or 1/3. The under-exposed image will then be on the West. After enlarging the picture the straight line NS is then drawn which passes through the points where the two circular images intersect. The precision of this method is of the order of 1/2 degree. If drawings are to be made from a solar image projected onto a screen the East/West direction may be obtained by clamping the telescope and marking the position of a small spot at two different times as the image drifts across the field.

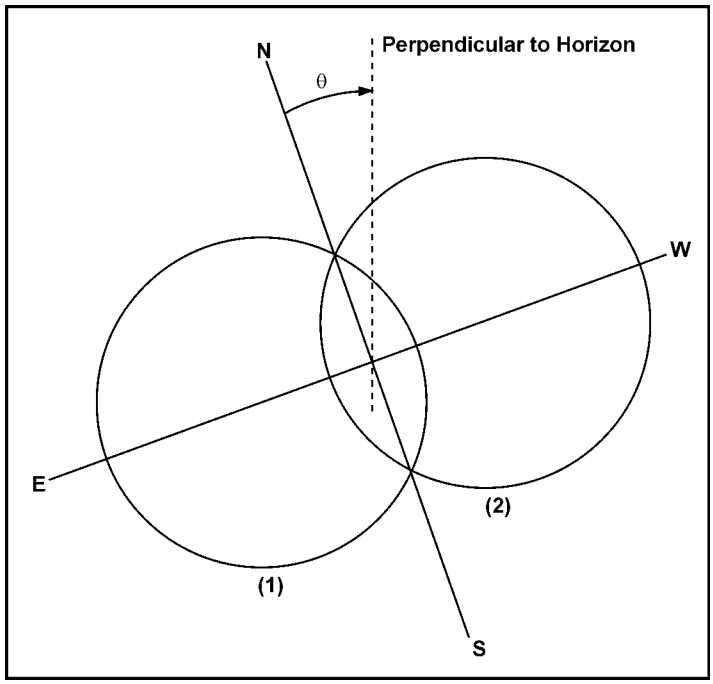


Figure 8.2: Determining the North/South geographical axis of an observation.

8.4.2 Mounting and Orientation of the Image

If an equatorially mounted telescope is used the motion of all the objects in the sky are then strictly related to a specific line in space, the North/South axis. Consequently, the telegraphical North/South orientation of the field of the instrument remains unaltered. The orientation procedure is to be carried out once each time the mounting is altered. With an Altazimuth telescope local vertical serves as a reference for the diurnal motion. The direction of geographical North varies with respect to the vertical over the course of a day and the field appears to rotate.

In order to calculate the field rotation, we consider Fig. 8.3. In the figure the latitude of the observing site, δ = declination of the Sun on the day of observation, H = Hour Angle of the Sun at the time of observation, N_P = the North Celestial Pole, Z = the Zenith, and NS = projection of the meridian on the plane of the horizon.

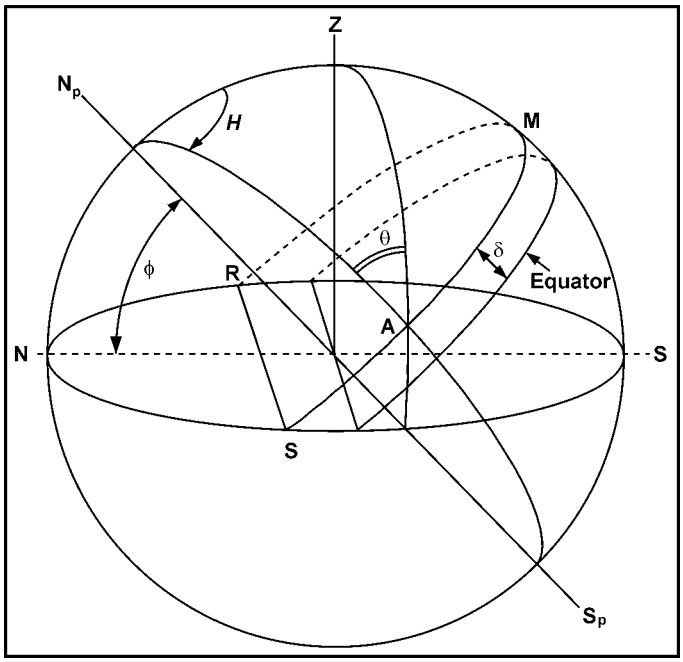


Figure 8.3: Defining the events to find the orientation of the observation of an observation made with an Altazimuth instrument.

The path followed by the Sun across the sky during the day is the circular arc R.M.S. If the angle is between the geographical North/South direction (i.e., the meridian N_PAS_P) and the vertical then it can be shown that

$$\tan\theta = \sin H\,(\cos\delta\tan\phi - \sin\delta\cos H)^{-1}. \qquad (8.1)$$

Fig. 8.4 shows how the North/South line rotates with respect to the vertical during a day. The plane of the diagram is the plane tangent to the celestial sphere at the center of the solar disk. The rotation of the North/South line takes place in the same direction as the hands of a clock.

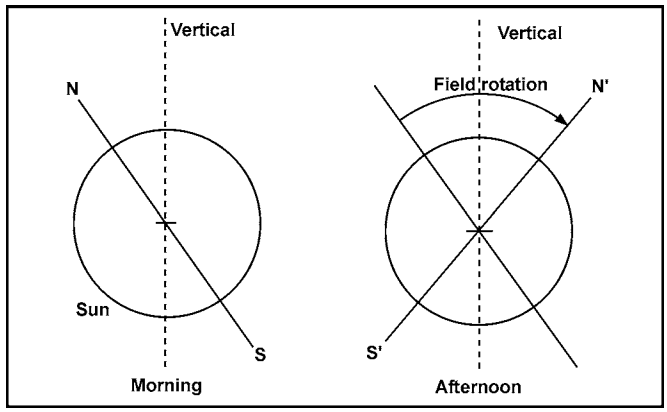


Figure 8.4: Variation in the geographical North/South direction as a function of time of day for an Altazimuth instrument.

8.4.3 Position of the Sun's North/South Axis

The angle between the Sun's North/South axis as indicated by NP-SP in Fig. 8.5 is given as P in ephemeredes for the physical observation of the Sun. This angle is measured positively in the trigonometrical sense. The diagram shows the two extreme positions (P = 26.32) of the NP-SP axis during 1987.

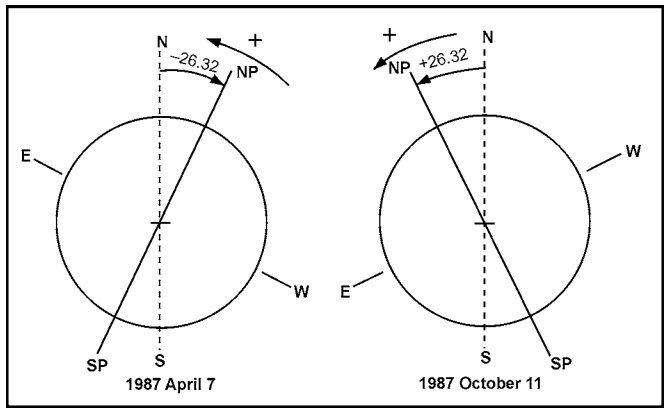


Figure 8.5: Determination of the direction of the Sun's axis of rotation with respect to geographical North/South line.

8.4.4 Measurement of the Positions of Solar Features

Accurate measurement of positions may be obtained to about 0.50 either by reading them directly from a coordinate grid or by calculation. In both cases the accuracy depends on the quality of the images.

8.4.5 Identification of a Feature at the Limb

The concept above is shown in Fig. 8.6. Only one value is to be determined. This is the position angle (marked PA) measured positively in the trigonometric sense from 0 to 360°, where geographical North (N) is taken as the origin and the feature in question (F) as the extent.

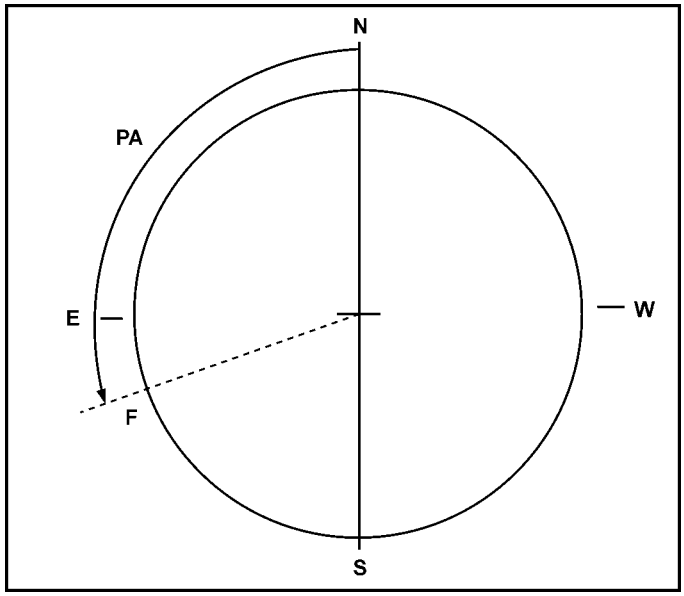


Figure 8.6: Convention for referring to objects at the solar limb.

8.5 AMATEUR-OBSERVING PROGRAMS

After showing the technical problems curiosity leads observers to watching sunspots. They soon start to count the spots each day, trying to recognize them and follow their changes. This gives rise to two specific programs: (*i*) a program for deriving the Wolf Number and (*ii*) one for obtaining the active area.

8.5.1 Counting Sunspots at the Wolf Number

On an image of the whole disk, found either by projection onto a white screen or by a photographic enlargement, sunspot groups may be identified. The relative Wolf Number is an arbitrary number R obtained by the equation

$$R = k(10g + s). \tag{8.2}$$

Here, k is a coefficient taking account of the observer, the quality of the image, and of the instrument; g is the number of spot groups; and s is the number of individual spots. The constant k is required in standardizing observations in any collective program but for individual work it may be ignored.

8.5.2 Calculation of the Active Area

This program can also examine the variations in one of the factors related to sunspots. Although in this method the measurement of the apparent area of spots is relatively easy, measurement of the actual surface area needs determination of the heliographic coordinates of each spot. As a matter of fact a correlation may be applied to take account of the foreshortening caused by the Sun being spherical. If θ is the angle subtended at the center of the Sun by the spot and the center of the disk, then the correction may be expressed as actual surface area = apparent surface area/cos θ. It is obvious that errors measuring the position or apparent area of spots close to the limb may cause major inaccuracies. The area of sunspots is generally expressed in millionths of the Sun's visible hemisphere (10^6), or in millionths of the total solar surface area.

8.5.3 Measurement of the Proper Motion of Spots

During the lifetime of active areas, sunspots move in both longitude and latitude across the surface of the Sun. The motion in longitude is usually greater than the motion in latitude. The apparent motion of any spot results from the combination of the Sun's differential rotation at the spot's mean latitude and its own proper motion.

8.6 MONOCHROMATIC OBSERVATION OF THE SUN

8.6.1 The Solar Spectrum

The Sun emits radiation at all wavelengths (λ) of the electromagnetic spectrum, but from the ground we have access through two windows only: the optical window in

the visible region and the radio window at longer wavelengths. These two spectral regions include the wavelengths where most of the energy is radiated from the photosphere into the lower corona. The chromosphere is almost transparent and is visible against the disk. After the light has passed through the photosphere and chromosphere, its spectrum exhibits a large number of absorption lines called the *Fraunhofer spectrum* having highly variable widths, relative intensities, and profiles. Measurement and analysis of their precise wavelengths throughout the spectrum enable us to find the Sun's chemical composition. Solar absorption lines contain more than 25,000 different lines in the region accessible to photography alone. Identification is made by comparison with laboratory spectra of atoms and molecules and about 80 different elements have been recognized in this way, the most predominant being hydrogen, calcium, and helium. The strongest lines are the H and K lines of ionized calcium.

8.6.2 Specific Instrumentation

Different types of instruments have been developed to get monochromatic images of the Sun, or of specific features, such as spots, faculae, or prominence. Two types of equipment may be sharply distinguished. The first consists of spectro-heliographs capable of covering a wide spectral range. The second type has a range restricted to one or more specific, neighboring lines and is capable of resolving very rapid events.

8.7 SPECTRO-HELIOGRAPHS

In the design of a spectrograph an objective forms an image of the object and a slit selects a narrow region of this image. The light from this strip is dispersed by a prism or grating preceded by a collimating lens and followed by an appropriate *camera* lens. The dispersion varies proportionately to the wavelength and it takes place in a direction perpendicular to the axis of the slit. This produces a spectrum at a specific wavelength of the area covered by the entry slit. If one spectral line is isolated by a second slit, the image falling on the latter represents the area of the Sun selected by the first slit at the wavelength selected by the second slit. Thus, the result obtained is a monochromatic image. The image of the Sun may be trailed slowly across the entry slit and at a result it sweeps the whole of the solar image at constant speed. If the film behind the second slit is moved simultaneously in the same direction and at the same speed, it captures an image of the solar surface at the desired wavelength. Such a device is called spectro-heliograph (Fig. 8.7).

Solar spectro-heliographs are heavy instruments and cannot be mounted on a refractor. The spectro-heliograph is generally a fixed instrument and light is fed to it via a coelostat. The following are the sources of error with spectro-heliographs:

(*i*) ***Permanent.*** If the displacements of the slits are not precisely parallel, or if the first and second slits have unequal residual curvature the errors are permanent.

(*ii*) ***Random.*** If there is a variation in the rate at which the slits move relative to the fixed parts of the equipment causing wear in any part of the variable speed drive train, periodic error arises from the lead-screws and the clutch for the worm and nut, or any other periodic error arises from inaccuracies in any part of the whole drive train.

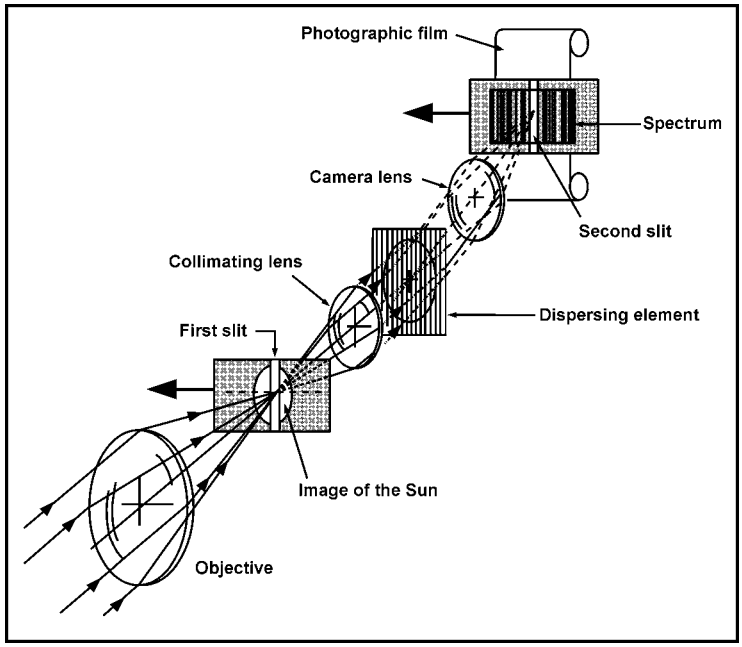


Figure 8.7: Principles underlying the design of a spectro-heliograph.

8.7.1 Coelostat

The coelostat has two mirrors. The first one called the *primary* rotates at half the Sun's diurnal rate around an axis parallel to the Earth's axis and located in the same plane as the reflecting surface; consequently, it reflects light into a fixed direction. A secondary mirror, on the other hand, subsequently reflects the beam toward the optical axis of the objective at the entrance to the spectroscope. In order to allow for the Sun's changes in declination, the secondary mirror may be shifted in a North/South direction so that it intercepts the beam from the primary mirror. In the winter season when the Sun is low, the secondary is very close to the primary and may even cast a shadow on it. When this occurs, the primary is shifted toward

the East in the morning and toward the West in the afternoon. The field rotation that results needs a correction to the Sun's position angle to be made.

If undertaken on an amateur scale, a portable coelostat may permit the less robust part of the instrument to be permanently mounted under cover. But alignment should be done with great care since the coelostat may introduce errors, particularly if observation includes long exposures. These are as follows:

(*i*) ***Permanent.*** Errors owing to the setting of the hour angle if observations are not always carried out at the same time of day and errors due to the poor mechanical state of the tangent screw in the drive strain are considered permanent. Another error is that of variations in the speed of rotation of the hour axis due to the variations in the frequency of the electricity supply.

(*ii*) ***Random.*** An error in alignment is not highly significant in summer when the primary mirror coincides with the optical axis of the instrument. The same does not apply in the winter, when the coelostat is East or West of the optical axis. Any error is responsible for the field rotation and thus changes the position angle.

8.8 HELIOGRAPHS

If the wavelength to be studied is predefined or when just one specific line is to be studied in a monochromatic observation, it is possible to do away with the cumbersome spectro-heliograph by using suitable filters. Whether commercial interference filters or the more sophisticated professional filters such as the Lyot, Halle, or other types are used, the method may be adapted to a simple refractor fitted with small aperture, long-focus optics. For amateurs this is the simplest way to study features in the chromosphere, such as filaments/prominences, spicules and flares at both the limb and on the disk, for example, in H light. Interference filters are expensive and they are fragile. They have a limited lifetime, but the results obtained with them are highly exciting.

8.9 AMATEUR INSTRUMENTS

Attempts to construct spectro-heliographs or heliographs required for monochromatic observation of the solar disk are very rare among amateurs. The former needs complicated mechanical arrangements. For the heliographs, an interference filter is required with a very narrow bandwidth (less than 0.1 nm). This requires care in handling and is also very demanding in use.

Monochromatic observation of prominences is easy to undertake. Amateurs make this form of observation using one of three techniques:

(*i*) A prominence telescope, the optical design of which resembles that of the coronagraph invented by B. Lyot. Here, use of an interference filter is essential to avoid any risk of burning the retina. The image-resolution depends on the quality of the filtration used.

(*ii*) A monochromatic interference filter with a narrow bandwidth of less than 0.5 nm located at the focus of a reflector or refractor.

(*iii*) A prominence spectroscope, located at the focus of a refractor or a reflector and using a widened slit. This method is very effective, and has the advantage of being adaptable to a reflector or refractor that is not specifically designed for solar observation.

8.9.1 Optics and Mounting

The object glass of a refractor or the primary mirror of a reflector forms, at its focus, an image of the Sun with a definite diameter. This image is projected onto a slit that has a variable width. In front of this there is an occulting cone with a diameter slightly less than that of the image of the Sun at the focus. The cone and slit assembly is held in place in the focal plane by a collimating lens. Behind the latter lies a 600-lines/mm, reflecting diffraction grating for directing the solar spectrum toward a small telescope.

The various elements in this type of prominence spectrograph (Fig. 8.8) may be constructed in different ways.

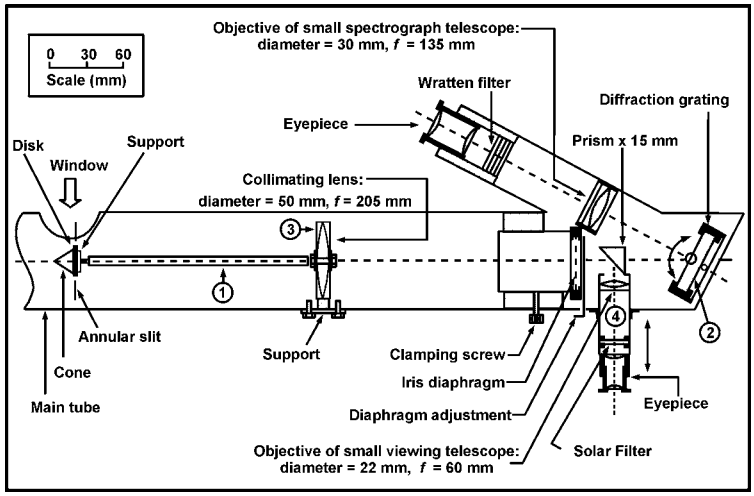


Figure 8.8: A schematic diagram of a prominence spectroscope made by Francois Castard.

The whole spectroscope is required to mount on a refractor or reflector that has a very stable mounting and a very even drive. The slightest vibration or inaccuracy in driving will seriously degrade the image quality. To avoid excessive beating of the spectroscope, its refracting or reflecting primary should have a long focal length. Its focal ratio should be around $f/15$. The occulting cone is made by applying a file, held at 45°, to a cylindrical bar of aluminum, held in the chuck of a drill, running at high speed. The slit is the most important part of the design. The use of an annular slit that may be adjusted in width is recommended. The advantage of following the curvature of the chromosphere is that long prominences may be seen over the whole length. The collimating lens should have a diameter two to three times that of the Sun at the focus. A hole may be drilled through its center using a sensitive drill-stand and a tungsten-carbide drill fed with 180-minute carborundum. The optical surfaces should be protected beforehand with either a coat of varnish or masking tape.

Adjustments. To adjust the spectroscope, observers should take precaution to avoid any chance of burning the retina if any technical hitch should occur. The image of the solar disk projected onto the occulting cone is particularly blinding and presents a real danger. Aside from any filtration, solar telescopes should always be fitted with a filter when settings and adjustments are being made. Adjustments should always be made in the following order:

(*i*) ***Step 1*:** The annular slit should be set to lie exactly at the focus of the collimating lens. Adjustments to obtain a sharp image of the solar limb projected onto the edge of the occulting disk are made by changing the position of the whole spectroscope.

(*ii*) ***Step 2*:** The position of the diaphragm should be adjusted to give a sharp image of the objective of the main telescope using tracing paper held temporarily against the blades of the iris diaphragm. The contrast is made by varying the aperture of the Lyot diaphragm according to the width of the slit and clarity of the sky.

EXERCISES

1. What are the dangers posed by direct observations of the Sun? What are the precautions to be taken to avoid any possibility of damage to the eyes?
2. Describe the method of direct visual observation of the Sun. What are two reasons for which this method is not recommended?
3. How is the Sun observed by projection? Describe the method.
4. Why is the direct photography of the solar image (at the primary focus of the instrument) not wise? What is a filter and where it is to be placed?

5. Why is high contrast not the most important factor in solar photography? How is the geographical North-South determined using the photographic method?
6. Discuss the mounting and orientation of the image.
7. Explain with a suitable diagram the variation in the geographical North-South direction as a function of time of day for an Altazimuth instrument.
8. How are sunspots watched by amateur astronomers?
9. How is the Wolf Number related to the number of individual spots? How is the active area calculated?
10. How do you measure the proper motion of sunspots?
11. What are the instruments developed to get monochromatic images of the Sun? Give the underlying principle.
12. Describe the construction of spectro-heliographs.

Chapter 9

Observing the Sun with a Coronagraph

9.1 INTRODUCTION

The corona is only about one millionth of the brightness of the Sun, so its observation with an ordinary telescope, outside of eclipses, is quite impossible. It was realized about a century ago that the corona is the Sun's outer atmosphere. Many astronomers tried to study it outside eclipses. Total eclipses are rare and visible from very restricted areas of the Earth. Initially, astronomers succeeded only to observe prominences, at the very base of the corona. It observed at wavelengths corresponding to their emission line, that they look much brighter. The instrument that allows observation of the corona outside eclipses is the coronagraph. Invented by Bernard Lyot in 1930, it succeeded in eliminating the source of scattered, stray light, which is stronger than the corona if no precautions are taken.

Observing white-light corona using a coronagraph is a difficult operation. It requires both a perfect instrument and an ideal site, but it is still possible to observe monochromatic emission from the corona under slightly less perfect observing conditions with a suitable monochromatic filter. This requirement became less

costly with the development of the interference filter around 1950. Subsequently, a small-sized coronagraph suitable for amateurs was devised at the beginning of the 1960s. Twenty-five years later the study of the corona has taken on new life with the development of radio methods and through ultraviolet observations from artificial satellites.

9.2 BACKGROUND OF CORONAGRAPH

A background on the principles behind coronagraphs is needed here. These instruments give the false impression of being simple. The mental picture that generally springs to mind is like that shown in Fig. 9.1. An object glass forms an image of the Sun and its surrounding corona at a focus, where an opaque disk is located. If we photographed the resulting image of the Sun, then we ought to obtain an image of the corona. However, the basic instrument (Fig. 9.1) will reveal a bright halo of the stray light that has been scattered by the Earth's atmosphere and by the instrument itself. The problem is to eliminate the various sources of scattered light one by one. Many interesting observations then become possible in clear weather. The failure to consider any particular source of scattered light, however, could ruin all attempts at observation. So a systematic approach is essential for solving instrumental problems.

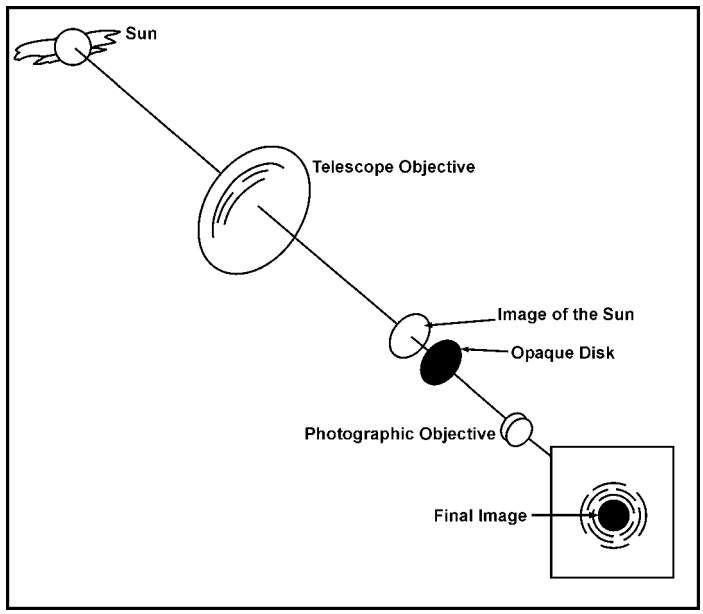


Figure 9.1: An arrangement often initially considered to observe the Sun's outer atmosphere.

9.3 REDUCTION OF THE HALO OF STRAY LIGHT

We put a small opaque disk D at S′ to produce an image of L_1 on a screen E with another lens L_1. The disk D should block all the light from S at S′, so we might expect to see no light at all on screen E. If source S is sufficiently intense a very detailed image of lens L_1 is observed at E. Thus, defects in the lens L_1 produce scattered light. The same thing will apply if the source of light is the Sun. Even after the disk of the Sun is blocked at S′ there will still be various sources of stray light within the instrument that will create a luminous halo, which will seriously interfere with observations.

The bright spots and lines seen on the image of L_1 are caused by defects in the lens; the material may be imperfect and contain bubbles or striae; the surfaces may not be perfectly polished and may have small pits or scratches; and the surfaces may not be perfectly clean and have greasy spots or specks of dust, etc., on it. To reduce the halo of scattered light, first, the lens should be made from a piece of glass that is completely homogeneous, and second, it should be ground and polished with extreme care. These two conditions can be met, but this assumes that the optical worker making the coronagraph's lens has powerful inspection methods available. The difficulties in meeting the objective may be minimized by using a simple lens made of a fairly hard glass that is not prone to scratching when cleaned. The arrangement of Fig. 9.2 is ideal for testing any lens offered by a manufacturer.

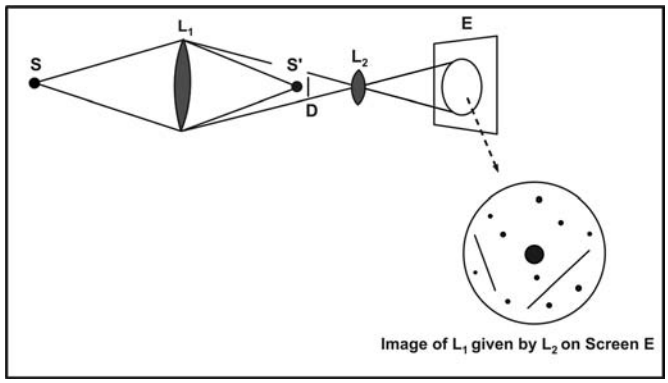


Figure 9.2: Using this laboratory arrangement, the various forms of stray light that occur within the lens L_1 become visible.

We now turn to the brilliant spot of light in the center of the image of the lens L_1 that is observed on the screen E. Fig. 9.3 shows a cross-section of the objective L_1, which is a simple lens, normally chosen to be plano-convex. The major part of the light incident on the lens will form an images S′ of the light source, and this

will be hidden behind the disk D. But transmission of a beam of light through a lens is not perfect, and partial reflections arise at the entry and exit surfaces. In the objective, a small fraction of light will be reflected by the outermost surface and will be sent back toward the Sun. This will, of course, not cause any problems. An approximately equal amount of light will be reflected at the glass-air exit surface of Fig. 9.3. This portion will suffer a second partial reflection at the entry surface, and this will send a small fraction of the beam back to form a secondary image of the Sun, somewhere near S′. This image is very faint as it is formed by a double reflection, and its intensity is approximately $5/100 \times 5/100 = 25/10000$ of that of the principal beam. But it is particularly inconvenient as it should be formed at a point where it cannot be hidden by the opaque disk D. The bright spot seen in the center of the image of L_1 (Fig. 9.2) is caused by this light that has undergone a double reflection from the surfaces of the objective.

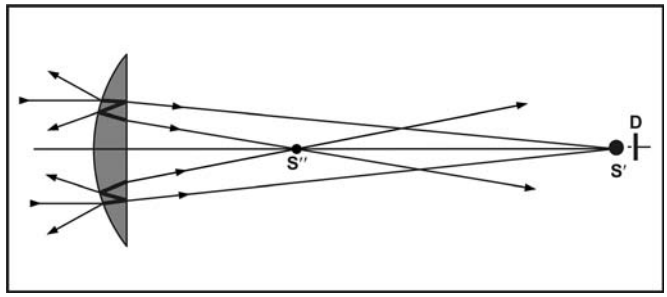


Figure 9.3: The geometrical paths of rays doubly reflected by the surfaces of a simple lens.

A simple solution to this problem was devised by Lyot. An image of the objective L_1 is formed at a point within the coronagraph assembly. A small disk, placed at that point, will block out the troublesome bright spot caused by the double reflection within the lens.

9.3.1 The Role of Diffraction

Let us now consider the next source of stray light that we identified as the brilliant ring surrounding the image of the objective L_1 on the screen E (Fig. 9.2). The somewhat obscure explanation is that this consists of light diffracted by the circumference of the lens. But the description is not completely satisfactory. It is well known that the image of a star given by a perfect astronomical objective is not an infinitely small spot of light, but appears as the airy disk, which is a small central disk surrounded by rings of light. This is shown in Fig. 9.4, where the diagram on the right reveals the variation in the intensity of light that is measured if the image is scanned in the direction indicated by the arrow (top figure). We assume that we are observing not a star but a straight line. This time (Fig. 9.4, center) the image

given by the objective will be a slightly widened, bright line, bordered by weaker fringes with minima where the intensity does not completely drop to zero. Finally, we consider the image of a luminous area that is sharply bounded at its right-hand edge (Fig. 9.4, bottom). We may consider the luminous area as consisting of numerous narrow, luminous lines side by side. The image given by the objective cannot be properly represented in a drawing, but it will gradually diminish. This means that a certain amount of light will reach areas of the image that ought to be completely black.

The disk of the Sun resembles a luminous spot, the brightness of which falls to zero as soon as we pass the limb. Although the Sun may be like that, because of the diffraction effects, the Sun's image will not be sharply defined but will fall off gradually. This stray light will not be blocked by the opaque disk inside the coronagraph; rather, it will be mingled with the weak light of the corona we want to investigate.

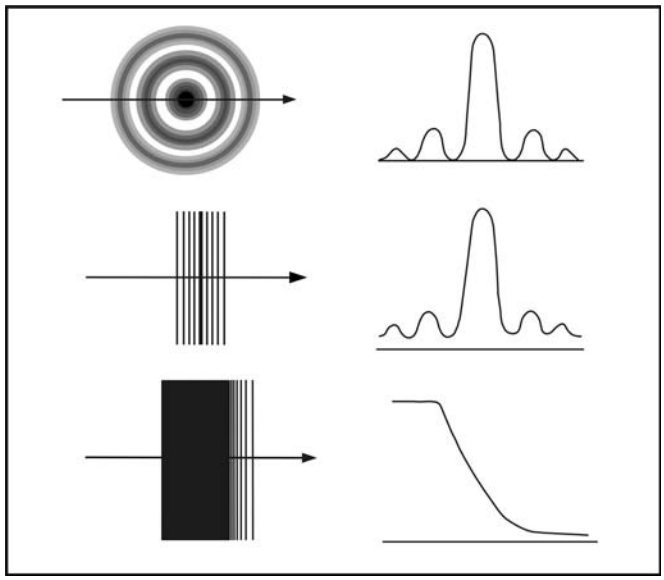


Figure 9.4: Due to diffraction phenomena, even a perfect lens does not produce an infinitely small luminous point, but instead produces the well known pattern (top) of a disk surrounded by rings. Similarly, the image of an infinitely fine line is enlarged and bordered by figures (center). Further, the image of a bright area with a sharp edge will appear as a more gradual decline (bottom). This causes considerable degradation of observations at the solar limb.

The problem of suppressing stray light caused by the diffraction fringes is well known to optical workers. The stray light may be suppressed by masking the brilliant fringe around the image of L_1 by a slightly smaller diagram. In practice, a

distinctly smaller diaphragm (about 2/3 or 3/4 of the size of L_1's image) is used. It is wise to lose a little light rather than to run the risk that any slight misalignment will allow part of the bright fringe.

9.4 DESIGN OF A LYOT-TYPE CORONAGRAPH

Fig. 9.5 shows the design of the coronagraph as devised by Lyot. The objective L_1 is a simple plano-convex lens, perfectly polished, with a focal ratio of about 1/20. Due to this long ratio, geometrical aberrations are negligible. In addition, problems of heating at the primary focus may be controlled without much complication. The metal disk D is calculated to be the same size as, or slightly larger than the image of the solar disk that it is to hide. The apparent diameter of the Sun varies during the course of the year as the distance of the Earth from the Sun varies. So a different disk must be provided to correspond to the size at different times, and to cater to the various wavelengths being studied. This second point arises because the objective L_1, being a simple lens, suffers from chromatic aberration.

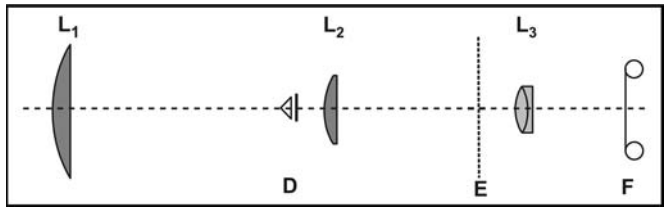


Figure 9.5: A schematic diagram of a coronagraph.

Fig. 9.6 reveals a method of mounting the reflecting cone and occulting the disk D that has proven successful in practice. The two are held by a central rod several centimeters in front of the lens L_2, which has been drilled through the center. An insulating sleeve protects the glass from the heat transmitted by the cone. To avoid thermal problems, the cone should be highly polished, or even aluminized. Lens L_2 forms an image of L_1 at the diaphragm E. Fig. 9.7 shows schematically the image that should be visible at E when the coronagraph is properly adjusted. The diaphragm should have a diameter somewhat less than the full aperture of L_1. This means that some light is lost, but more stray light will be lost than light from the corona, which is fundamental. The resulting arrangement is also reasonably insensitive to small errors in alignment. The central disk should not be in contact with any optical component.

Behind the diaphragm E in Fig. 9.5, all the instrumental sources of scattered light have been eliminated. It only forms an image of the *eclipsed* Sun onto a recording medium, e.g., onto a film F, by a photographic objective L_3. To avoid

increasing the length of the overall instrument, it is advisable to mount L_3 just behind E. In addition, the focal length of L_3 will determine the magnification of the final image, with a focal length equal to half that of the distance between L_3 and the occulting disk. If everything is favorable the level of scattered light is only one millionth of the flux from the non-occulted Sun. With a good quality lens it is certainly possible to obtain scattered-light levels that are $1/10000^{th}$ to $1/100000^{th}$ of the non-occulted flux form the Sun.

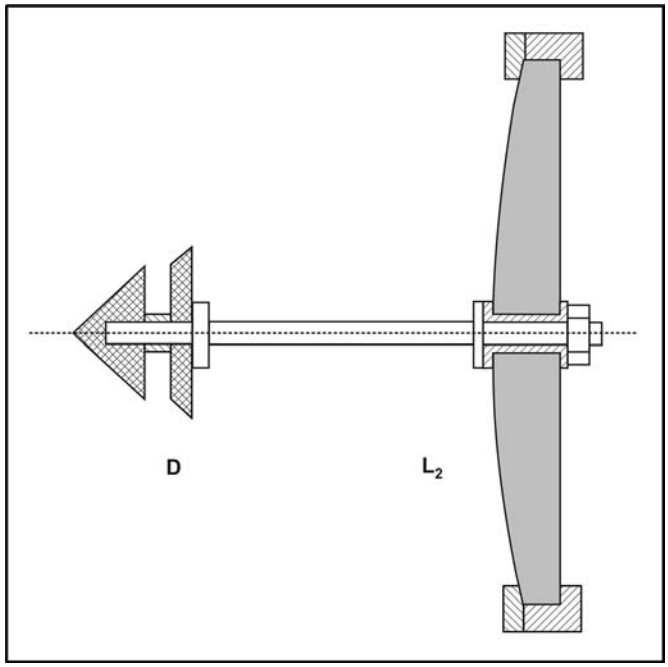


Figure 9.6: Method of mounting the cone and occulting disk D in front of lens L_2 in a coronagraph.

9.5 THE ATMOSPHERIC HALO

The bold line in Fig. 9.8 shows the variation in intensity from the center of the solar disk to the white-light corona. The scale here is logarithmic, where the lower of the fine lines shows the brightness of the sky close to the Sun during a total eclipse. This is about $1/1000^{th}$ of the intensity of the inner corona. On the right-hand side of the figure, the brightness of a blue sky, several lens of degrees away from the Sun, is indicated. This light characteristic of a clear blue sky arises from the scattering of sunlight by the molecules forming the Earth's atmosphere. It is seen from the figure that its brightness is more or less equal to that of the white-light corona. The

sky close to the Sun is not blue; the disk of the Sun behind a distant object is a very brilliant, white halo around the object. This is indeed an atmospheric effect and its contribution is shown in Fig. 9.8 (upper line). This is, in fact, an atmospheric corona, but the term halo is used to differentiate it from the solar corona.

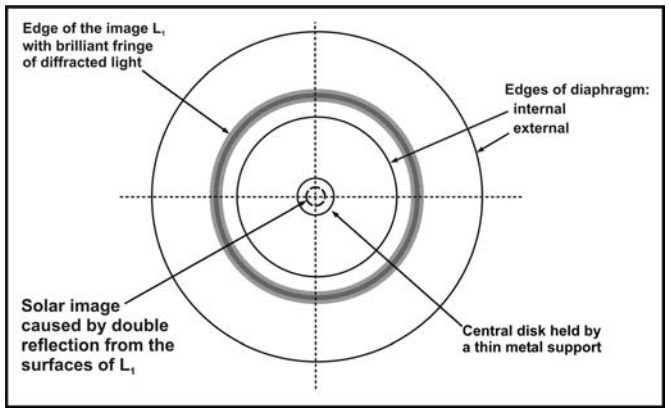


Figure 9.7: Appearance at the plane of the Lyot diaphragm in a properly adjusted coronagraph.

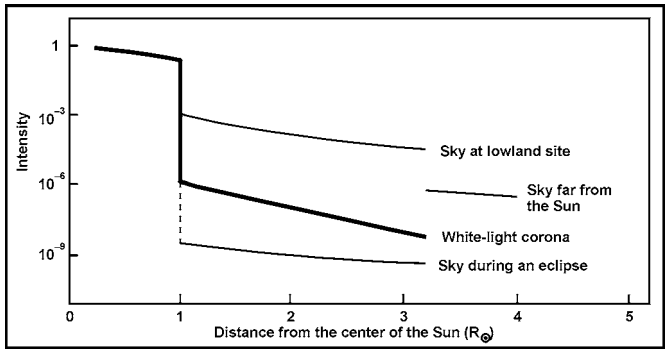


Figure 9.8: Ordinate shows the luminous intensity on a logarithmic scale with respect to the intensity at the center of the solar disk. While the abscissa reveals distances from the center of the Sun, expressed in solar radii, the curves show the variation in the intensity of the Sun's outer atmosphere (the white-light corona) and that of the sky through which it is observed both during and outside an eclipse.

The halo, due to scattering by aerosols, is not the only detrimental effect arising in the Earth's atmosphere. Fig. 9.9 shows (solid line) the brightness of the atmospheric halo as observed at a high-altitude site with a good quality coronagraph that does not add any scattered light. The ordinate gives the luminosity expressed on a logarithmic scale with respect to the center of the solar

disk. The abscissa represents the distance from the Sun's limb in seconds and minutes of arc, also on a logarithmic scale. The solid line shows the variation in the scattered corona at the final focus in a properly adjusted coronagraph. The dashed line shows how this scattering changes between good seeing conditions and those with a sky that causes a slight degree of scattering. The objective of a coronagraph is generally a simple lens (Fig. 9.10), so it is impossible to block out the whole solar disk at all wavelengths. Here, the occulting disk is placed at the focal plane for red light (R), and it can be seen how part of the blue image (B) is not obstructed.

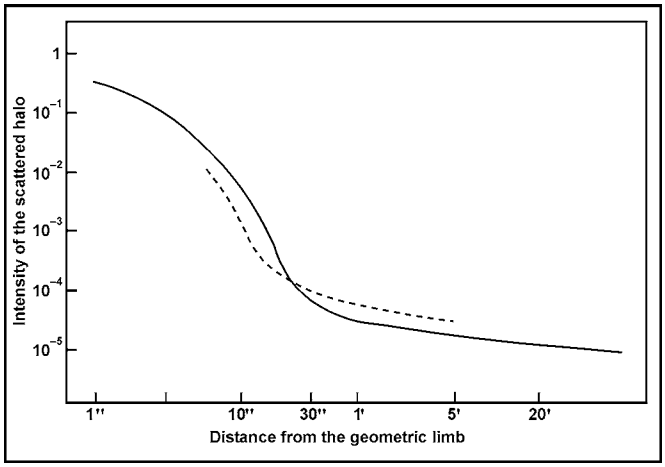


Figure 9.9: Ordinate shows the luminosity expressed on a logarithmic scale with respect to the center of the solar disk, while the abscissa represents the distance from the Sun's limb in seconds and minutes of arc, also on a logarithmic scale.

9.5.1 Removal of Stray Light Due to Chromatic Aberration

We know that a coronagraph's objective should be a simple lens. A doublet has a number of sources of additional scattering in the four glass surfaces and in the cement used to assemble the two lenses. But the simple lens is affected by chromatic aberration. The image of the Sun in red light (R in Fig. 9.10) is larger and farther from the lens than it is in blue light (B in the figure). The distance between these two points is about 20 mm for a simple lens with a focal length of 1 m. The occulting disk D should be located in the primary-image plane corresponding to the working wavelength, which slightly complicates the design. But if the occulting disk is placed over the red image of the Sun, some of the blue light will not be interpreted. Thus, the artificial eclipse will only be complete for a small range of wavelengths and the level of scattered light in the coronagraph will increase sharply at other colors. When the occulting disk is adjusted to fit the

red-light image of the Sun, there will be a flux of non-occulted blue light within the instrument. In the best case (Fig. 9.9) the scattered halo at 1 arc-minute from the solar limb has an intensity of $1/100000^{th}$ of the solar disk. Because the white light corona is about $1/1000000^{th}$ of the brightness of the Sun, it is difficult to detect without employing special methods.

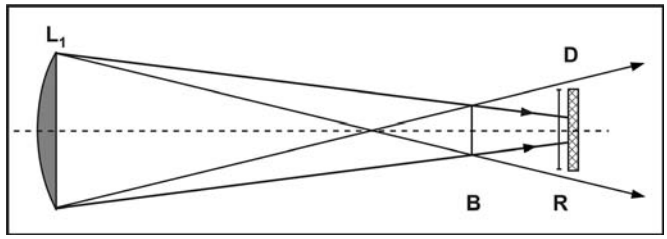


Figure 9.10: The objective of a coronagraph is generally a simple lens. Here, the occulting disk is kept at the focal plane for red light (R). Notice how part of the blue image (B) is not obstructed.

Fig. 9.11 shows a tracing of a spectrum of the corona. Apart from the continuum, there are narrow but strong emission lines in the spectrum. The principal one is in the green at 530.3 nm; another is visible in the red at 637.5 nm. At the wavelength of an emission line, the corona is as much as 100 times stronger than the continuum. The diffraction halo arising in the Earth's atmosphere typically shows a continuum with narrow absorption lines. Although the corona is weaker than the scattered halo, it predominates if it is possible to make observations at the exact wavelength of emission lines.

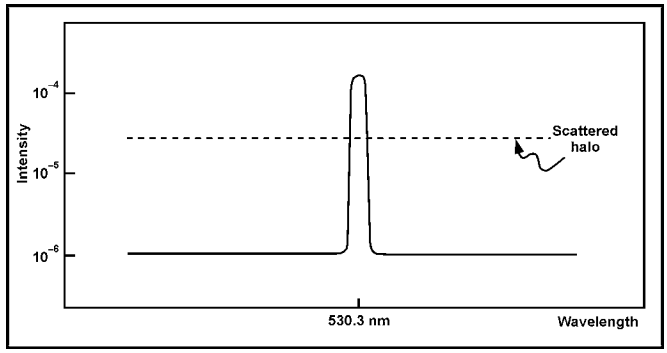


Figure 9.11: The solid line is a schematic representation of part of the coronal spectrum. In the presence of scattered light, it is easier to see the corona if it is possible to isolate wavelengths close to an emission line using a filter.

It now remains to find a good filter capable of selecting a suitable coronal-emission wavelength. The gelatin or glass filters that are used in photography

are far from being sufficiently selective (curve G in Fig. 9.12). On the other hand, interference filters give maximum reduction of the scattered halo (curve 1 in Fig. 9.12). It should be noted that it is difficult to use filters of bandwidths less than 1 nm because narrowband filters only give the desired bandwidth at a specific temperature and for a truly parallel beam of light. To detect coronal emission it is essential to reduce the scattered light as much as possible, so a filter with a bandwidth of 0.2 nm must be used. This will require the use of a temperature-controlled enclosure. The filter should be installed in a beam that converges or diverges as little as possible. A solution (Fig. 9.13) is to use two doublets instead of the lens L_3 and to mount the interference filter between them.

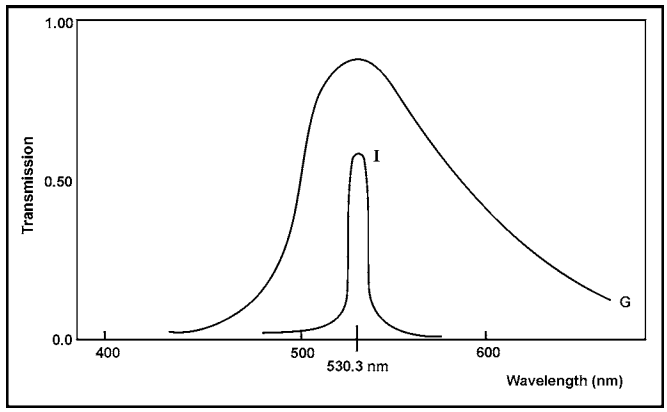


Figure 9.12: General shape of the bandwidth of a filter made from colored gelatin (G) and of an interference filter (I).

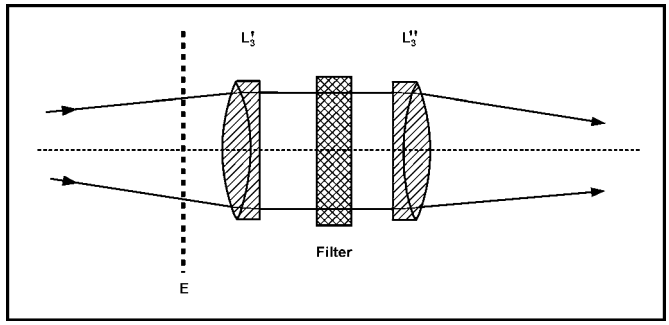


Figure 9.13: Interference filter should be located where the beam of light is essentially parallel. This is achieved by splitting lens L_3 as indicated.

The situation is similar with respect to prominences. Observation is greatly facilitated in this case by using a filter that can isolate the emission lines. Fig. 9.14

shows a schematic representation of the most intense emission lines in the visible region. Observations may be made easily if the H_α hydrogen-line can be isolated by using a filter centered on 656.3 nm. A bandwidth of 5 to 10 nm suffices to show prominences clearly, although a narrower filter allows observation when there is greater scattering from the sky.

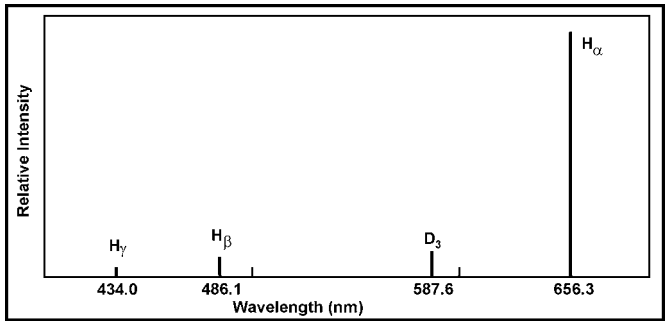


Figure 9.14: In the visible region the spectra from prominences mainly consists of hydrogen lines (H_α, H_β, H_γ, etc.) and a helium line (D_3). The red H_α line is the strongest, and so it is easiest to see.

9.6 OBSERVATION OF PROMINENCES AT THE LIMB

In all solar investigations, observers must not forget that precautions are to be taken so that the intense flux may not damage the instrument, or worse, the observer's eye. These warnings are valid for observations with a coronagraph. Because the blinding solar image is *normally* hidden by the occulting disk, it is tempting to look at the prominences without any additional attenuation of the beam of light. But any accidental decent ring of the coronagraph by an involuntary movement, a gust of wind, or failure of the drive motor, etc., would uncover the Sun and could immediately cause a burn, which might be severe. This only serves to emphasize that special attention should be taken to the equatorial drive of a coronagraph. A good drive is, of course, essential if good photographs are to be taken. Here the potential problem is more serious. If someone is observing a planet through a telescope that is tracking badly, there will be a slow shift of the object in the eyepiece and this is not too inconvenient for visual observation or short exposures. With a coronagraph, any fault in tracking will cause the Sun to appear in the form of a dazzling crescent, preventing any useful observation, and possibly putting the observer's eyesight in danger. So the golden rule for the observer is not to put an eye to the eyepiece unless the reduction in the beam of light is such that the non-occulted image of the Sun can be born without any danger. If an interference filter is used, then the reduction in the amount of sunlight is achieved by virtue

of the filter's narrow bandwidth. However, it is essential to know whether the interference filter has a secondary bandwidth in the infra-red. Such radiation may be dangerous, even though it is invisible. Heat filters are commercially available to provide the desired protection when used in conjunction with the interference filter used to isolate the light from the prominences.

The precautions necessary to safeguard the observer's eyes also applies to ancillary equipment that may be installed behind the coronagraph. Photoelectric receptors may be protected by some form of cut-out, which interrupts the supply in case of accidental overload. It should be remembered that a considerable amount of light and heat is rejected by the occulting disk at the coronagraph's primary focus, and that adequate free space must be maintained around the disk and the occulting cone.

9.7 TYPES OF OBSERVED EMISSION

The curve shown in Fig. 9.9 shows the variation in the brightness of the halo of stray light observed at the final focus of a good coronagraph. This is repeated in Fig. 9.15, which shows the variation in brightness of prominences observed in H_α, H_β, and D_3. In Fig. 9.15, the relative intensities of the stray light and the emissions correspond to those observed through narrow-pass interference filters. There is no doubt that observation of the corona in monochromatic light is not simple and should only be attempted by an experienced observer having the benefit of

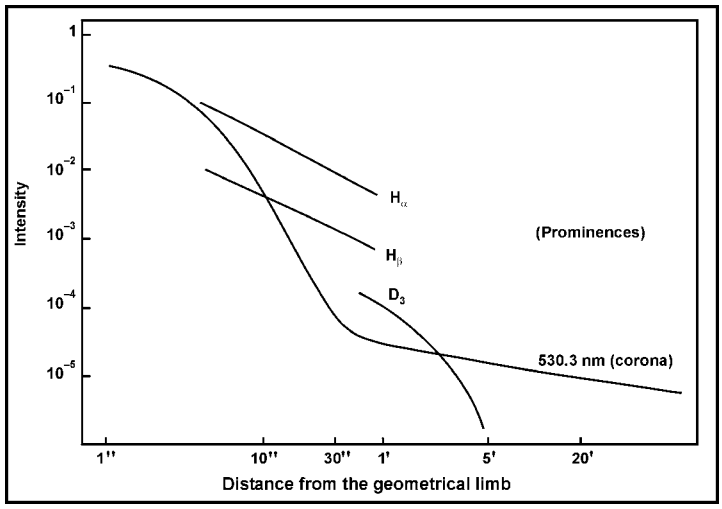


Figure 9.15: The intensity of scattered light compared with the intensities of prominence and coronal emission lines observed through an interference filter.

an excellent site. We know that observation of the white-light corona when the atmospheric corona cannot be reduced by proper filtration is scarcely feasible. The images seen in monochromatic emission only show coronal material at a very specific temperature, depending on the wavelength studied, so the visible features entirely depend on the filter employed. The white-light corona exhibits material over a wide range of temperatures simultaneously and thus gives a broader-based picture.

The best way to become familiar with a corona graph is to observe prominences in H_α light. It may be noted that on some days the Sun is practically featureless. Prominences are constantly varying features in the solar atmosphere. Some of their characteristics are described below.

9.7.1 The Telescopic Appearance of Prominences

One interesting characteristic of prominences is their extremely fine structure. The photographs taken at different observatories show only a 0.5 arc-second across and even these are probably enlarged by turbulence. This means that the detail accessible with an amateur coronagraph is likely to be limited by the resolving power of the objective or in most cases by turbulence. It should be noted that average daytime turbulence amounts to several seconds of arc, and it is rare to have both a very clear sky and stable images at the same time. In any case, the most favorable conditions occur in the early morning hours. Photography of prominence in H_α light needs films sensitive to red light.

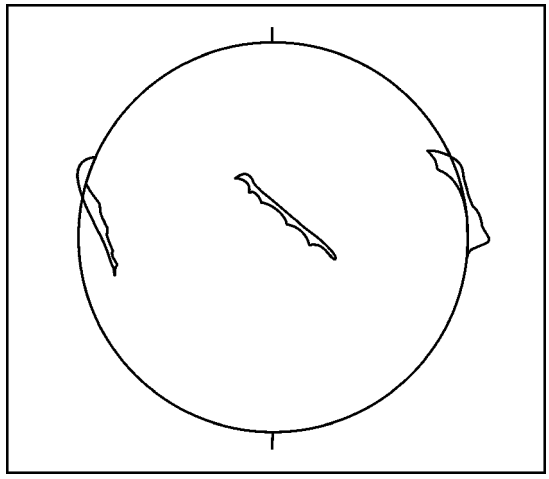


Figure 9.16: Due to the rotation of the Sun, prominences may be seen with a coronagraph at the eastern limb, then observed against the disk with spectro-heliograph, and finally become visible at the western limb before disappearing behind the Sun.

When observed through a telescope, prominences give the observer the impression of being located on the plane of the sky. But in reality prominences are generally regions of relatively cool gases, which have thicknesses of about 10,000 km, and a length 10 times of that amount. When viewed in H_α light against the solar disk prominences appear as dark filaments. These characteristically look rather like a long, thin viaduct. They extend farther in longitude than in latitude. When they cross the limb they are seen in profile, and their appearance is greatly altered because of perspective. Reconstructing their true shape in space would entail having knowledge of their appearance against the disk, and also knowing that they are evolving very slowly with time. These two conditions are very rarely met (Fig. 9.17).

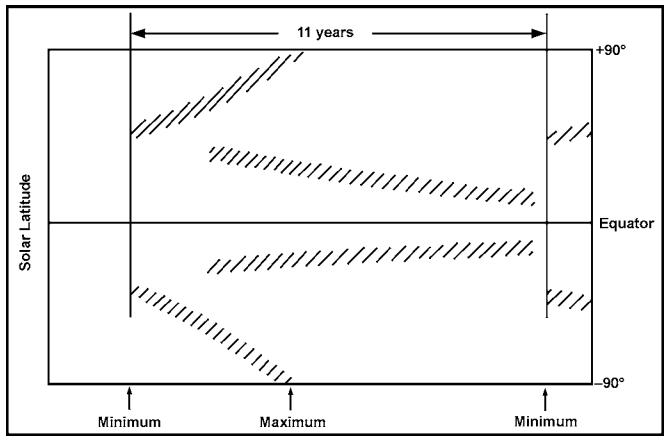


Figure 9.17: Bands of latitudes where prominences most frequently appear (shown by the shaded zones in the figure) vary over the course of the 11-year cycle.

9.7.2 Changes in the Appearance of Prominences

The appearance of prominences changes over three typical time-scales. Under very stable seeing conditions, the finest detail in prominences is found to alter over a few minutes or tens of minutes, sometimes reproducing an earlier state almost exactly. The cause of these changes is not clearly understood. Contrary to appearances, prominences are not hot *flames* extending from the Sun into cold empty interstellar space. Instead, they are relatively dense, cool regions (8000 K) immersed in the Corona, which is far too hot (2,000,000 K) to emit any radiation at the H_α wavelength. Prominences are thus the site of extremely complex thermal-instability phenomena. From one day to another, prominences change their appearance. But here it is difficult to separate real changes from variations caused by the alteration in perspective. The Sun rotates by about 13° per day, and even a

very high prominence does not remain visible for more than about 2 to 3 days each time it crosses the limb. The lifetimes of prominences may exceed a month. Hence, it is possible to see a prominence disappear over the western limb and then recover it at the same latitude at the eastern limb 14 days later after the Sun has completed half a rotation. Despite this, the appearance is very different after a period of two weeks. The third time-scale of changes in prominences arises from the 11-year solar cycle. Prominences are less frequent around the maximum, when the disk has most spots. Moreover, the latitudes at which prominences appear depend on the phase in the cycle. If the latitudes are plotted throughout the cycle, we may get a diagram shown in Fig. 9.17. At a minimum there are two bands of prominences on the Sun, at about 45° latitude in each hemisphere. Prominences may therefore be seen crossing the limb in the Northeast, Northwest, Southeast, and Southwest. As the cycle progress, some of the prominences shift toward higher and higher latitudes forming what is known as the polar cap and ultimately more or less reaching the poles at the time of maximum. It is then possible to see prominences rotating in one spot over a considerable period of time. A second zone corresponds to the latitudes at which spots are seen (i.e., starting at about 40° at the beginning of the cycle and ending at about 10° just before the minimum). The pattern shown in Fig. 9.17 permits one to predict reasonably well which regions of the Sun are likely to show prominences.

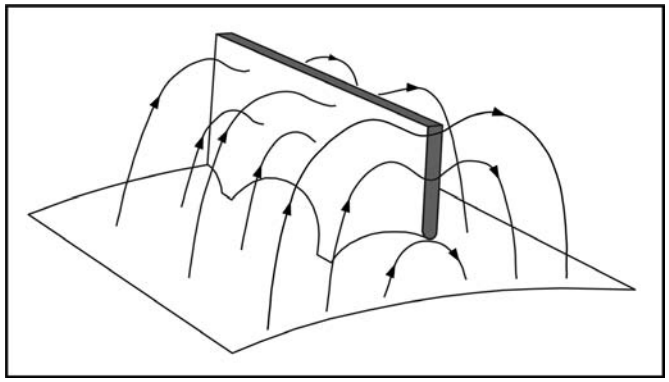


Figure 9.18: The model illustrates that the relatively cold, dense material in prominences is held in a sort of magnetic trap by the lines of force.

9.8 RAPID EVENTS

Anyone viewing the Sun through a coronagraph should not expect to get similar events. Two main categories of events deserve to be monitored. When an active

area crosses the solar limb it shows small, active prominences, which exhibit rapid changes, often in the form of mass ejections of greater or lesser frequency. In specialized works these are usually referred to as surges, sprays, etc. When an area's activity is most intense, one may see a chromospheric flare. Through a coronagraph one may observe, very close to the Sun, arches and large and small clumps of gas that appear very bright. Some hours later, these may have turned into spectacular loops, which look like fountains of light, with material appearing in the corona and raining down onto the surface of the Sun. Another category of event is the sudden disappearance of a quiescent prominence. This may happen to prominences at any latitude, but most frequently affects high prominences that are close to active areas. Through the coronagraph, the prominence may be seen rising to considerable heights and progressively dissipating while some of the material appears to fall back on each side.

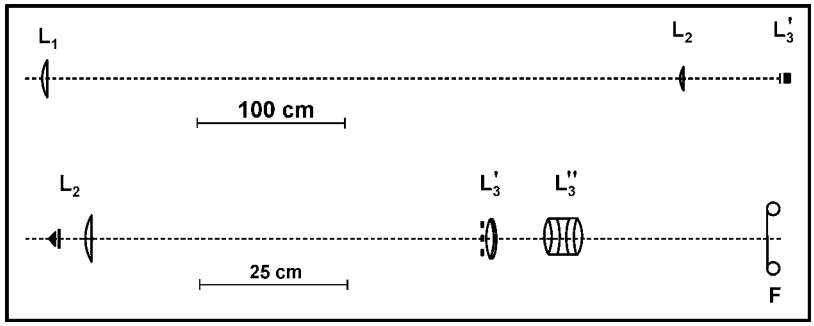


Figure 9.19: Optical diagram of a coronagraph used at the Pic du Midi to photograph the solar corona.

9.9 A TYPICAL CORONAGRAPH

Details of one of the coronagraphs used at the Pic du Midi for photographing monochromatic emission from the solar corona are presented here. Fig. 9.19 shows a schematic diagram of the various elements. Table 9.1 shows the corresponding various parameter settings of the coronagraph.

Table 9.1: Parameter settings of one of the coronagraphs used for photographing the monochromatic emissions from the solar corona.

Primary lens L_1	Plano-convex, borosilicate-crown lens; diameter: 200 mm; focal length: 4007 mm at 637.4 nm, 3968 mm at 530.3 nm.

Occulting disk D	One of a set of disks in 1-mm thick brass, with diameters: 36.0, 36.2, 36.4, ..., 39.0 mm. It is located at the focus of lens L_1 for the working wavelength, and has ahead of it a cone of polished dural, 34 mm in diameter. (The distance between the cone and the occulting disk is about 10 mm.)
Field lens L_2	Plano-convex lens in ordinary crown; 80 mm; focal length: 630 mm. It is drilled in its center in order to carry the occulting disk on the end of a metal rod. The distance between the disk D on and the lens L_2 is 40 mm.
Annular diaphragm E	Situated 660 mm behind L_2. The image of L_1 given at E by L_2 is 33 mm in diameter. The aperture of the diaphragm has a diameter of 25 mm. The central disk 6 mm in diameter hides the image of the Sun caused by double reflections from the surface of L_1, which has a diameter of about 2 mm.
Lens L_3	Situated 10 mm behind E. A clave doublet: diameter 60 mm; focal length 720 mm.
Lens L_3''	Situated 100 mm behind L_3''. The interference filter, within its oven, is located in the space between these two lenses L_3' and L_3''. The lens is a Benoist-Berthiot triplet diameter 55 mm; focal length: 350 mm.

NOTE

Final image of the Sun. This is formed about 360 mm behind L_3'' on the film, which is carried in an Exakta 24 × 36 mm body. The diameter of the (occulted) solar disk is 18 mm, with a longitudinal adjustment of L_3''. Focusing is carried out by longitudinal adjustment of L_3''.

EXERCISES

1. What is meant by corona of the Sun?
2. Discuss the principles behind the coronagraph.
3. Explain how the reduction of the halo of stray light can be done.
4. Draw the geometrical paths of rays doubly reflected by the surfaces of a simple lens, and discuss.

5. Using a diagram, show the design of the coronagraph as devised by Lyot. Next describe the different sections pointing out a method of mounting the reflecting cone, and occulting the disk in front of the lens in a coronagraph.
6. Show the variation in intensity from the center of the solar disk to the white-light corona, and explain.
7. Discuss how the stray light is removed owing to chromatic aberration.
8. Draw a coronal spectrum and explain the role of a filter.
9. Explain how one can detect the coronal emission successfully.
10. How is the interference filter located where the beam of light is essentially parallel?
11. What are the precautions used to safeguard the observer's eyes when studying with a coronagraph?
12. Draw the curves showing the variation in the brightness of the halo, if stray light is observed at the final focus of a good coronagraph, and discuss the results.
13. What is the best way of becoming familiar with a coronagraph?
14. Discuss some interesting characteristics of the telescopic appearance and prominences.
15. Briefly discuss the changes in the appearance of prominences over three typical time-scales.
16. Explain the modes to illustrate the relatively cold, dense material in prominences.
17. What are the active and quiescent prominences?
18. Present the details of the coronagraph used at the Pic du Midi for photographing monochromatic emission from the solar corona.

Chapter 10

Solar Eclipses and Studying the Corona

10.1 INTRODUCTION

Total solar eclipses are impressive astronomical events that suddenly reveal the solar corona. During totality the black disk of the Moon is ringed by bright pink light and surrounded by a silvery halo. The phenomenon enables extremely valuable astrophysical observations to be made. The solar corona is the source of the solar wind and strongly scattered light within our solar system. Total eclipses of the Sun allow thorough investigations of the corona.

10.2 CIRCUMSTANCES OF ECLIPSES

Fig. 10.1 shows the Earth and the Moon in space illuminated by the rays of the Sun. The Moon completes one rotation around the Earth in 27.3 days and has an orbit inclined at an appreciable angle (5°09′) to the plane of the Earth's orbit around the Sun. The Earth-Moon distance varies around the lunar orbit and the apparent size of the lunar disk is similar to that of the Sun (about 32 minutes of

arc). This happy coincidence arises from the fact that the ratio of the distance from the Earth to the actual diameter has a similar value for each of the two bodies. When the apparent size of the Moon is larger than that of the Sun, there is a total eclipse. If the apparent size is smaller, the eclipse will be annular. The duration of totality is governed by the ratio of the apparent diameters of the two bodies, which determine the magnitude of the eclipse. When the Earth's rotation is taken into account this duration never exceeds 8 minutes. These few minutes are extremely precious for observers, as they offer the finest view of the corona. Anyone within the shadow cone observes the brightness of the daytime sky fall to about $1/10000^{th}$ of its normal value. Prominences may sometimes be seen around the lunar disk. Surrounding everything is the magnificent halo of light scattered by the plasma in the solar corona and by particles in the interplanetary medium. The conditions required for any eclipse to be total are: (*i*) it must be a New Moon; (*ii*) the Moon must be near the line of nodes; and (*iii*) the apparent size of the lunar disk must be larger than the apparent size of the Sun.

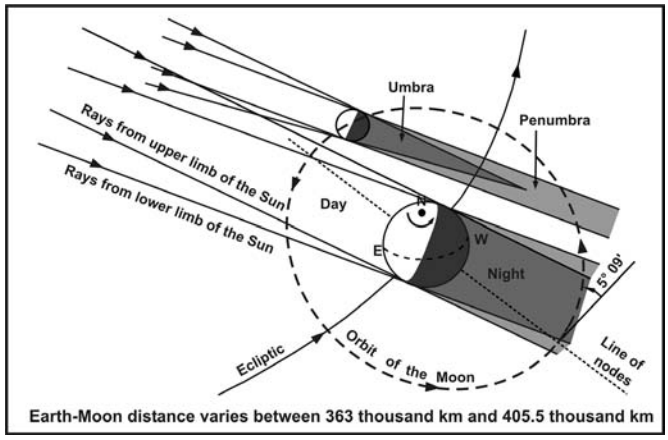


Figure 10.1: A schematic diagram showing the conditions under which the Moon's shadow may be intercepted by the Earth, producing a total eclipse of the Sun.

10.3 MEASUREMENTS OF THE SOLAR DIAMETER

Fig. 10.2 shows the various phases of a total eclipse, assuming the apparent disks of the Sun (filled) and of the Moon (open) are perfectly circular. The disk of the Moon crosses in front of the disk of the Sun, so the latter is completely covered between the 2^{nd} and 3^{rd} contacts. This lasts for several seconds because the lunar limb is very irregular (Fig. 10.3). Variations in the profile of the Maria and Highlands may permit a certain amount of the solar photosphere to be seen, giving rise to Baily's

Beads. The difference between the apparent motions of the two bodies across the sky gives a typical relative velocity of 1 second of arc in 2 seconds of time. When variations in the profile are considered, it is seen that the duration of the visibility of Baily's Beads may be taken into account. This is one of the difficulties of measuring the duration of eclipses.

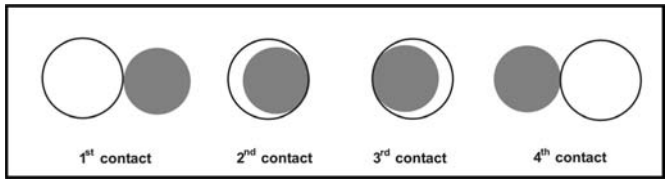


Figure 10.2: Various phases of a total eclipse.

Published ephemeredes for the Sun and the Moon permit the average apparent diameters of the two disks as seen from the Earth to be determined very accurately from the topo-centric Earth-Moon and Earth-Sun distances. By making the assumption that the diameter of the Moon remains constant over a time-scale of a few centuries and by measuring the precise interval between the 2^{nd} and 3^{rd} contacts, it is possible to determine if the diameter of the Sun is truly constant at least from one eclipse to the next. There are some doubts about the constancy of the solar diameter. Provided allowance can be made for the profile of the lunar limb, a precise measurement is possible. The intervals to be measured are generally of the order of a few minutes and accuracy of the order of 0.02 seconds of time is expected. This corresponds to measuring the Sun's diameter to 0.01 arc-seconds. If the observer is not on the central line, it may or may not be necessary to introduce a correction for the relative positions of the disks. Using the calculated ephemeredes, the interval of time may be converted into a value for the solar diameter. The methods of carrying out this experiment are similar to those used for studying the form of the Moon and for verifying ephemeredes.

10.4 IMPROVING EPHEMEREDES BY CONTACT-TIMING

The times of contact for a specific place may be calculated from ephemeredes, which have a limited accuracy. This produces errors in the times or contact that are of the order of a few seconds of time. One of the sources of error lies in our insufficient knowledge of the form of the Moon, and also of the diameter of the Sun. Several other factors also affect the calculations. As a result, the times of contact calculated from the two sets of ephemeredes may differ by two seconds of time. Contact times may be measured to a high degree of accuracy in a simple experiment. A photometer was required in the past to measure the solar luminosity during the

few seconds before and after totality. It is also possible to use a microcomputer to provide an adequate time-base. In recent years it has become possible to do far better, either by using a cine camera synchronized with the clock, or a video camera. This method has an advantage over the photometer method, where the total flux from the solar crescent is measured. It allows the problem posed by the irregular edge of the Moon (Fig. 10.3) to be tackled by providing a visual record of the Baily's Beads produced by gaps in the Moon's profile. To get a good-quality

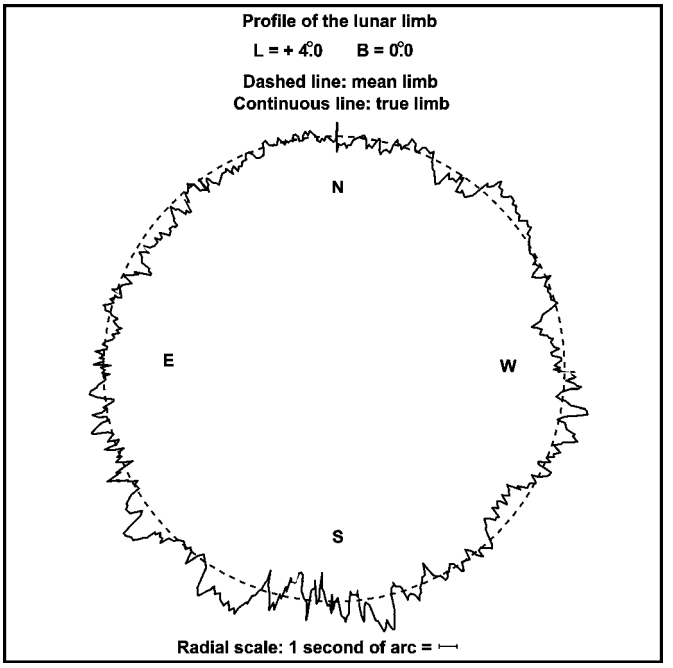


Figure 10.3: Typical profile of the lunar limb as calculated for the total eclipse of November 23, 1984. Note that the radial scale is greatly exaggerated (about 60 times).

image, the equipment should be installed on an equatorial mounting, or else fed by a coelostat, and provided with filters. The video camera's second channel may be used to record the time signals, or these may be superimposed on the image. A form of mounting is shown in Fig. 10.4.

10.5 INNER CORONAL PHOTOGRAPHY AND PHOTOMETRY

The observation of the corona during totality should be made with the greatest care so that experiments function without fail. It is preferable to undertake reliable experiments than to attempt some special feat. Even if the resulting photograph

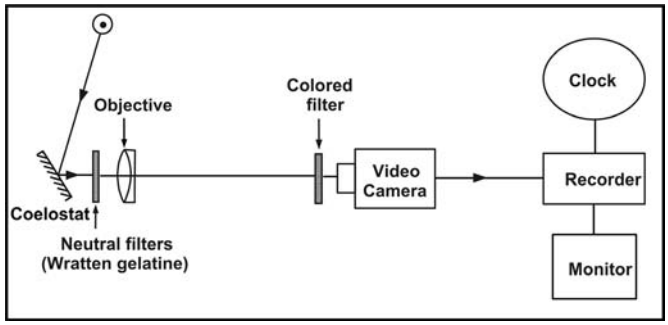


Figure 10.4: A schematic diagram of an experiment designed to measure the 2nd and 3rd contacts of an eclipse.

is unique it cannot be used unless properly calibrated and all the instrumental factors are well known. Fig. 10.5 shows the origin of the sky illumination during an eclipse. A represents primary scattering by the ground and the lower atmosphere of sunlight, which then illuminates B, the line of sight in the Earth's atmosphere. This further scatters some of this radiation toward the observer at C.

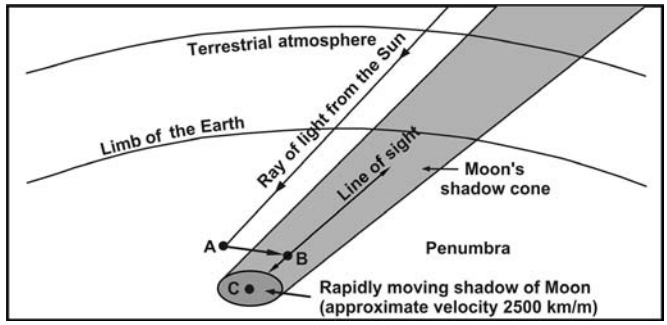


Figure 10.5: The origin of the sky illumination during an eclipse.

Fig. 10.6 shows why the solar corona may be observed during totality. The sky becomes sufficiently dark for the contrast between the various features to be perceptible. Even the faint Earthshine on the face of the Moon may be measured at that time.

10.5.1 Structures in the Inner Corona

In addition to the data acquired at total solar eclipse numerous experiments in space and on the ground, particularly with large coronagraphs (200-mm diameter or more), have discovered very complex features in this region of the solar atmosphere, which reveals a wide range of plasma phenomena. Fig. 10.7 shows the principal features of this part of the corona, which has both quite cool regions such as prominence material,

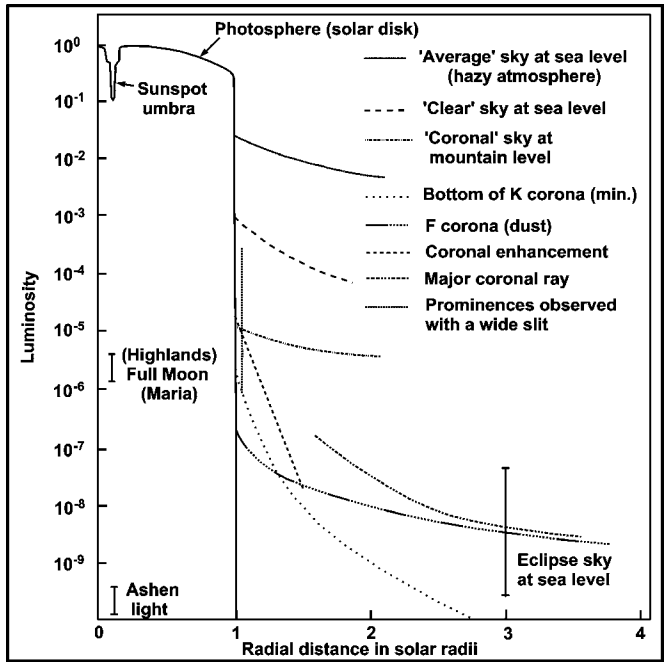


Figure 10.6: The radial distribution of intensity close to the Sun of the daytime sky and of various objects observed during an eclipse.

super-spicules, etc., and very hot ones like coronal arches and flares. These extremes are separated by only a few hundred kilometers. Recent observations show highly variable features, which may change rapidly with waves of ionization, excitation, condensation, etc. This is particularly noticeable when they are cool and low in density. Others are violent ejections of material at velocities of several hundred km/s forming very narrow, long jets. Above active regions, coronal enhancements show that changes are continually taking place especially in the shape of temporary loops that form and dissipate repeatedly and that may even be the site of violent

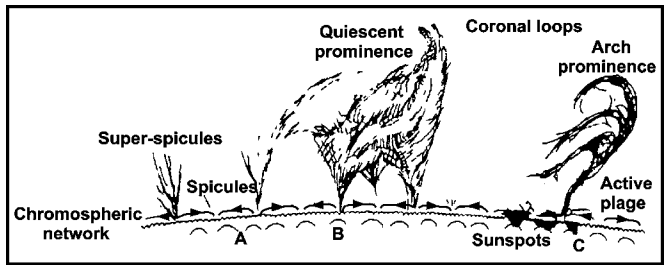


Figure 10.7: Diagram showing the main structures in the inner corona and the chromospheric extensions over various quiet (A and B), or active areas (C).

explosions. This is essential if we are to understand how the corona is heated to such high temperatures and why such a corona exists at all. Thermodynamic properties that vary with time are important factors to be taken into consideration.

10.5.2 Observations at the Extreme Limb

Observations at the extreme limb are similar to those made at the contacts, except that selective filters should be used. The details depend on whether the desired spectra are choromospheric or photospheric (the latter are emission spectra). To study the extreme outer limit of the photosphere it is preferable to work in the red and the near infra-red, where there are spectral windows that are totally free from lines. These permit the *true* continuum to be measured and cover fairly wide spectral regions. It thus ensures that there is a good signal-to-noise ratio for photometric measurements. If the photographic method is applied, exposure times should be chosen to give a correct rendition of the solar limb. To ensure that measurements may be made at the extreme limb, where the intensity is less than 1% of that at the center of the disk, contrast should be kept low ($\gamma < 1$). Also, careful calibration should be carried out, so that the profile of the brightness at the extreme edge of the solar photosphere may be deduced.

The narrow layer, the chromosphere and the spicules, up to about 10 arc-seconds from the limb, may be studied with an appropriate filter. Investigation of the super-spicules that extend out to 20 to 30 arc-seconds may also be attempted. But this is too difficult to do photographically. Visual methods may be tried with a telescope, using the optimum magnification. This type of observation could also be made with a telescope fitted with a spectro-helioscope, but the resolution is likely to be less. The best way is to use a very broad-pass filter perhaps even a Wratten gelatine one that transmits H_α lines, and concentrate on studying extremely fine detail.

10.5.3 Observations of the Inner Corona

There are different features that may be studied, provided good resolution is attainable. For these purposes photography gives the best results. To distinguish cool materials, whose radiations are dominated by the orange and red emission lines of D_1, D_2, and D_3, and (above all) H_α from ionized material at 2×10^6 K with *white* emissions, color emulsions (i.e., color sheet-film) should be used. With a refractor designed for visual use, radiation in the blue is more *diffuse* because of chromatic aberration. Cool material emits some strong lines in the blue. If these wavelengths are removed by filtration, better discrimination between coronal and cooler material is achieved. Using this method, the green layer of the emulsion will be dominated by coronal radiation and the red layer by radiation from cooler

material. The cool material is in ceaseless motion. It is easy to detect changes during totality. The same does not hold for the ionized corona, which is fixed by the magnetic field. It is not likely to show any variations during totality, except the extremely low amplitude oscillations that are predicted by certain theories.

To obtain the maximum coverage, observations should begin several seconds before 2nd contact, when most of the corona is visible on the opposite side of the Sun, to a few seconds after 3rd contact. If the shutter release mechanism does not allow repeated exposures to be made at a sufficiently high rate, it is preferable to wait until the disappearance of the last Baily's Bead at 2nd contact before making the first exposure. The success depends on the care with which the photographic equipment has been set up. Ensure that the image is of high quality and has an adequate scale, so that the resolution will not be too dependent on the film's grain size. A long focal length is also required, preferably at least 5 m. A focal length of 10 m with 13 × 18 sheet film will give better results. A simple achromatic doublet will give better results because the focal ratio is so small (around $f/100$). With fine-grain film, exposure times between 1/4 and 1 second are required. This means that guiding must be accurate if second-of-arc resolution is to be preserved.

10.6 PHOTOGRAPHING THE K (OR PLASMA) CORONA

We come to the field involving study of the outermost region of the solar atmosphere; the K or plasma corona, the region that consists of completely ionized material. It contains mainly protons and electrons trapped by the Sun's overall magnetic field lines. This also entails investigation of mass loss from the Sun and of its counterpart in the interplanetary medium, the solar wind. Fig. 10.8 shows the principal features involved; very narrow features, like the exceptionally long coronal steamers, which fade at several solar radii into another, tenuous corona. Light arises from scattering of solar radiation by free electrons in the ionized gas (Thomson scattering). This radiation is so polarized and measurements of the degree of polarization enable the location of the features in space to be determined and enable them to be seen out to greater distances. The shape of the corona is of great interest as the shape depends on the distribution of the magnetic field lines. In fact, this shape is difficult to determine during an eclipse, as all that can be seen during an eclipse is the projection of the various features around the Sun onto the plane of the sky. Different features appear superimposed on one another along the line of sight. The features may, in fact, be intertwined. Most features do have one predominant direction, and this is generally radial, but they are also nearly always curved to a greater or lesser extent and many even form arches. These arches are difficult to detect. They demonstrate the existence of greatly extended magnetic field lines, the bases of which lie in the deeper layers of the Sun. It is possible to

observe transient features, large-scale mass-ejections, which resemble enormous bubbles or loops the size of the Sun, which travel through the corona at speeds of about 400 km/s. Such events occur when the Sun is at sunspot maximum, and may then account for as much as 15% of the mass loss that occurs through coronal expansion. Observations of the corona, particularly in x-rays, in front of the solar disk have established a direct relationship between coronal features and measurements of the solar wind. Regions of the solar atmosphere that are least active most closely correspond to the high-speed component of the solar wind. These regions, known as *coronal holes*, coincide with very faint regions on eclipse photographs.

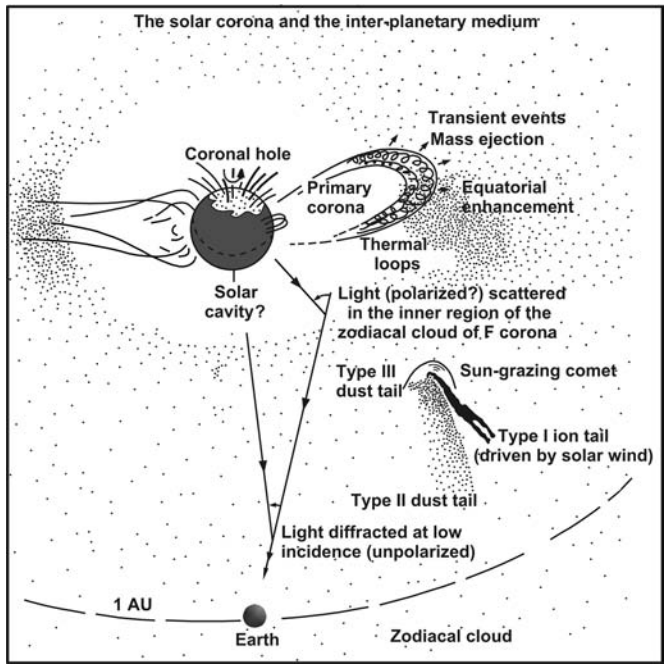


Figure 10.8: Schematic representation of the solar corona and its extensions into the interplanetary medium.

10.7 METHODS OF PHOTOGRAPHING THE PLASMA CORONA

Interesting studies may be carried out with modest equipment, if the film is calibrated correctly. By making certain assumptions we may deduce the total mass of the corona at the time of observation. Also, the records may be used to create a series of isophotes from which the degree of flattening of the plasma corona may

be established along with drawings of specific structures. Elaborate methods have also been developed to carry out precise photometry of coronal structures.

10.7.1 Wide-field Photographs

It is possible to carry out wide-field photography with just a fixed mounting, without having to compensate for diurnal rotation. With a 1000-ISO film or faster and an aperture of $f/5.6$, for example, typical exposure times are between 1/1000 and 1/15 seconds. Care should be taken to frame the corona properly and to use a focal length of at least 300 mm to avoid the damaging effects of vignetting. Using a longer-focus lens and a 6 × 6 camera will give results with even better resolution, but not necessarily with greater photometric accuracy. Sensitometry calibration should be carried out on the same film before and after the eclipse. Photographs should be greyscale, and uniformly illuminated by sunlight reduced by a factor of about 50,000. To obtain a proper calibration, it is preferable to construct an *absolute* sensitometer (i.e., one without any optics).

10.7.2 High-resolution Photography

The problem posed by obtaining images of the white-light, plasma corona is more complicated as the field to be covered is much greater. In addition, the exposure times required to photograph the outermost features are 100 to 1000 times as long for a given aperture and film speed. Such observations require particularly accurate guiding. A special photographic camera is shown in Fig. 10.9. This camera (essentially an eclipse coronagraph) is fed by an astrograph-type, a photographic objective of fairly large diameter and with a focal length longer than one meter. This camera could be supported on an equatorial mounting provided this was solid enough to track the solar corona. The enormous dynamic range of luminosity

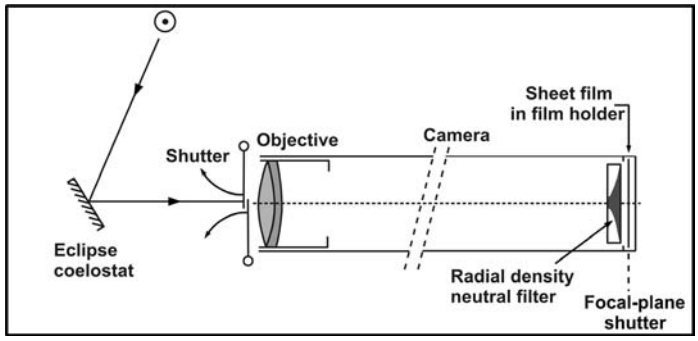


Figure 10.9: A schematic diagram for an eclipse experiment designed to photograph the features in the solar corona with a high spatial resolution.

in the corona gives rise to a strong radial intensity gradient. To avoid gross over-exposure of the inner corona, a radial mask or a radial neutral-density filter is required. The structures within the corona exhibit a very low contrast.

10.8 SPECTROSCOPIC STUDY OF THE CORONA

Different experiments are possible according to whether one is most interested in the corona, the transition region (i.e., with its 'flash' spectrum) the dust corona in searching new lines, in simply identifying them, or even in measuring specific thermodynamic properties. Fig. 9.10 shows an example of the arrangement of the various components in an experiment designed to observe part of the coronal spectrum. To capture the flash spectrum of the chromosphere, the widened entry slit should be placed parallel to the limb of the Moon, first at 2nd contact, and then toward the end of totality, at the 3rd contact position. They should also be taken in succession to obtain good spatial resolution. Photographs of the area being studied may be obtained at the same time as the spectrograms by capturing the image reflected from the front side of the slit. The plane diffraction grating may be replaced by a concave grating. It is also possible to replace the mirrors (objective and collimator) with lenses and to add components to correct aberrations to make the final part of the system more efficient. Spectroscopic analysis of the solar corona should preferably be made of a bright area of the inner corona. To get maximum coverage of the spectrum during totality, it is possible to make slight adjustments to the grating. It is also possible to record the whole spectrum at once, but with a lower spectral resolution. To identify the lines and to determine their positions, it is necessary to use the same system to photograph lines from a reference source.

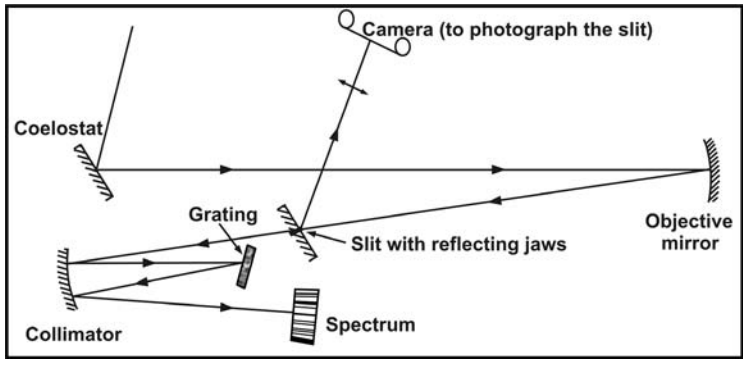


Figure 10.10: An example of the arrangement of the various elements in an eclipse experiment designed to carry out spectroscopic analysis of the solar corona.

EXERCISES

1. Draw a diagram showing the Earth and the Moon in space illuminated by the rays of the Sun, and then explain the circumstances of eclipses.
2. Show and explain the various phases of a total eclipse (for the measurement of solar diameter).
3. What are the required conditions for any eclipse to be total?
4. Draw a schematic diagram of an experimental design to measure the second and third contacts of an eclipse, and then interpret briefly.
5. What cautions should be taken for the observation of corona during the totality of an eclipse?
6. Explain why the solar corona may be observed during totality.
7. Show the main structures in the inner corona and the chromospheric extensions over the various quiet or active areas, and discuss.
8. To study the photosphere why is it preferable to work in the red and near infra-red?
9. Discuss briefly the method of observing the lunar corona. Explain how the photographing of the inner corona is done.
10. Explain with a suitable diagram how the solar corona may be extended to a interplanetary medium.
11. Describe the method and equipment for photographing the plasma corona, pointing out the calibration to be made.
12. Why is the photograph of a plasma corona complicated? What type of special photographic camera is recommended for the purpose?
13. Discuss the arrangement of the various components in an experiment to observe part of the coronal spectrum.

Chapter 11

MEASUREMENT OF SOLAR FLUX DENSITY

11.1 INTRODUCTION

When solar radio waves arrive at the Earth, one can measure the following properties: (*i*) *flux density* (i.e., intensity of the solar radiation) at various single frequencies, (*ii*) *frequency spectrum*, (*iii*) *polarization*, and (*iv*) *source brightness distribution* at one or more frequencies. Due to the transient nature of many solar phenomena it is necessary to observe and record these parameters simultaneously and continuously. The CSIRO observatory at Culgoora has comprehensive facilities for ground-based meter-wavelength solar observations. Its main instruments are:

(*i*) Radiometers for measuring flux density at 43, 80, 160, and 327 MHz.

(*ii*) Two-antenna short-baseline interferometers for radiometer calibration and for quiet-Sun flux density measurements.

(*iii*) Scanning radio-spectrograph, recording from 8 to 8000 MHz the dynamic spectrum. The recording is on 70-mm film and on a facsimile paper chart.

(*iv*) Acousto-optical radio-spectrograph, recording the dynamic spectrum of very weak bursts in selected frequency bands between 30 and 335 MHz.

(*v*) Spectro-polarimeter, recording on color films the dynamic polarization spectrum over the range 25 to 667 MHz.

(*vi*) Radio-heliograph, producing and recording radio *pictures*.

11.2 ANTENNAS FOR SOLAR OBSERVATIONS

In this section first we will describe the basic principles of antennas as applicable to radio astronomy. Next, we will discuss the antenna requirements from the point of solar observations.

11.2.1 Basic Principles for Using Antennas

The performance of a transmitting antenna is specified by the flux density of the radiated field at a distant point when a given power is supplied to the antenna terminals. If the input power W (in watts) produces a flux density F (in watts per square meters) at distance r (in meters) then,

$$F = \frac{GW}{4\pi r^2}. \tag{11.1}$$

The constant of proportionality G is the gain of the antenna. G is a function of the direction of transmission and has a maximum value G_0 along the axis of the main antenna beam.

In receiving antennas, including radio telescopes, it is more appropriate to specify its properties in terms of the effective area. If w is the power received by an antenna of effective area A (in square meters) when exposed to a plane wave with flux density F, then

$$w = AF. \tag{11.2}$$

The effective area is a function of direction and is proportional to the corresponding gain

$$G = \frac{4\pi A}{\lambda^2}. \tag{11.3}$$

Here, λ is the wavelength.

In solar radio astronomy we are concerned with sources of wideband *noise-like* radiation. The received power w can be specified by equating it to the thermal noise power obtained by replacing the antenna with a resistor at the antenna temperature T_a and is expressed as,

$$w = K T_a \Delta f. \tag{11.4}$$

(K = Boltzmann's constant and Δf = the receiving bandwidth.)

For a noise source, we may define the flux density per unit bandwidth at the receiving antenna as S (in W m^{-2} Hz^{-1}), so that

$$F = S \Delta f. \tag{11.5}$$

It is seen from Eq. (11.2) through (11.5) that, if the receiving aperture area is A perpendicular to the direction of the source and the source is small compared to the antenna beam width, then

$$S A = K T. \tag{11.6}$$

The received power is obtained by the angular distribution of brightness in the source. For each element of the source, the brightness temperature T_B is defined as the temperature of a black body for which the brightness of the thermal radiation would equal the brightness observed actually. T_B may change with frequency, but, like S, it is assumed to be constant over the receiving bandwidth Δf. The power received from a source element subtending a solid angle $d\Omega$ is

$$dw = \frac{KT_B A}{\lambda^2} \Delta f \, d\Omega. \tag{11.7}$$

Hence, the total received power is given by

$$w = \frac{K \Delta f}{\lambda^2} \iint_{4\pi} T_B \, A \, d\Omega. \tag{11.8}$$

These equations are used to find expressions for the signal received from a uniformly bright source of angular size considerably less than that of the antenna beam. The source is assumed to be located on the axis of the beam, where the effective area has its maximum value A_0. Then

$$S = \frac{K T_B \Omega}{\lambda^2} \text{ (where } \Omega = \text{angular size)} \tag{11.9}$$

and

$$T_a = \frac{A_0 T_B \Omega}{\lambda^2}. \tag{11.10}$$

For the special case of a circular source of angular diameter d,

$$S = \frac{K T_B}{\lambda^2} \frac{\pi d^2}{4} \tag{11.11}$$

$$T_a = \frac{A_0 T_B}{\lambda^2} \frac{\pi d^2}{4}. \tag{11.12}$$

If an antenna is surrounded by a black-body enclosure at uniform temperature T_B it can be shown by a thermodynamic argument that

$$T_a = T_B. \tag{11.13}$$

Hence, we can write,

$$\iint_{4\pi} A \, d\Omega = \lambda^2. \tag{11.14}$$

Eq. (11.14) shows a relation between the antenna beam-width to the effective area (or to the gain). Therefore, for beams of circular cross-section we have,

$$A_0 = k \frac{4\lambda^2}{\pi \alpha^2} \tag{11.15}$$

and

$$G_0 = \frac{16k}{\alpha^2}. \tag{11.16}$$

Here, α is the half-power beam-width, k is a constant for a given beam shape, with $1 > k > 0$.

11.2.2 Antenna Requirements for Solar Observations

Antenna requirements for solar observations are summarized as follows:

(*i*) The antenna-receiver combination must give a high ratio of solar signal-to-noise ratio from all other sources in order to enable measurements to be made with the desired accuracy.

(*ii*) The response should be approximately uniform over the whole angular extent of the radio Sun.

(*iii*) The antenna beam must be steerable (either mechanically by turning the antenna or electrically by phase adjustments) so that it may be pointed at the Sun throughout the period of observation.

(*iv*) The sensitivity to the signals arriving in directions away from the Sun should be kept low to minimize interference and other spurious effects, particularly those due to reflections at the ground.

These specifications are interrelated and at the same time conflicting. The sensitivity requirement implies a wide receiving area and a narrow antenna beam.

The need for uniform response sets a minimum beam-width and thus limits the receiving area. The absolute sensitivity of the receiving system is obtained by the relation between the *wanted* signal from the Sun and the *unwanted* statistical fluctuations owing to noise from the bright sky within the antenna beam and from the receiver itself. This noise component is expressed as an equivalent antenna temperature T_N. The receiver contribution to T_N at meter wavelengths is typically of the order of 300 K, since the sky brightness temperature T_{SK} is usually uniform over a solid angle greater than the antenna beam. The effect of the sky radiation is to increase the antenna temperature by an amount T_{SK}. Sky temperatures are shown in Fig. 11.1. The figure shows the following: (*a*) upper limit and (*b*) lower limit of sky noise component; (*c*) typical receiver noise, expressed as an equivalent antenna temperature; (*d*) and antenna temperature corresponding to quiet Sun at sunspot minimum, for an antenna with an effective area of 20 m.

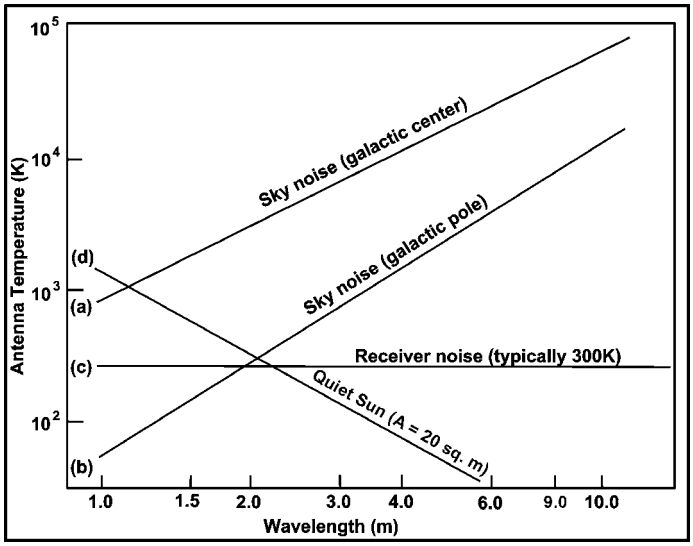


Figure 11.1: The antenna temperature in a solar radiometer.

In wideband antennas for radio-spectrographs there is an unavoidable compromise between uniformity of response at the shortest wavelength and sensitivity at the longest wavelength.

The ground-reflection problem is common to all mechanically steerable radio telescopes. This can be minimized by careful design (e.g., paraboloid reflectors should be designed to minimize side-lobe responses, even at some sacrifice of efficiency). A substantial improvement can also be obtained by using vertical rather than horizontal polarization. The reflectance of the ground at large incident angles (i.e., at low solar elevation) is smaller for vertically than for horizontally polarized

waves. At elevation angles less than about 40° the vertically polarized ground reflection is from 3 to 20 dB weaker than the horizontally polarized component.

11.3 PARABOLOID-REFLECTOR ANTENNAS

A convenient form for a solar radio telescope antenna at wavelengths from 1 to 5 meters is a steerable paraboloid reflector, illuminated by a *feed* antenna located at the focus. The polar diagram and receiving area of this type of antenna depend on the diameter of the aperture and also on the distribution of illumination over the aperture. This is determined by the radiation pattern of the feed. In order to produce a polar diagram with low side-lobes it is usual to design the feed so that the illumination at the edge of the reflector is about 10 dB less than the central maximum value. When this is done, the effective area of a reflector of diameter D is about $\pi D^2/8$ (i.e., the aperture efficiency is about 0.5 while the half-power beam-width is about $70\lambda/D$ degrees).

We know that it is required to reject ground-reflected waves and therefore it is advantageous to use a very deep reflector so that the edges of the reflector shield the feed from radiation incident at large angles to the main beam. Alternatively, a cylindrical *skirt* may be attached to the edge of a comparatively shallow reflector to achieve this result more easily. The long wavelength operating limit for a paraboloid-reflector antenna is set by interaction between the feed and the reflector. If D/λ is too small, this interaction leads to high side-lobe levels and to poor impedance characteristics caused by standing waves between feed and reflector. For satisfactory performance use a reflector with $D/\lambda \geq 3$.

At comparatively long wavelengths an open wire mesh rather than a continuous metal sheet can be used as a reflecting surface. These relaxations lead to savings in cost, weight, and wind-loading. A suitable meter-wave feed for a paraboloid is a dipole with a reflecting plate, or a simple two-element or three-element Yagi-Uda array. For broadband operation a log periodic feed is the appropriate choice.

11.4 BROADBAND ANTENNAS

For spectrographs and multi-frequency radiometers we need antennas that operate over a wide frequency range. In this context rhombic antennas used in the early spectrographs are of historical importance. They have now been entirely displaced by more modern designs, of which the Log-Periodic Antenna (LPA) is the most widely used. The LPA as shown in Fig. 11.2 is an end-fire array of dipoles, graduated in size and connected to the receiver by a common transmission line.

With a suitable choice of dimensions this array operates over a band slightly less than the range between the resonant frequencies of the longest and shortest dipoles. Bandwidths of from 3:1 to 10:1 are readily obtained. At any one frequency the antenna current is concentrated in the dipoles, which are close to resonance. The LPA exhibits almost constant gain (typically 5 to 10 dB) and beam-width over its operating band; the effective area is directly proportional to λ^2.

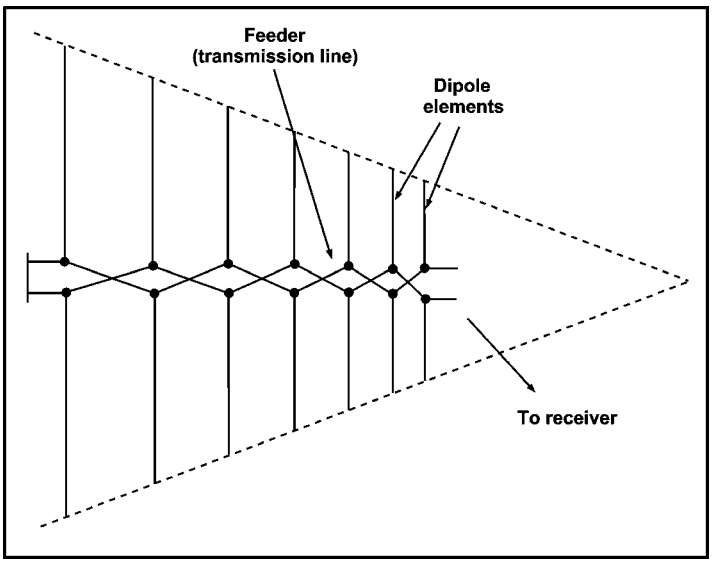


Figure 11.2: A log periodic dipole array. The spacing between adjacent dipoles is proportional to the dipole length. The feeder (twin-wire transmission line) is transposed as indicated between successive dipoles to provide the required phase relations. The feeder is terminated by a short-circuit about $\lambda_{max}/8$ beyond the longest dipole, where λ_{max} is the longest operating wavelength.

The conical helix is a wire helix wound to fit a conical surface. It is a broadband antenna somewhat analogous to the LPA. The receiver is connected to the apex of the helix. The response is circularly polarized corresponding to that of the winding. Antennas of this type have been used as the elements of the Clark Lake radio-heliograph and the operating wavelength range is 2.5 to 15 m. At wavelengths up to a few meters a paraboloid reflector with an LPA as its feed is a good broadband receiving antenna. The effective area is constant but the beam-width is proportional to wavelength. The position of the phase center of the LPA feed varies with wavelength causing defocusing. However, the resulting deterioration in the antenna pattern can, in general, be made very small.

11.5 DIPOLE ARRAYS

At wavelengths larger than about 5 m paraboloid-reflector antennas suitable for solar observations are very costly. Arrays of dipoles in various forms are more practical as receiving antennas. A single log-periodic antenna is used when fairly low gain and directivity are acceptable. In order to provide a narrower antenna beam and to reduce the unwanted sky radiation component in the received signal, arrays of several LPAs may be suitably used. The resultant polar diagram is then the product of the interference pattern of the array and the polar diagram of an individual LPA. The 4-m to 12-m antenna for the Culgoora spectrograph uses a pair of LPAs. The structure is mechanically steered to point the beam at the Sun. Fixed arrays of closely spaced horizontal dipoles supported above a reflecting ground plane give a large receiving area and narrow beam-width. The beam can be steered by adjusting the relative phases of the elements. This arrangement has the advantage of freedom from ground-reflection errors, since the image of the dipoles in the ground plane form part of the array. At large zenith angles the receiving area is reduced by foreshortening. So the sensitivity and beam-width deteriorate as the Sun approaches the horizon. An antenna that can be regarded as a form of dipole array is the corner reflector. In this case, a single dipole is supported in the aperture of a V-shaped reflector consisting of two plane surfaces. The beam is formed by interference between the fields due to the current in the feed dipole and the images of this current in the reflecting surface. The number of images is determined by the angle between the reflectors. Corner reflectors have been used as radiometer antennas and also as elements in a large array at $\lambda = 7$ m.

11.6 RADIOMETERS

Radiometers are designed for measuring accurately the total solar flux density at a single frequency. At meter wavelengths this measurement presents problems not encountered at shorter wavelengths. The flux from the quiet Sun is smaller, the background sky brightness is greater, and the receiving antennas required are larger. The high fluxes encountered during solar bursts also introduce problems of calibration and dynamic range.

The parameters involved are shown in Fig. 11.3. This shows typical flux densities for quiet Sun and for large solar bursts. Maximum and minimum galactic background flux densities are also revealed. The variation of the background radiation around the ecliptic at 80 MHz is shown in Fig. 11.4. In the presence of the Sun this background is difficult to measure as it varies with a scale size similar to the beam-widths of typical antennas. Other characteristics of solar bursts affecting the design of radiometer receivers are the rapid variation of flux with

both time and frequency. Rise rate as high as 10^4 solar flux units per second are sometimes noticed. Time constants and sampling intervals between 1 second and 1 millisecond are usually chosen. Good rejection of signals is also important as instantaneous variations in flux up to 30 dB in a few megahertz span may also occur. Individual bursts have bandwidths in the range 0.1 to 100 MHz or more. The receiver bandwidth is thus a compromise between sensitivity and the need to measure the true flux of narrow-band features. Bandwidths from 0.1 to 10 MHz are used in general.

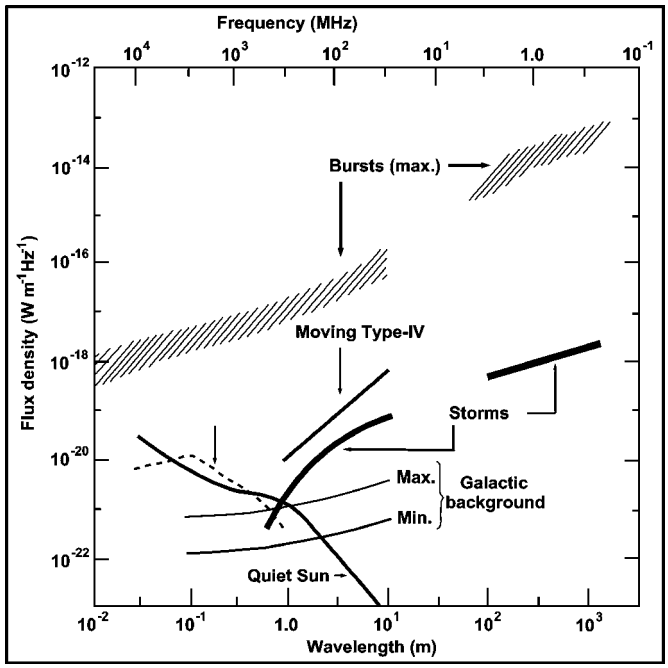


Figure 11.3: The flux densities of the quiet Sun (at sunspot minimum) and of large solar bursts of several types. The range of the galactic background flux is also shown (20° beam-width is assumed).

11.7 DYNAMIC RANGE

In order to accommodate a wide range of fluxes encountered between quiet Sun and intense bursts, radiometer receivers need dynamic ranges of ≈ 40 dB. The recording and presentation of data require special attention. If the range of fluxes is 40 dB and the lowest flux is recorded with a precision of 1%, then a total recording range of 10^6:1 is required if a proportional response is used. Even with digital recording

an elaborate system using 20-bit binary numbers would be needed to give the precision over the whole range. One method of compressing the output range is to use a form of automatic gain control. In it the receiver gain is automatically reduced in one or more steps when the input signal exceeds predetermined levels. This method is unsuitable if direct analog recording, e.g., chart or film, is used. The automatic gain changes make such records almost indecipherable when rapidly varying bursts are recorded. Alternatively, two separate receiving and recording channels can be used. One of these records the weaker signals accurately while the latter are recorded on the second channel of much lower gain.

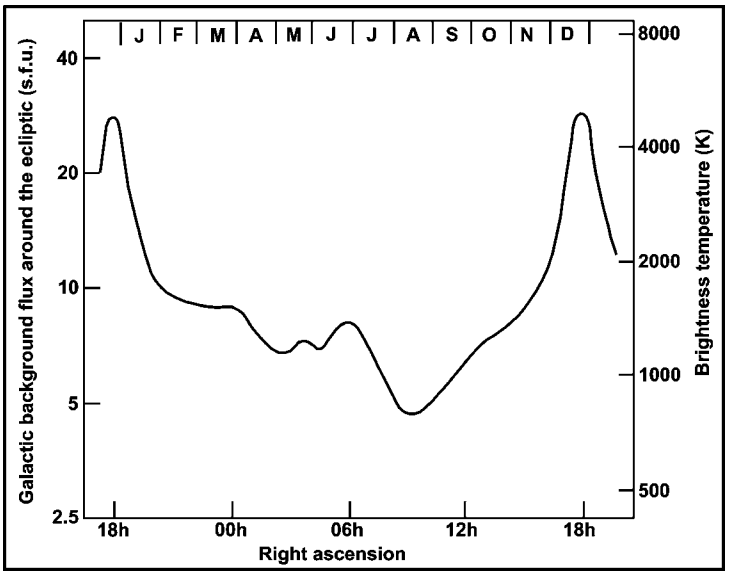


Figure 11.4: Galactic background flux and brightness temperature measured around the ecliptic at 80 MHz. Both measurements are averaged over a 20° beam-width. The position is given as right ascension on the ecliptic; the upper horizontal scale indicates the right ascension of the Sun month by month.

Very high precision is not required for the measurement of bursts. For these the recording scale can be greatly compressed by using a receiver with logarithmic response (i.e., one with the output voltage proportional to the logarithm of the input power). Since high precision is required for quiet-Sun signal levels, a suitable compromise is to use a receiver response that follows a steep logarithmic law at low signal levels and changes progressively to a flatter one as the signal increases. Such a response is ideal for display purposes. It has a minor disadvantage of requiring a complicated step-by-step calibration procedure, as shown in Fig. 11.5.

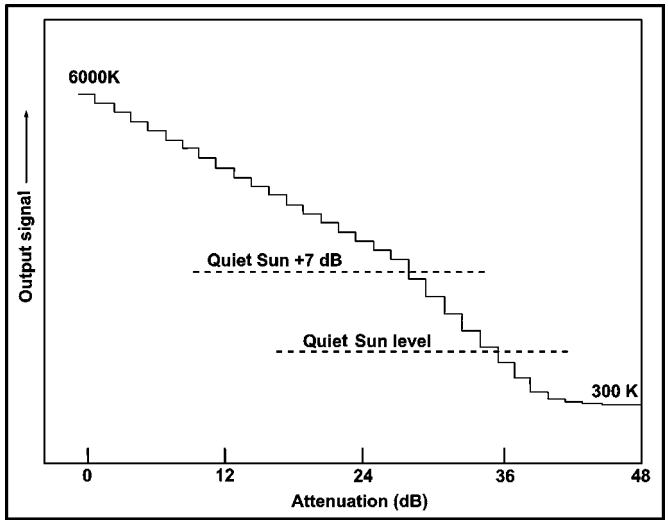


Figure 11.5: A typical calibration for the Culgoora radiometers, which employ dual-slope logarithmic responses. The output of the 6×10^6 K Zener diode noise generator is attenuated in steps of 1.5 dB.

11.7.1 Calibration

To convert radiometer output signals into absolute values of flux density a calibration of the instrument is required. This can be done by comparing the output signal produced by the Sun with that obtained when the antenna input is replaced by the signal from an absolutely calibrated noise generator. This method requires accurate calibration of the noise source and also the measurement or calculation of the effective area of the antenna. Both these intermediate steps must be repeated periodically to guard against errors due to any changes in the equipment parameters so that this *absolute* calibration method is laborious and subject to error. It is more reliable to use as a calibrator one of the strong cosmic sources of small angular size for which the flux density has been accurately determined. Usually it is not possible to observe the cosmic sources during the solar observing period; the Sun is much stronger than the calibration sources, so that solar radiation picked up in the antenna side-lobes prevents an accurate daytime measurement of the calibrator. It is therefore better to use an artificial noise generator as a secondary standard, comparing its output first with the calibrator and then with the Sun. When this is done, it is necessary for the noise generator to be stable in output over periods of hours or days; it need not be calibrated absolutely.

If an artificial noise generator is used as the primary calibration standard for a radiometer with a linearly polarized receiving antenna, the calculated solar flux

must be doubled to allow for the rejected polarization component. This correction is, of course, not to be applied when the primary standard is a cosmic radio source. The widely used and by far the most intense calibration source is Cassiopeia-A. This is a recent supernova remnant. Its intensity is decreasing at about 2% per year. The flux density and its secular variation have been accurately determined. Accepted values are:

$$\left.\begin{aligned}\log S_{\text{CasA}} = {} & 1.745 - 0.0080(t - t_0) \\ & - \{\, 0.770 - 0.00126(\, t - t_0)\,\} \log f \\ \text{for } 200 \text{ MHz} \le {} & f \le 30 \text{ GHz}\end{aligned}\right\} \tag{11.17}$$

$$\left.\begin{aligned}\log S_{\text{CasA}} = {} & 1.540 - 0.0080(t - t_0) \\ & - \{\, 0.681\text{-}0.00126(t - t_0)\,\} \log f \\ \text{for } 20 \text{ MHz} \le {} & f \le 200 \text{ MHz}\end{aligned}\right\} \tag{11.18}$$

where S is in solar flux units and f in MHz; t is measured in years and t_0 is 1980.

The next strongest source is Cygnus-A. This source has a curved spectrum, which is fitted at meter wavelength by the expression:

$$\log S_{\text{CygA}} = 0.695 + 0.085 \log f - 0.178 \log^2 f. \tag{11.19}$$

Both Cassiopeia-A and Cygnus-A are far North (declinations +58° and +40°, respectively) and are useful as calibrators only for observatories in the Northern hemisphere. For Southern latitudes suitable sources are Taurus-A, Virgo-A, and Hercules-A. At meter wavelengths the expressions are satisfactory:

$$\left.\begin{aligned}\log S_{\text{TauA}} &= -0.182 - 0.27 \log f,\ f > 30 \text{ MHz} \\ \log S_{\text{TauA}} &= 0.650 - 0.83 \log f,\ f < 30 \text{ MHz}\end{aligned}\right\}. \tag{11.20}$$

The spectra of both Hercules-A and Virgo-A maintain fairly constant slope. The flux density at 178 MHz has been used, together with the slope between 38 and 750 MHz:

$$\left.\begin{aligned}\log S_{\text{HerA}} &= 0.796 - 1.008 \log f \\ \log S_{\text{VirA}} &= 0.799 - 0.790 \log f\end{aligned}\right\}. \tag{11.21}$$

The problems of accurately correcting for sky noise are more serious with weak cosmic sources than with the Sun. This means that total power observations of cosmic sources are simply not possible with solar radiometers. An alternative is to determine the effective area of the antenna by observing cosmic sources using two identical antennas as an interferometer. This can be done even when the source flux densities are very much less than the background sky flux. Provided the antenna system is stable the measured effective area can then be used to determine solar flux densities when one of the antennas is used alone as a total power radiometer.

Two types of artificial noise source are frequently used at meter wavelengths: (*i*) Thermionic diode and (*ii*) Zener diode noise sources.

11.8 SPECTROMETRY

A radio-spectrograph is an instrument used for recording the spectrum of radiation at radio frequencies. In radio astronomy, radio-spectrographs are important in solar observations and also in the study of atomic and molecular spectral lines. The requirements and techniques are different in the two cases. Since the solar spectrum varies with time a solar radio-spectrograph must record the dynamic spectrum (i.e., the solar activity as a function of both frequency and time). It is almost universal practice to present the data whether on film, as a facsimile record, or as some form of contour map as a plot of the intensity on a frequency-time plane.

11.9 CLASSIFICATION OF RADIO-SPECTROGRAPH

Three types of radio-spectrographs can be recognized. The first group consists of scanning instruments, in which the output of a single narrow bandwidth receiver is recorded while the receiving frequency is rapidly swept across the frequency range to be observed. Second is the multi-channel spectrograph, consisting of numerous narrowband receivers each of which accepts one segment of the required frequency range, and the output of each receiver provides one point in the spectrum. In the Acousto-Optical Spectrograph (AOS), a more recent development, the broadband radio signal, which is to be analyzed is converted to an acoustic wave in a transducer cell. When coherent light is passed through this cell the first order diffraction pattern delineates the spectrum of the radio signal.

11.9.1 Scanning Radio-Spectrographs

These were the first solar radio-spectrographs. Here we will use, the Culgoora spectrograph as an example of a fairly elaborate scanning system. The Culgoora instrument records spectra over the continuous frequency band from 8 to 8,000 MHz. The frequency range of an individual scanning receiver is limited to less than 10:1 by the practical difficulties of designing antennas and pre-amplifiers for larger frequency spans. The full frequency band is subdivided into seven sub-ranges, each of about 3:1 in frequency. This is shown in Table 11.1. Five range units are used; each of the two highest frequency ranges is split into two sections, which are scanned in alternate scanning cycles. In the three meter-wavelength bands

(24 to 666 MHz) the dynamic range required is very large. The spectrum in these ranges is recorded at two sensitivity levels differing by 10 dB.

Table 11.1: The Culgoora scanning radio-spectrograph.

Freq. span (MHz)	Antenna	Receiver noise fig. (dB)	Bandwidth[12] (MHz)	Time per scan (seconds)
8-24	Horizontally polarized LPA	10	0.033	0.25
24-74	Two-element array of cross-polarized LPAs	6	0.1	0.25
74-222	13-m-diameter paraboloid, cross-polarized LPA feed	6	0.38	0.25
222-666	6-m-diameter paraboloid, cross-polarized	6	2	0.5
666-2000	LPA feed			0.5
2000-4000	1.8-m-diameter paraboloid, LH polarized	13	30	0.5
4000-8000	Conical-helix feed			0.5

Thus, a total of 10 spectra have to be displayed for photographic recording. Fig. 11.6 shows the components of a single range unit, while Fig. 11.7 shows the control system, which synchronizes the various scanning waveforms and organizes the sampling and display of the output signals. The receiving frequency is scanned by varying the local oscillator frequency in a super-heterodyne mixer. Up-converter mixers are best suited for broadband applications. They may be either low-noise parametric mixers or conventional mixers. The latter have poor noise figures, so that some pre-amplification of the input signal is required. Whatever arrangement is used, the frequency control waveform is shaped to give a suitable variation of receiving frequency with time. This waveform is chosen so as to make $(df/dt)/f$ approximately constant over the whole scanning range.

[12] In each frequency span, the spectrum is sampled at 200 points.

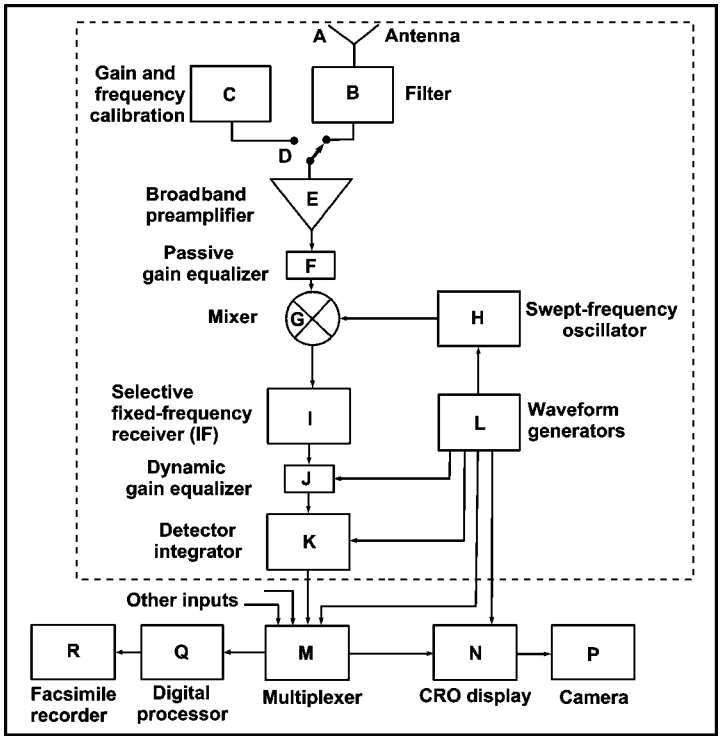


Figure 11.6: Block diagram of a scanning spectrograph showing components of a single range unit, A-L (in the dashed box) and the display and recording system, M-R used at Culgoora.

The scanning mixer (see G, Fig. 11.6) is followed by a fixed-frequency superheterodyne receiver (see I, Fig. 11.6) in which the final IF amplifier determines the instantaneous bandwidth (i.e., frequency resolution) of the system. The resolution for each receiver (see Table 11.1) is typically 1% to 2% of the corresponding scanning range.

11.9.2 Multi-channel Spectrographs

To observe weak solar bursts, the inherently low sensitivity of the scanning spectrograph is a serious limitation. One way of getting high sensitivity is to use a multi-channel instrument. This consists of a set of N narrowband radiometers, each tuned to a different fixed frequency so that together they provide continuous measurement of signal intensity at N points, which cover the required frequency range. The number N of data points is limited to the number of receiving channels. For this reason a scanning spectrograph is a more appropriate choice when a very wide frequency range is to be covered.

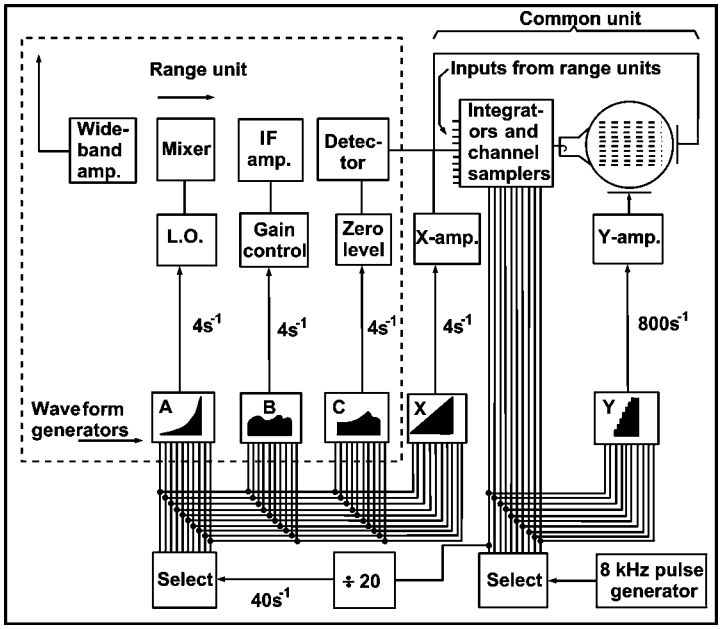


Figure 11.7: The Culgoora spectrograph showing in schematic form, which shows the various control waveforms applied to a single range unit (inside the dashed box) and the x and y waveforms applied to the CRO display. Other range units are multiplexed to the CRO display via the integrator and channel samplers shown. The separate spectra can be displayed on the screen in parallel; by use of a system of mirrors these spectra are arranged in a single line and in the correct frequency sequence to be photographed.

Several large multi-channel radio-spectrographs have been built and operated. The Nancay (France) radio-spectrograph shown in Fig. 11.8 operates in the frequency range 110 to 470 MHz. This range is subdivided by a binary *tree* of filters with some amplification to compensate for losses. The main amplification of the signals takes place at a stage where the frequency range has been subdivided into 36 bands each of 10 MHz bandwidth. This arrangement greatly reduces inter-modulation by strong interfering signals. Each of the 10 MHz-bandwidth amplifiers feeds a set of 10 mixers, with local oscillator frequencies, which differ from each other in steps of 1 MHz. This matrix of mixers produces 360 signal output, with the center frequencies effectively spaced at 1 MHz intervals over the 360 MHz input frequency range. A bank of amplifiers with 1 MHz bandwidth can be connected to any 120 of these mixer outputs. These amplifiers thus provide 120 points in a spectrum with 1 MHz resolution. A further frequency subdivision is available with the simultaneous use of 100 more ampliflers of 200 kHz bandwidth to provide higher frequency resolution in a restricted frequency range.

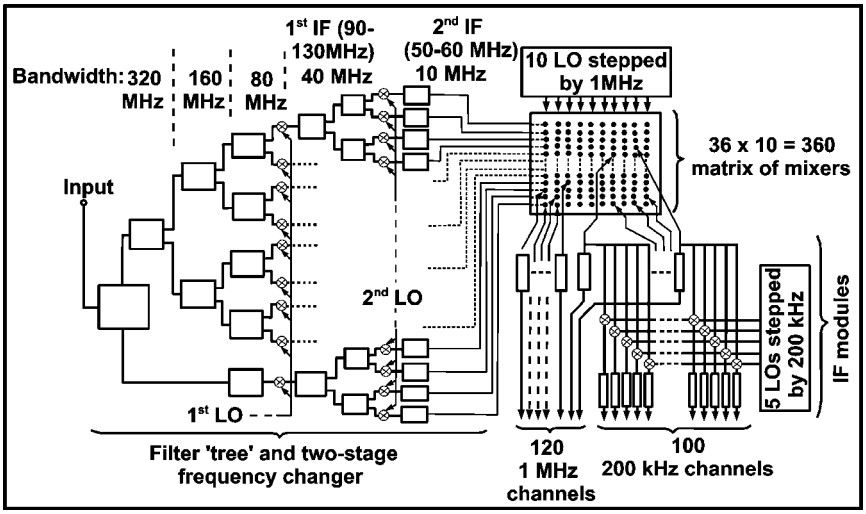


Figure 11.8: Diagram of the RF-section of the Nancay multi-channel radio-spectrograph.

11.9.3 Acousto-Optical Spectrograph (AOS)

The AOS, originally called the electro-optical spectrograph, uses the diffraction of coherent light waves to produce the spectrum of a broadband radio signal. The arrangement of a typical AOS is shown in Fig. 11.9. Light from a laser is expanded into a parallel beam, which passes through a transparent modulator containing traveling ultrasonic pressure waves. The latter are generated from the amplified broadband radio signal by a piezo-electric transducer bonded to one end of the modulator cell. The density variations due to the ultrasonic waves produce phase variations in the light so that the modulator acts as a phase diffraction grating. The angle to the first-order diffraction pattern of the coherent light is proportional to the radio frequency applied to the transducer. The light intensity in this pattern is proportional to the input signal intensity. The first-order diffraction pattern is the spectrum of the radio signal. It can be focused by a lens for display or recording. The spectrum at any instant is that of the time segment T of the signal, which is at that instant within the modulator, so that the frequency resolution is 1/T.

An AOS has been in operation at Culgoora to record the spectra of weak bursts, which can be recorded by the radio-heliograph, but not by the comparatively insensitive wide-range scanning spectrograph. This AOS has an 80 MHz frequency range and a frequency resolution of 160 kHz.

The AOS combines high sensitivity and resolution with a moderately large frequency range. Improvements can be expected in the critical optical components.

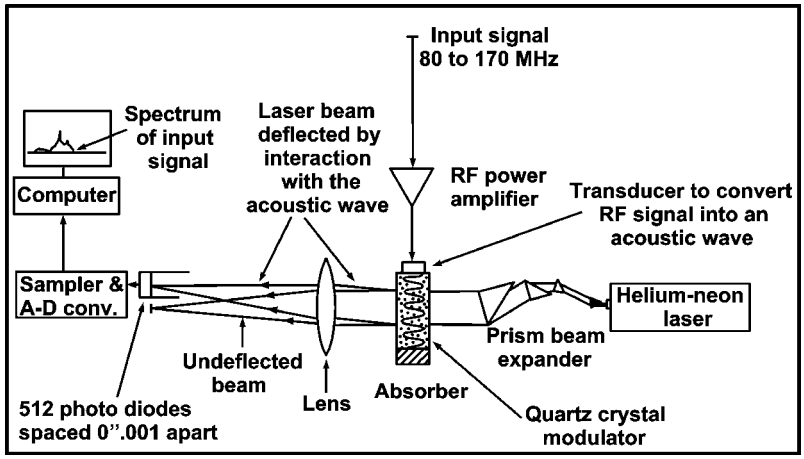


Figure 11.9: A schematic diagram of an acousto-optical radio-spectrograph.

For all practical purposes the available recording media for modern solar radio-spectrographs, with their high data rates, are photographic film, facsimile records, and magnetic tape.

11.10 RADIO-SPECTROGRAPH ANTENNAS

A suitable antenna for solar radio telescopes should have sufficient directivity to discriminate against non-solar radiation, while at the same time its beam must be wide enough to give uniform coverage to all parts of the radio Sun at the operating wavelength. In conjunction with the receiver to be used the antenna should have sufficient receiving area to detect the quiet Sun.

The threshold antenna gain G_t required to detect the quiet Sun in the presence of other noise sources is

$$G_t = \frac{4\pi(T_g + T_r)}{\Omega T_d \sqrt{\tau\, \Delta f}}. \tag{11.22}$$

Here, T_g is the sky noise temperature, T_r receiver noise temperature, T_d apparent disk temperature of the quiet Sun, Ω solid angle subtended by the Sun, τ integration time, and Δf receiver bandwidth.

Fig. 11.10 (*a*) shows the characteristics of reflectors of various diameters, and of several arrays of log-periodic antennas. Fig. 11.10 (*b*) is a performance diagram for the antennas actually used in the Culgoora spectrograph. Other antenna configurations can also be used. For example, the Harvard spectrograph used a

single 8.4-m paraboloid for the frequency ranges 100 to 580 MHz and 2000 to 4000 MHz, and a simple array for 25 to 100 MHz.

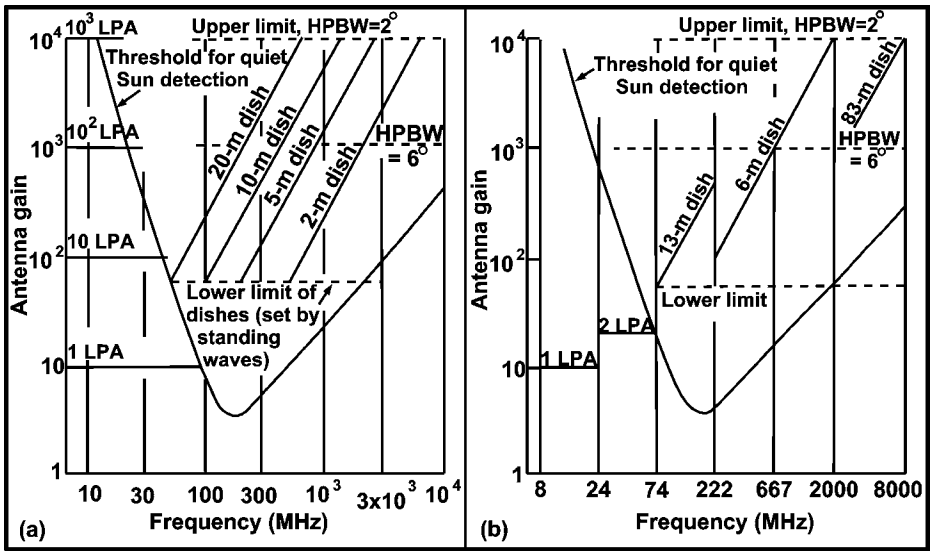


Figure 11.10: (a) Gain-versus-frequency characteristic for paraboloid dishes and various combinations of log-periodic antennas. (b) The antennas used for the 8 to 8000 MHz Culgoora radio-spectrograph compared with the threshold level of the quiet Sun.

EXERCISES

1. What are the properties that can be measured from solar radio waves arriving at the earth?
2. Mention the comprehensive facilities of the CSIRO observatory at Culgoora for ground-based meter wavelength solar observation.
3. Describe the basic principles of antennas as applicable to the radio astronomy.
4. What is 'noise-like' radiation?
5. Establish a relationship between the antenna beam-width to the effective area of the antenna.
6. Summarize the antenna requirements for solar observations.
7. The ground-reflection problem is common to all mechanically steerable radio telescopes. Explain how this can be minimized through a careful design.
8. Discuss why a steerable paraboloidal reflector is a convenient form for a solar radio telescope antenna at wavelengths from 1 to 5 meters.
9. What type of antennas is preferable for spectrographs and multi-frequency radiometers?

10. Give the principle of broadband antenna, clearly explaining the geometry. What is the utility of dipole arrays?
11. Discuss how radiometers are suitably used for measuring accurately the total solar flux density at a single frequency.
12. In order to accommodate a wide range of fluxes, what are the necessary modifications to be made in radiometer receivers?
13. To convert radiometer output signals into absolute values of flux density, a calibration of the instrument is required. Discuss how this can be done.
14. Describe a radio-spectrograph. Classify the three different types of radio-spectrographs.
15. Which was the first solar radio-spectrograph?
16. Mention the characteristics of antenna at different frequency bands of a radio-spectrograph.
17. Give the block diagrams of a scanning spectrograph showing the different components.
18. Explain how multi-channel spectrographs can be used to observe weak solar bursts.
19. What is an acousto-optical spectrograph? Draw a diagram to explain the principle.
20. What are the qualities to be maintained in an antenna for a radio telescope?
21. Show the gain vs. frequency characteristic for paraboloidal dishes and various combinations of log periodic antennas, and explain.

Chapter 12

Measurement of Solar Brightness Distribution

12.1 INTRODUCTION

Most solar features, particularly at meter wavelengths, develop and change rapidly so that high time resolution is necessary. The elegant but slow methods of Earth-rotation aperture synthesis are usually not applicable. In this chapter, we discuss radio-heliographs (i.e., radio telescopes that produce images of the Sun in real time and in rapid succession). To get useful information on the position and structure of localized radio sources on the Sun it is required to achieve an antenna beam-width no greater than a few minutes of arc. We need apertures of at least 1000 λ. At radio wavelengths longer than about 10 cm it is impracticable to build single filled-aperture antennas of sufficient size. At meter wavelengths, spaced arrays of many small antennas have been used to give the requisite overall dimensions. The elements of such arrays are directional, and for hour-angle coverage they must be steered during observations so that their beams are always pointed at the Sun. In many cases facilities are included for measuring the polarization of the radio sources by providing the antennas with crossed-dipole feeds and quadrature phase switches.

12.2 THEORY OF ANTENNA ARRAYS

In the following, we summarize the theoretical results needed for an understanding of the operating principles of radio-heliographs.

12.2.1 Aperture Illumination and Field Response Pattern

If an antenna is used for transmission, the current distribution over this aperture is specified by the grading $g(x, y)$. This is the complex amplitude of the current at any point (x, y) in the plane, normalized to a maximum value of unity.

The radiated field pattern V is given by the Fourier transform of the grading,

$$V(l, m) \propto FT\{g(x, y)\}, \tag{12.1}$$

where $FT(z)$ is the Fourier transform of z. The coordinates used are shown in Fig. 12.1, where $V(l, m)$ represents the field strength radiated in the direction specified by l and m, the direction cosines with respect to the X and Y axes. In terms of the polar coordinates θ and ϕ,

$$l = \sin\theta \sin\phi, m = \sin\theta \cos\phi.$$

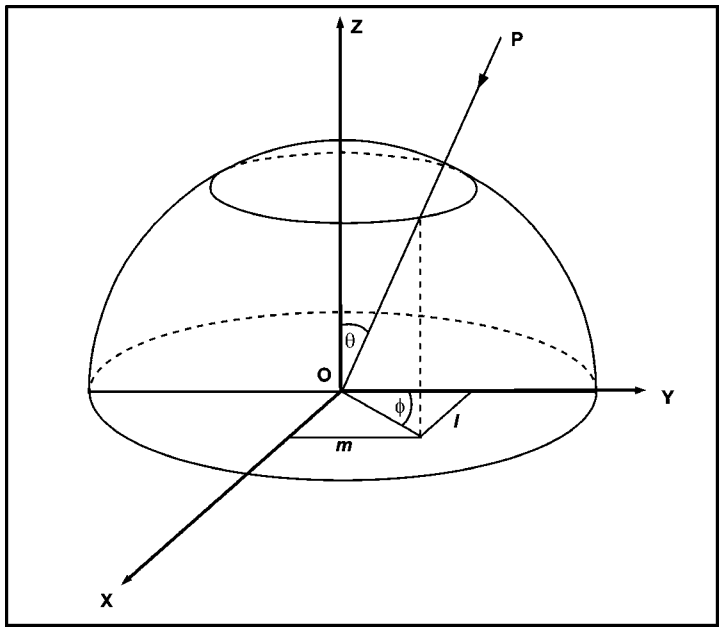


Figure 12.1: Coordinate system for the array geometry. The array is in the *x-y* plane. PO is the direction of the incident ray from the source; *l*, *m* are direction cosines of PO.

For a horizontal array, θ is the zenith distance and ϕ is the azimuth. It follows from the reciprocity theorem that the response pattern in reception is of the same form as the radiated field pattern in transmission. An example of a linear co-phased array of small antennas is shown in Fig. 12.2. The field pattern consists of multiple fan-shaped lobes. With a uniform grading the pattern exhibits large side-lobes, which can be reduced by altering the grading (i.e., changing the weighting of the fourier components). In this example the required reweighting can be produced by tapering the current distribution from the center to the ends of the array (Fig. 12.2 (*d*) and (*e*)).

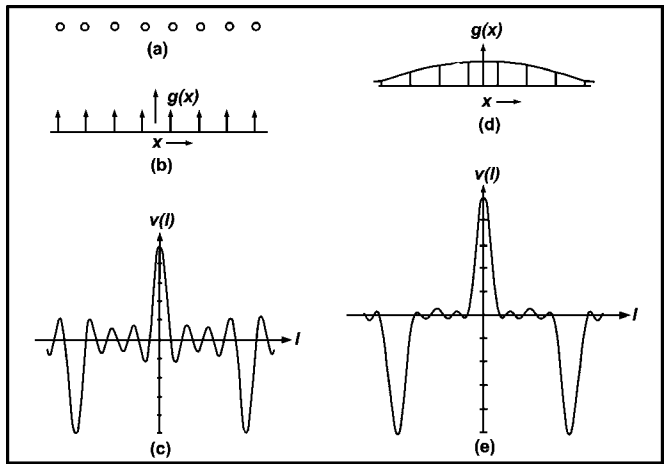


Figure 12.2: Grading and field response patterns for an array of eight small elements uniformly spaced along the x-axis: (a) uniform grading; (b) gives the pattern; (c) a suitably tapered grading; (d) gives a response; and (e) with low side-lobes. For nearly normal incidence the direction cosine *z* angle of incidence.

12.2.2 Power Polar Diagram and the Transfer Function

The power polar diagram of a receiving antenna is directly proportional to the square of the magnitude of the field pattern:

$$A(l, m) \propto |V(l, m)|^2, \tag{12.2}$$

where A is the receiving area in the plane (l, m). Then, using Eq. (12.1) and applying the autocorrelation theorem, we get,

$$FT\{A(l, m)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (g(x, y)\, g^*(x-u, y-v)\, dx\, dy = c(u, v), \tag{12.3}$$

where $c(u, v)$ is the transfer function of the antenna and is equal to the autocorrelation function of the grading and g^* is the complex conjugate of g. For an array of small spaced elements the grading g is zero everywhere in the x-y plane except at the location of each element. It is seen from Eq. (12.3) that the transfer function $c(u, v)$ is non-zero only at a set of discrete points in the u-v plane, each at a position where the u and v coordinates are respectively equal to the x- and y-components of the line joining a pair of the array elements. Fig. 12.3 shows the transfer function of an annular array with uniform grading.

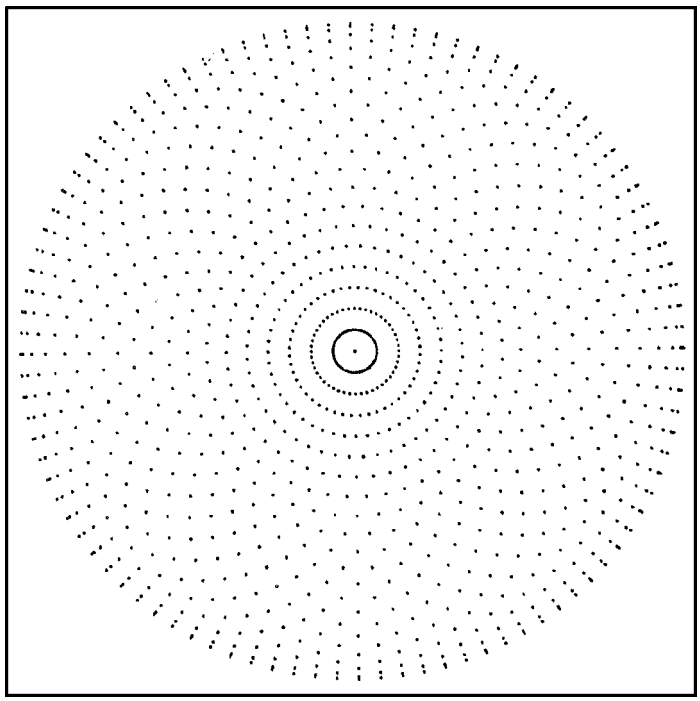


Figure 12.3: Transfer function of an annular array of 48 uniformly spaced elements.

12.3 IMAGE FORMATION

In order to produce an image of the sky brightness distribution we must record and display the signals corresponding to each of a raster of points in the required field of view. The sampling theorem shows that to avoid loss of information the angular separation of these sampling points should not exceed $\lambda/2D$, or about half the array beam-width.

12.3.1 Beam-scanning Methods

The common method of imaging is by pointing the array beam at each, in turn, of the required sampling points in the sky and recording the receiver output. For a one-dimensional array with a fan-beam response this is done by allowing the Earth's rotation to sweep the beam across the field. To get more adequate time resolution it is required to scan the beam rapidly across the field by introducing progressive phase-shifts among the array elements. Any scanning method has the disadvantage that where N sampling points are required each is only observed for $1/N$ of the total time. As a result the sensitivity is effectively reduced in the ratio $1{:}N^{1/2}$. This loss of sensitivity can be avoided by a multi-beam technique. In that technique the array elements are connected in N different phase combinations each of which directs the array beam at one of the N sampling points. Each output of the phasing network is connected to a separate receiver. This continuously measures the signal corresponding to one of the points in the field of view.

12.3.2 Acousto-optical Signal Processors

Another method of image formation uses AO (acousto-optical) techniques. For an array image processor, a set of AO cells is arranged as a scale model of the configuration of the elements in the array. Each cell is modulated by the radio frequency signal from its corresponding element. The model is placed in the aperture plane of a lens and is illuminated by a collimated laser light beam. The first-order diffracted wave from an AO cell reproduces in phase and amplitude the signal applied to its modulator. Thus, in the image plane of the lens the diffracted light waves from the array of cells combine vectorially to form an optical image of the observed radio brightness distribution.

12.3.3 Correlator Radio Telescopes

An alternative approach to image formation is to measure directly the correlations between the signals from each pair of elements in the array and to find the image from these correlation values. A measurement of the correlation between the signals received by a pair of antennas in an array provides the complex amplitude of one value of the spatial frequency spectrum of the power received from a distant source of incoherent radiation. The polar coordinates in the u-v plane of the point measured are given by the length and orientation of the baseline between the antennas. Consequently, its position in the plane is the same as that of the corresponding point in the array transfer function. If the complex correlation products are measured and mapped in the u-v plane for all the antenna pairs in the array, we get the spatial frequency spectrum of the response of the array to the sky brightness distribution. The response itself is the Fourier transform of

this spectrum and can be computed by one of the standard methods of numerical Fourier transformation.

The above method of image formation has some attractive features. As the Fourier components are separately accessible for processing, various operations such as the reweighting of components and the application of phase and amplitude correction can be carried out directly before performing the Fourier transform. All the correlations are simultaneously measured and thus no loss of sensitivity due to time-sharing occurs. In a typical receiving system each correlator has a circuit for forming the instantaneous product of the two signals at a sampling rate of the order of twice the radio frequency bandwidth. For solar observations the value is usually a few megahertz. For practical purposes two correlators are used for each pair of channels with a 90° difference in the relative phases of the input signals. From the two outputs the phase and magnitude of the complex correlation product can be achieved. Successive values of each correlation product are summed to improve the signal-to-noise value. At intervals equal to the time between successive images the integrated values are transferred to the device to perform the Fourier transform. Usually it is a general-purpose digital computer.

The correlator input signals are quantized by representing any positive value by 1 and any negative value by 0. This *one-bit* correlation procedure is valid for noise-like signals. After correction the final time-averaged result is the same as if exact values were used for the correlator inputs.

12.4 SOME TYPICAL RADIO-HELIOGRAPHS

12.4.1 Grating Arrays

A uniformly spaced linear array is similar to the plane diffraction grating in optics, and so is called the grating array. The power polar diagram of a grating array of N elements with spacing d between adjacent elements (Fig. 12.4) is expressed as,

$$A(\theta) = c(\theta)\frac{1}{N}\left\{\frac{\sin(\pi N d \sin(\theta/\lambda))}{\sin(\pi d \sin(\theta/\lambda))}\right\}^2, \tag{12.4}$$

where $c(\theta)$ is the power polar diagram of each of the elements and θ is the angle between the array from the source and the array normal. For an untapered array at small angles of incidence, $A(\theta)$ corresponds to a set of fan beams with half-power beam-width $0.88(N\text{-}1)\,\lambda/(ND)$ radians, where $D=(N\text{-}1)d$ is the overall length of the array. Adjacent beams are λ/D radians apart. With the increase of incident angle, both the beam-width and beam spacing increase in proportion to the foreshortening of the array.

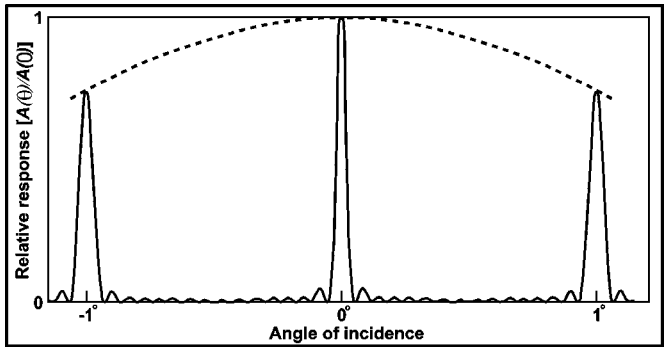


Figure 12.4: Power polar diagram of a uniformly fed grating array, with $N = 16$, $d = 57.2\ \lambda$. The broken line is the polar diagram of the array elements, which are directed toward the zenith.

For solar applications the beam spacing should be greater than the diameter of the radio Sun. Generally, beam spacing of the order of 1° has generally been used at wavelengths up to 1 m, so that the element spacing d is $\approx 60\ \lambda$. At longer meter wavelengths, where the diameter of the radio Sun is over 1°, spacings of 20 to 30 λ are more appropriate. Since angular resolution of a few minutes of arc is required for solar studies, the aperture D requires the order of 1,000 λ. These considerations determine the required number of elements which has been from 8 to 32 in grating array radio-heliographs.

An early meter-wavelength grating radio-heliograph was developed at Nancay, France. This East/West array was operated at 169 MHz and used 32 paraboloid antennas on a 1550-m (870 λ) baseline, giving 3′.8 fan beams, 2°.01 apart.

12.4.2 Circular Arrays

The Culgoora radio-heliograph uses annular antenna arrays instead of usual linear arrays. The array geometry is shown in Fig. 12.5. The main array for 80, 160, and 327 MHz uses 96 steerable 13-m paraboloids. For convenience, a separate array of 48 simple corner-reflector antennas is used at the lowest operating frequency, 43.25 MHz. In the power polar diagram (Fig. 12.6) the central part is a circularly symmetrical pencil beam similar to the response of a continuous annular aperture of the same diameter,

$$A(\theta) = J_0^2\left(\frac{2\pi a}{\lambda}\right). \tag{12.5}$$

Here, a is the array radius, θ is the angular distance from zenith, and $J_0(x)$ is the zero-order Bessel function of the first kind. As the array uses spaced elements, there are further grating responses. These are complicated ring patterns of radius $\theta_1, 2\theta_1, 3\theta_1, \ldots$, centered on the main pencil beam:

$$\theta_1 = \frac{\lambda}{d}. \tag{12.6}$$

Here, d represents the spacing between adjacent array elements.

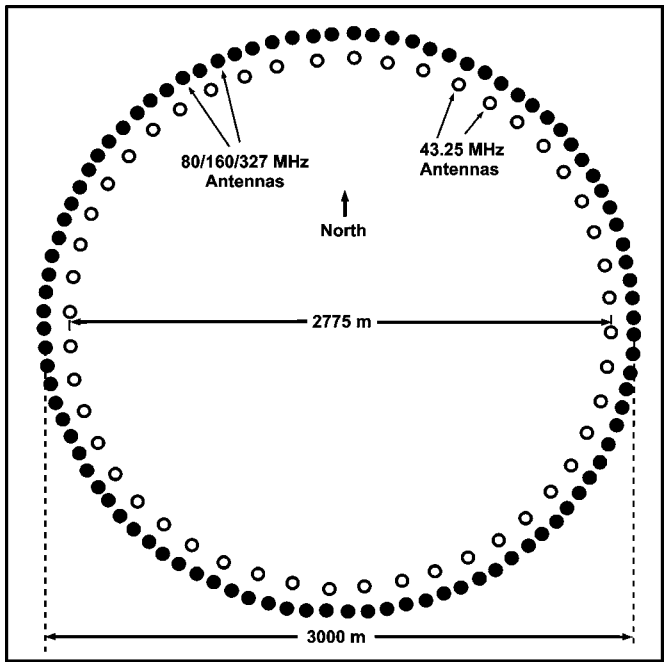


Figure 12.5: The Culgoora antenna arrays.

The dimensions of the array were chosen to give a field of view of approximately 2 clear of grating lobes, at the original operating frequency of 80 MHz. The fields and beam-widths now obtained at the four observing frequencies are shown in Fig. 12.7. At any one of these frequencies two images per second are built up, each with 60 × 48 picture points. A combination of multiple beams and phase scanning is used. A branching network forms a North/South row of 48 pencil-beam responses and these are shifted from East to West in 60 steps to form the picture. The two pictures revealing LH and RH circular polarization are produced quasi-simultaneously. The signals for each column of picture points are integrated for about 8 ms, then polarization switches and the antennas are reversed before the next 8-ms integrations. Thus, columns of points are recorded for each picture in turn. The operating frequency is changed from second to second in a sequence selected by the operator.

It is clear from Fig. 12.6 that the J_0^2 pencil beam has unacceptably high side-lobes. Therefore, some methods of correction equivalent to reweighting the spatial

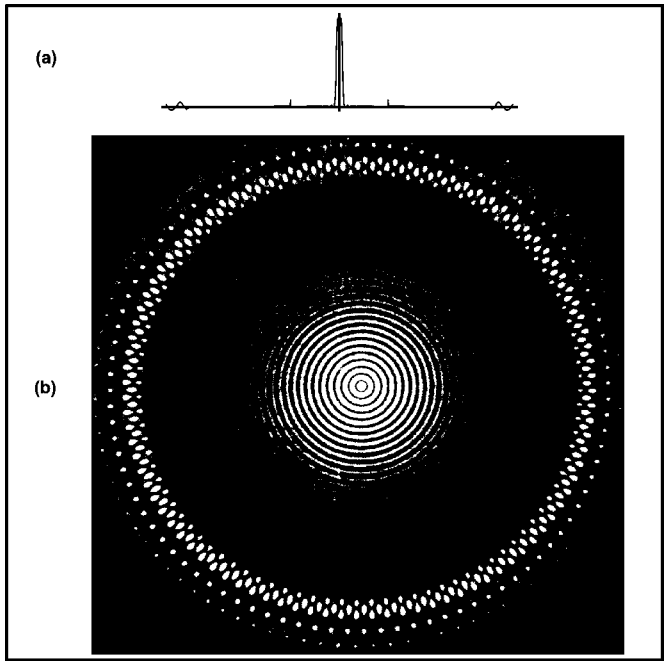


Figure 12.6: The power polar diagram of the Culgoora 96-element circular array with co-phased elements: (a) shows the radial cross-section of the pattern, with the central J_0^2-beam and first-order grating response. The pattern in two dimensions is produced in (b), which is a photograph of the diffraction pattern of an optical model of the array.

Fourier components of the image are required. To do it, a series of circularly symmetrical polar diagrams can be synthesized by introducing progressive phase gradients around the array and that these can be added to the J_0^2 response in suitable proportions to produce a response with acceptably low side-lobes. The correction process is alone during the 8-ms period used for the formation of each column of picture-points by introducing a rapid sequence of phase shifts to produce each of the required polar diagrams in turn. Each pattern is set up for a time, which is proportional to the weighting required. By integrating the receiver output over the whole 8-ms period, the desired response with side-lobes suppressed is obtained. The picture-forming process is controlled by a computer and applies to each antenna channel. The phase changes required to scan the picture field to keep this field centered on the Sun and provide the polar-diagram correction. The 48 picture-point intensities for each column of the image are digitized and transferred to a second computer. This prepares them for recording on magnetic tape. Due to high data rate, it is required to put the data into a format that uses the recording medium efficiently. The computer is used to perform a data compression process. The data-compression computer examines the data for each compete picture

and determines which picture points are brighter than a certain threshold. The intensities and positions of bright points are recorded on the magnetic tape. When the recorded picture is reconstituted from the record the intensity of each of these discarded points is set at zero. The computer finds the threshold level for each picture as a prescribed fraction of the intensity of the brightest point in the picture. Fainter points are below the level of side-lobes from the strongest source and so unreliable.

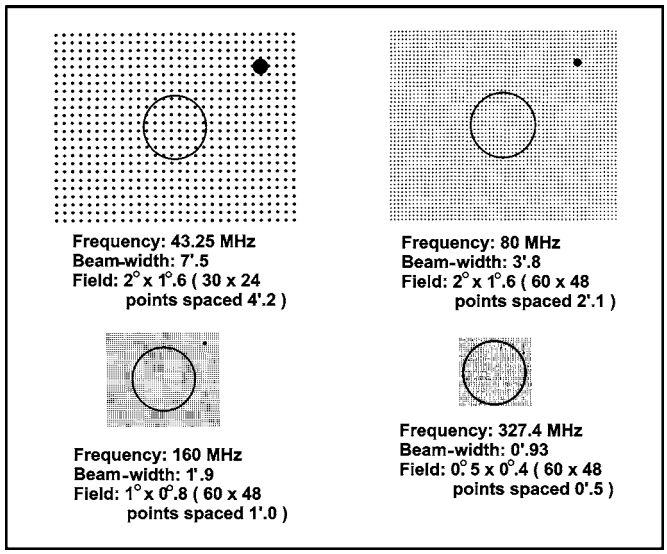


Figure 12.7: The field of view and picture format for the Culgoora radio-heliograph. For each frequency the recorded field of view is shown with the outline of the optical Sun. The black filled circle shows the half-power contour of the array beam on the same scale.

12.4.3 Cross-correlation Arrays

The grating-array radio-heliographs as explained are examples of total-power receiving systems in which the response is derived from the correlations between the signals from all the elements of the array. We now discuss radio telescopes where the desired response pattern is achieved by cross-correlating the signals from two separate arrays. The Fourier transform of the image is made up of the correlations between the signals from each element of one array and those from each element of the other array. The transfer function is the cross-correlation of the gradings of the two arrays and the power polar diagram, which is obtained in the magnitude of the product of the field patterns of the two arrays. The process of correlation and image formation is performed with the type of digital-correlator receiver system. An alternative is the *phase-switched receiver* method, where the

outputs of the two arrays are connected to a common receiver via a switch, which reverses their relative phase at some convenient audio frequency. The switching frequency component of the output is isolated by a synchronous demodulator circuit. If V_1 and V_2 are the instantaneous signal voltages from the two arrays it is easy to show that the amplitude of this modulated component is proportional to the time-average of V_1V_2. Thus, the modulated receiver output corresponds to a power polar diagram, which is the product of the voltage polar diagrams of the two arrays. An example of a cross-correlation system is the Mills Cross, which uses two orthogonal linear arrays with a common phase center. This is shown in Fig. 12.8 (*a*) while typical voltage polar diagram is shown in Fig. 12.8 (*b*). The product of the two fan-beam patterns is essentially zero except where they intersect in the central region, so that the cross-correlation power polar diagram is a pencil beam corresponding to the hatched area of Fig. 12.8 (*b*).

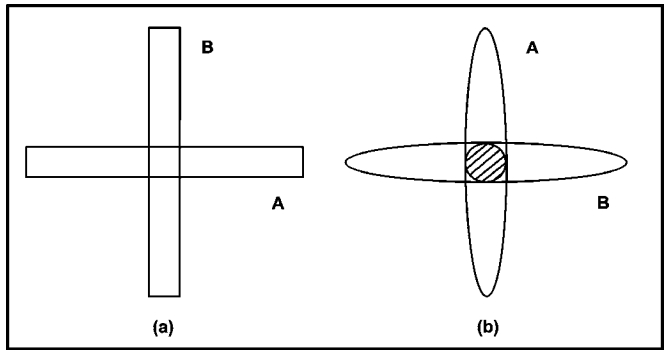


Figure 12.8: (a) A pair of linear antennas arranged as a Mills cross. (b) A cross-section of the fan-beam field response patterns of the two arrays. The pencil-beam power polar diagram of the cross-correlation response is revealed by the hatched area.

12.4.4 The Crossed-grating Array

The filled array of the Mills Cross is replaced by a pair of grating arrays (Fig. 12.9 (*a*)). The power polar diagram (Fig. 12.9 (*c*)) is a family of pencil beams located at the intersections of the two sets of fan beams. The power response profile of each of these pencil beams corresponds to the voltage polar diagram of either array. It is thus important to taper the grading of the latter to secure a pencil beam free of excessive side-lobes. The angular beam-width of the crossed grating is about twice that of the fan-beam responses of the component arrays. Since each of the element spacing is repeated twice in the cross, the same angular resolution can be attained with a *T* array, where one arm of the cross has been omitted (Fig. 12.9 (*b*)). The *T* makes more economical use of antenna elements than the cross. On the contrary, the cross configuration has been preferred in a number of cases, particularly in

centimeter-wavelength applications. This is because the phase adjustment between the arrays is much less critical for the cross than for the *T*. Mills Cross arrays have been used as radio-heliographs at 34.5 MHz and at 25 MHz. The Clark Lake radio-heliograph operates as a digital correlation system with time resolution of about 1 second. Rapid fan-beam operation has also been used with the East/West array only where the broadband response makes it possible to record the variation in position of radio burst sources with both frequency and time.

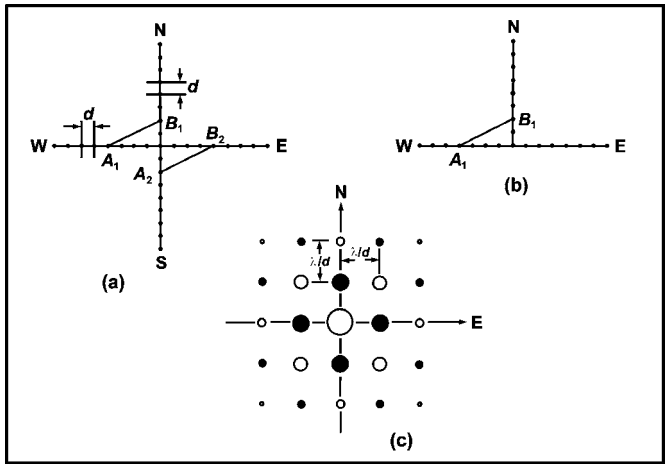


Figure 12.9: The crossed grating interferometer. (a) Geometry of a crass array. Each baseline occurs twice, e.g., A_1B_1 and A_2B_2. (b) The equivalent *T* array. (c) Central part of power polar diagram. Circles show positions and relative responses of pencil beams. Alternate pencil beams give positive and negative responses—filled circles represent negative response.

12.4.5 Low-redundancy Arrays (The Compound Interferometer)

In principle, the angular resolution of the linear grating array can be increased by extending the baseline *D*. The unit spacing *d* is fixed by the requirement for sufficient separation between the fan beams, so that a very long uniform array needs an inconveniently large number of antenna elements. In a uniform array, each of the spacing is repeated by more than one pair of elements. As a result the response from such an array contains a large amount of redundant information. However, it is possible to construct an array with suitably chosen unequal spacing where this redundancy is greatly reduced. But optimum arrays have not been used much in real-time imaging. For radio-heliographs, a more practicable low-redundancy arrangement is the *compound interferometer*. In this, two collinear grating arrays are used, and the image is formed by cross-correlating the signals from them (Fig. 12.10). The array *AB* has close spacing, corresponding to the

required separation between adjacent fan beams. The second array *CD* consists of two widely spaced elements. The cross-correlations between *AB* and *CD* give Fourier components corresponding to all the multiples of the unit spacing *b*. The length of the array is thus increased effectively from $(n-1)a$ to $2na$, and the angular resolution is approximately doubled by the addition of only two extra elements. Several compound arrays have been used as radio-heliographs.

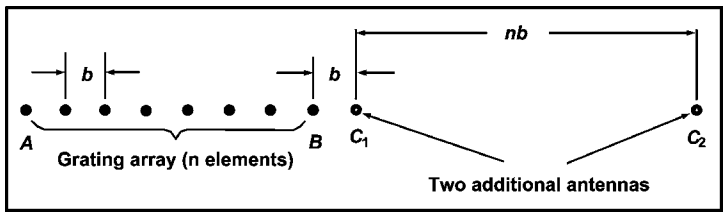


Figure 12.10: A compound interferometer array.

An acousto-optical image processor is shown in Fig. 12.11. The array of transducers is scaled so that a separation of 4 mm represents the unit spacing in the antenna array. The incident light is collimated by a set of slits, which define these spacings. The image is formed on an array of photo-diodes whose outputs give the signal intensities for 256 sampling points in the one-dimensional field of view. The phase-switching method of correlation is retained in this mode of operation. The required modulated component of the output is extracted by recording the differences in the photodiode outputs between alternate half-cycles of the switching waveform.

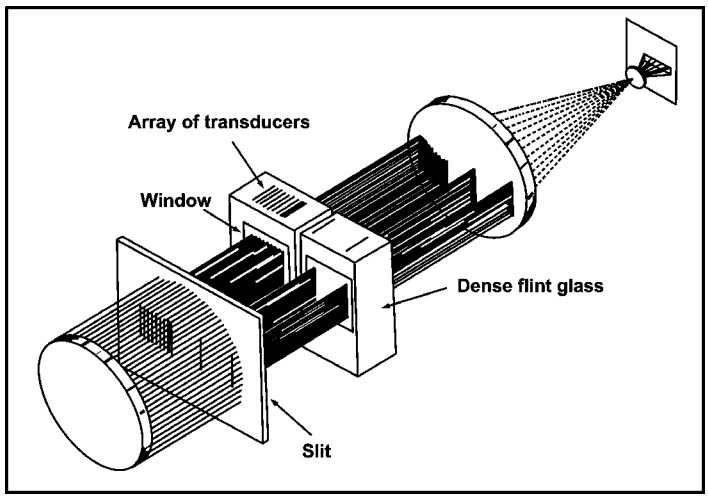


Figure 12.11: An acousto-optical image process for a compound interferometer array.

12.5 CALIBRATION OF RADIO-HELIOGRAPH ARRAYS

In any radio-heliograph an important performance criterion is the dynamic range so that the ratio of intensities of the brightest and faintest sources can be simultaneously recorded. This is usually determined by the side-lobe level of the instrument. Sources fainter than the side-lobes of the strongest source are obscured by the latter. The dynamic range is also limited to the range of signal levels, which can be processed and recorded. In some of the instruments (e.g., the Culgoora radio-heliograph) an automatic gain control is used to ensure that the brightest sources present are always within this range.

Side-lobes are produced by errors in the phase and amplitude relations of the incoming signals from the elements of the array. Such errors are due to variations in transmission lines and electronic circuits between the antennas and the central point at which the signal components are combined. In order to eliminate these error side-lobes, it is required to use some method of calibration to equalize the phase shifts and amplitude changes among the antenna channels. Amplitude calibration is straightforward and is performed by comparing and equalizing the signal levels in all the elements of the array. The measurement of phase shifts is rather difficult due to the need to provide a phase reference signal at each element. In early radio-heliographs phase calibration was performed with the aid of reference signals transmitted to each antenna element by cable from a central oscillator. This method has the disadvantage that its accuracy depends directly on the uniformity and stability of the cables, which carry the reference signals. In other methods the effect of cable length is eliminated from the results. However, errors may still be introduced by variations in impedance-matching in the cables and in the coupling of the reference signals to the array. An alternative approach is to make use of signals from a celestial radio source as the phase reference. This is done by seeing a suitable source in a separate calibration measurement. From the results, the deriving phase corrections are to be applied during array configurations, the phase calibration information can be actual solar observations. With suitable achieved at the same time as the solar image is recorded. The phase-closure method depends on the presence of a sufficient degree of redundancy (i.e., repetition of element spacing) in the array. When this condition is satisfied, the phase errors in the array are determined from simultaneous measurements of the phases of the interference patterns of the signals from each pairs of array elements. The relative phases of all the elements are corrected to eliminate side-lobes due to phase errors. This correction process does not require any knowledge of the sky brightness distribution. As it measures only relative phases, the precise position of the array beam cannot be determined without some additional information. The latter is provided by an *a priori* knowledge of the brightness distribution or by making an absolute phase calibration between one or more pairs of elements.

12.6 ATMOSPHERIC PROPAGATION EFFECTS

In general, tropospheric refraction is not important in meter wavelength radio-heliographs. It is seen that refraction by a horizontally stratified troposphere creates no collimation error in a phased array the plane of which is horizontal. Also, departures in density and humidity from horizontal stratification are too small for significant phase errors in radio-heliograph arrays. On the other hand, ionospheric refraction has serious effects, particularly at longer meter wavelengths. The Culgoora radio-heliograph has shown that at 43.25 MHz ($\lambda \approx 7$ m) collimation errors in excess of 10′ arc are common. Also, the measured positions of a source are often scattered over a range of 10′ to 20′ during a few hours of observation. Such errors are large enough to cause great difficulty in interpreting the solar recordings.

EXERCISES

1. What are radio-heliographs?
2. Related to the theory of antenna arrays, explain the following with suitable mathematical consideration:
   (*a*) Aperture illumination.
   (*b*) Field response pattern.
3. What is suggested by the sampling theorem regarding image formation?
4. Describe briefly the beam-scanning method, pointing out how the loss of sensitivity can be avoided.
5. Explain the principle of the acousto-optical technique related to image formation.
6. What are correlator radio telescopes, and how they work?
7. Give the principles of :
   (*a*) Grating arrays.
   (*b*) Circular arrays.
   (*c*) Cross-correlation arrays.
8. What are the differences between a cross-correlation array and a crossed-grating array?
9. Explain how the angular resolution of the linear grating array can be increased. Give the principle of compound interferometer.
10. Discuss how calibrations of radio-heliograph arrays are made.
11. In interpreting the solar recordings, explain ionospheric refraction effects.

Chapter 13

MEASUREMENT OF POLARIZATION

13.1 INTRODUCTION

To determine the polarization state of a radio signal, it is required to measure four independent quantities. In the case of solar radio emission we know that with the receiver bandwidths that are normally used there is no possibility of detecting any linearly polarized component and we are concerned solely with the analysis of circular polarization.

13.2 POLARIMETERS

The quantities measured are the received signal intensities I_R and I_L using right-hand (RH or R) and left-hand (LH or L) circularly polarized antennas, respectively. In radio astronomy, circular polarization of the incident wave is said to be RH if the field vector rotates clockwise when viewed in the direction of propagation. The degree of circular polarization $\boldsymbol{P}$ is defined as

$$P = \frac{(I_R - I_L)}{(I_R + I_L)}. \tag{13.1}$$

P is generally expressed as a percentage. When this is done it should be called the *percentage polarization*. If, for instance, $I_L = 3\ I_R$, then $P = -0.50$ and the percentage polarization is 50% LH. The nomogram shown in Fig. 13.1 provides a convenient means of calculating P from I_R and I_L. The degree of polarization thus obtained exhibits only the average polarization state of the total received signal. When multiple sources are present on the Sun with a different degree of polarization, this *global* polarization does not reflect the intrinsic polarization of the individual sources. Even if the polarimeter shows a 50% LH polarization we cannot conclude that there is a single partly polarized source on the Sun. It is possible that the source is bipolar with the LH component three times as strong as the RH component. Alternatively, LH and RH emissions may be coming from different active regions on the Sun. These possibilities cannot be distinguished with polarimeter observations alone. Higher angular resolution of a radio-heliograph is helpful in resolving such ambiguities.

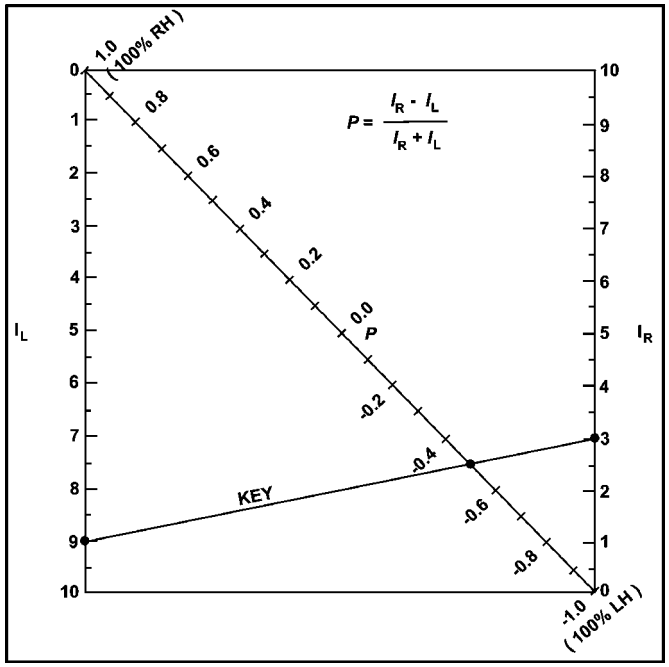


Figure 13.1: A nomogram for quick calculation of the degree of polarization.

13.3 SINGLE-FREQUENCY POLARIMETERS

Antennas that respond to RH and LH circularly polarized emissions are required to measure I_R and I_L. With solar polarimeters at meter wavelengths the usual practice

is to employ a pair of orthogonal linearly polarized antennas in conjunction with a 90° phase shifter. A pair of crossed dipoles feeding a deep paraboloid reflector is a suitable form for polarization observations. This arrangement minimizes the error due to ground reflections. At the longer meter wavelengths the use of paraboloid reflectors becomes impractical and antennas such as a crossed Yagi array is used.

In the meter-wavelength range for single-frequency applications it is usual to use a quarter-wavelength of cable to provide the required 90° phase shift. If the outputs of the crossed antennas are added vectorially with this extra length of cable in the feeder of one of the antennas, the system responds to circularly polarized emission. The sense of polarization received depends on which of the two signals is retarded. It is almost impracticable to use two separate pairs of crossed antennas and receivers, one for each circular polarization. Also, this method makes accurate balancing difficult. Instead, a single pair of crossed antennas is used for both polarizations on a time-sharing basis. The phase shifting cable is switched into one antenna feeder, giving RH polarization response for half of the cycle. It is then changed to the other feeder to give LH polarization for the other half-cycle. A single paraboloid reflector and receiver are used for both polarizations for making gain equalization between RH and LH channels. The switching period should be much shorter than the intended time resolution. For instance, a time resolution of 0.1 second would be appropriate for solar radiometer observations in the meter-wavelength range, but for polarization measurements a switching frequency of at least 200 Hz is essential.

For various practical reasons helical antennas, though inherently circularly polarized, are rarely used in solar polarimeters. In particular, it is required to use two separate helices for the RH and LH components, but it is physically impossible to mount two helices of opposite sense on a single paraboloid reflector. The sense of polarization of a helix is unambiguously determined by its windings direction. As a matter of fact this type of antenna is useful as a check on observations with a crossed-antenna system. In the design of the receiver, such as the bandwidth requirements, the necessary considerations are almost the same as in the radiometer. Square-law response (i.e., output voltage proportional to input power) is essential for satisfactory background subtraction, which is best provided by direct square-law detection. The alternative, linear detection followed by squaring during processing of the data provides a smaller dynamic range. In the polarimeter the response of the detector should be fast enough to pass the polarization switching frequency, while for a radiometer it is sufficient if the response of the detector is faster than the intended time resolution. The components I_R and I_L of the detector output are separated using appropriate signal gating circuits synchronized with the polarization switch. These two components are recorded on a chart by a double-pen recorder or digitally on a magnetic tape.

13.3.1 Background Intensity Problems

The greatest problem associated with the observation of polarization in solar emission is the presence of background components. These are receiver noise and all received emissions other than the solar event. If P is calculated using the raw observed values of I_R and I_L it is not equal to the P of the event. In some cases even the sense of polarization may be in error. To get a meaningful value of P, we must first subtract the background components from the raw observed values of I_R and I_L. In the case where there is no solar activity other than the event, the background is the superposition of the receiver noise, galactic radiation, and quiet-Sun radiation. The subtraction of these background components is simple, as they are steady in intensity. A real-time subtraction is possible, using an offset value present for the day. If a noise storm exists at the time of the event, the situation becomes quite different. We should treat this storm emission as part of the background and subtract its intensity from the raw observed values of I_R and I_L. In such cases the real-time subtraction is impossible, as one cannot estimate the probable value of the background intensity at the time of the *event*.

For short-duration events a satisfactory method is to estimate the instantaneous value of the background intensities by eye-inspection of chart records or computer-plotted graphs of raw observed values of I_R and I_L and then read off the intensities of the event by subtraction. This method becomes simple if square-law detection is used, because one can just measure the deflection above the estimated background level. Although in principle it is possible to program a computer to determine the background level in a similar way, in practice the situations encountered are so diverse that the seemingly subjective visual method yields more reliable results. If both event and background vary with similar time scales, it is almost impossible to separate the components with a simple non-directional polarimeter. Such cases can usually be analyzed with the help of a radio-heliograph with high angular resolution.

13.4 RADIO-HELIOGRAPH AS A POLARIMETER

A radio-heliograph as a polarimeter has significant advantages over a conventional polarimeter. The radio-heliograph measures the brightness at each point in the field of view, while the polarimeter measures the brightness integrated over all the picture points. As a result spatially diffused background components like galactic or quiet-Sun emission may significantly affect the polarimeter observation. If the background is a separate localized source then the polarization of the two sources can be observed separately by the heliograph. In this way polarization measurements with a radio-heliograph are easier to interpret and more reliable

than those with a conventional polarimeter. Pencil-beam systems tend to have inconveniently low time resolution for some polarization applications, which is often too slow to separate the bursts in a Type-I storm from the continuum component.

A higher time resolution could be attained with a correlator-type radio-heliograph, as the data for all picture points are then processed simultaneously instead of sequentially. One-dimensional instruments are more suitable in this respect as their time resolution is very high. The Nancay radio-heliograph provides one-dimensional scans of the Sun at a rate of 50 polarization pairs per second.

13.5 WIDEBAND SWEPT-FREQUENCY POLARIMETER

The Culgoora spectro-polarimeter shown in Fig. 13.2 is a swept-frequency polarimeter for recording the polarization of the solar radio emission over a wide frequency range. The output of this instrument is recorded on moving film in the same way as that of the radio-spectrograph. It gives a record of the degree of polarization as a function of frequency and time.

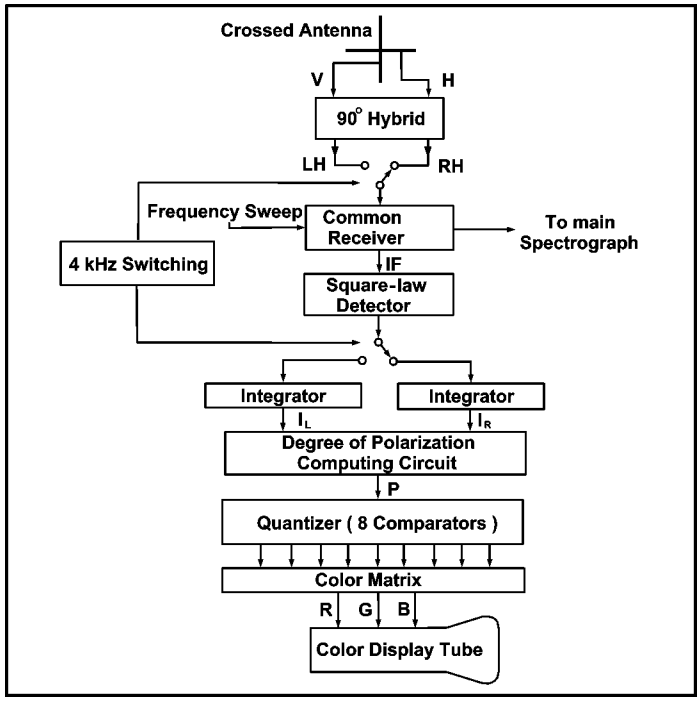


Figure 13.2: A simplified schematic diagram of the Culgoora spectro-polarimeter.

We know that the background cancellation is a major problem even in single-frequency polarimetry. In a swept-frequency system it is nearly impossible to devise a satisfactory method of background cancellation. The best method that can be done in practice is to remove the steady background component (e.g., quiet Sun, etc.). High accuracy in the recorded degree of polarization cannot be expected. The real advantage of the spectro-polarimeter is that we can get a general picture of the state of polarization over a wide range of time and frequency by visual inspection of the record. The antenna for a swept-frequency polarimeter should cover the necessary frequency range and be capable of receiving two orthogonal linear components.

A 90° phase-shifter is required, which works over a wide frequency range to derive the circularly polarized components from two orthogonal linear components. Quadrature (90°) hybrids are suitable for this application. In the meter-wavelength range hybrids, which cover 5:1 or 10:1 frequency ranges are readily available. As in a single frequency polarimeter, the degree of circular polarization is measured by switching between the two outputs of the hybrid, which provide signals for RH and LH circular polarization. At least one polarization switching cycle must be completed within each picture point time. This would determine the minimum switching frequency if the receiver used stepwise frequency scanning. In the Culgoora instrument continuous scanning is used. In it, a fairly high switching frequency of 4 kHz is adopted. This avoids errors as the receiver observes RH and LH components at slightly different times and thus at slightly different places in the spectrum. A polarization switch and a square-law detector capable of handling a high switching frequency are therefore used.

A real-time computation of P within one picture-point time of 1/400 s is required. It would be possible to perform the necessary computation with a high-speed microprocessor. To use this instrument it is necessary to display P as a function of both f and t in a simple manner. Black-and-white representation is not adequate. The P value varies between –1 and +1, so that in black-and-white we would have to set zero at a certain medium grey level. The resolution by eye-inspection would then be poor, around zero. In fact, a difficulty arises in distinguishing between slightly RH, zero, and slightly LH polarization. Color representation is more suitable. A color movie film is used in the Culgoora spectro-polarimeter. Color graphic computer printouts might also be used for the purpose.

To facilitate clear recognition P is quantized in nine different steps. These are 100% to 50% LH, 50% to 30% LH, 30% to 15% LH, 15% to 5% LH, 5% LH to 5% RH, 5% to 15% RH, 15% to 30% RH, 30% to 50% RH, and 50% to 100% RH. These steps are represented by nine easily distinguishable colors: Pink, red, orange, yellow, brownish-grey, green, blue-green, blue, and purple instead of using a continuously changing color-polarization scale. It is possible to record intensity information on

the color film of the spectro-polarimeter because color is used to represent the polarization. However, the latitude of color film is very narrow, and changes in density tend to cause color variations also. Thus, it is wise to record polarization on the color film and to retain the separate black-and-white intensity record.

EXERCISES

1. What are the four independent quantities required to determine the polarization state of radio signals.
2. Define mathematically the degree of circular polarization. What is percentage polarization?
3. Draw a nomogram for a quick calculation of the degree of polarization, and discuss.
4. Discuss single-frequency polarimeters.
5. "The greatness problem associated with the observation in solar emission is the presence of background components." Explain how the problems can be minimized in a radio-heliograph study.
6. What is a swept-frequency polarimeter?
7. Give a simplified schematic diagram of a spectro-polarimeter, and discuss its operation.

Chapter 14

RECEIVERS AND ANTENNAS FOR RADIO ASTRONOMY

14.1 INTRODUCTION

Radio astronomy is the study of the radio waves naturally emitted by astronomical objects like the Sun, Moon, planets, our own galaxy and other galaxies, and clusters of galaxies far distant from our own. The amounts of radio energy received are very small. The science has grown mainly because experimenters have built antennas of very large collecting area and have observed with receivers of very high sensitivity and stability. This chapter describes the principle behind the design of such antennas and receivers and shows how such instruments have been built and used in recent years.

14.2 RECEIVERS FOR RADIO ASTRONOMY

A radio astronomy receiver (or radiometer) is the device for measuring the radio frequency power available from the antenna system. The receiver amplifies the power over a range of frequencies centered on the frequency being used. It thereby

detects the power, and investigates or filters the result before recording it. Fig. 14.1 shows the general elements of such a receiver. We define some of the important parameters of the receiver below:

(*i*) **Central Frequency (*f* MHz).** The output of the antenna (Fig. 14.1) is fed to the preamplifier to provide gain over a range of radio frequencies. The central frequency is the frequency at which observations are made and is usually at the center of the preamplifier frequency response.

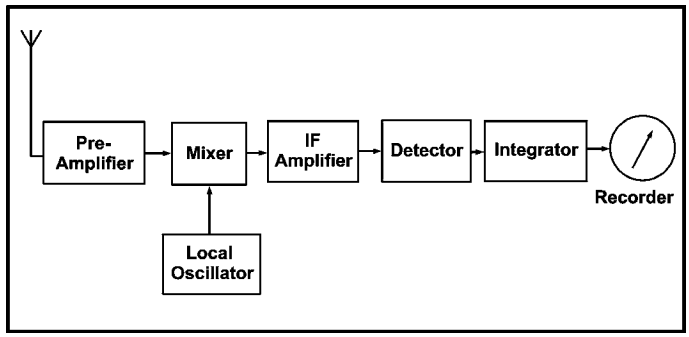


Figure 14.1: Block diagram of a radiometer.

(*ii*) **Gain (*G*).** The receiver is intended to convert the very small change in power of about 10^{-17} W due to a radio astronomy source in the antenna beam into a power capable of being recorded. The gain may be defined as,

$$G_{\text{total}} = \frac{\text{Output power to recorder}}{\text{Input power to receiver}}. \tag{14.1}$$

Since about 10^{-3} W will operate a conventional pen recorder, G_{total} might be 10^{14} expressed as a ratio, or 140 dB. The total gain of the receiver is provided before the detector by the preamplifier and the IF or main amplifier and after detection by low-frequency or direct current-amplifiers. The gain will depend on the radio frequency fed to the receiver and is a function of the size of the input signal to the receiver.

(*iii*) **Bandwidth (*B* MHz).** The gain of the whole receiver is large over a relatively narrow range of frequencies called bandwidth. It is symmetrically placed with respect to the central frequency. In a receiver such as Fig. 14.1 the bandwidth would be determined by the intermediate frequency amplifier, although receivers with the bandwidth set by the properties of the signal frequency amplifier are known. The bandwidth B also termed as the *pre-detector bandwidth* is determined at some point in the receiver before the signal reaches the detector.

(*iv*) **Time Constant (τ sec).** The signal leaving the detector is still fluctuating with time. It is required to smooth out the fluctuations before recording the signal. In its simplest form, this smoothing is done by a resistor and capacitor network forming a low-pass filter or by making a time integral of the detector output over a fixed integration time. Whatever means of smoothing is adopted, the result is to provide a number of statistically independent measures of the detector output per second of time. The reciprocal of this number is referred to as the time constant (τ) of the integrator.

(*v*) **Effective Input Noise Temperature (T_e° K).** Practical receivers generate noise within themselves and also the antenna may generate or receive unwanted noise. A good radiometer is one which allows radio astronomical signals detected most easily and measured most accurately in the presence of such noise. To describe this noise we consider the behavior of a simple receiver (Fig. 14.2 (*a*)). Assuming that the receiver generates no noise, and that it is connected not to an antenna but to a matched resistor in an enclosure at T° K, the available noise power, which such a resistor can deliver to the receiver within its bandwidth B is

$$P = kT\,B \text{ watts.} \tag{14.2}$$

In the above equation, k is Boltzmann's constant and the output meter will show a deflection due to this input power. Eq. (14.2) is derived from Nyquist's formula for the thermal noise power generated in a resistor. If the temperature of the resistor were reduced to 0° K no noise power would be delivered to the noiseless receiver (Fig. 14.2 (*b*)) and the output meter deflection would be zero. Under these conditions, assume the receiver will become noisy (Fig. 14. 2 (*c*)), generating its own noise power and deflecting the recorder by δ. Then we assume that the receiver returned to its noise-free state and the temperature of the resistor increased to a value T_e so that the same deflection δ occurs (Fig. 14.2 (*d*)) as with the noisy receiver. The effective input noise temperature of the noisy receiver is defined as T_e.

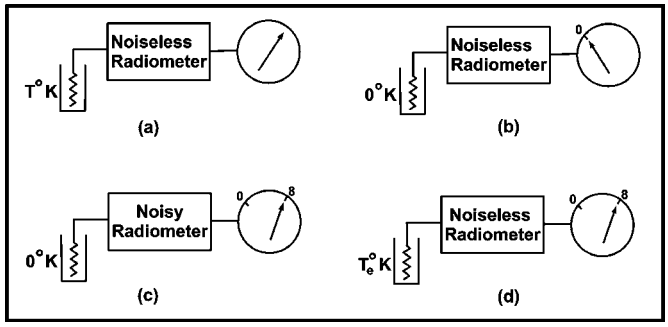


Figure 14.2: The noise temperature T_e of a radiometer.

This definition of the internal noise of a receiver in terms of an equivalent noise input power is related to the well known noise factor (F) of the receiver by the relation

$$T_e = (F - 1) \times 290^\circ \text{ K} \tag{14.3}$$

14.3 RADIO ASTRONOMY SIGNAL

It can be said that all radio astronomy signals are *white* noise, which has the following important characteristics:

(a) Its power spectral density function is constant over all frequencies.

(b) Its autocorrelation function is a constant times the unit impulse function.

If the radio astronomy signal were truly white noise as it impinges on the antenna, its characteristics would be modified by the antenna and receiving system. Before the noise signal goes the detector, it is passed through various parts of the receiver. It is now nearly true to describe the noise power as *bandwidth-emitted white noise*. The degree to which this approximation approaches the truth depends on:

(a) The power spectral density function of the noise before it reaches the antenna, and

(b) The width of the pre-detector bandwidth.

Evidence for (a) is derived by measuring the power received from a radio source within a finite bandwidth whose center is placed at various frequencies throughout the usable radio spectrum. The type of spectra measured is shown in Fig. 14.3, which illustrates the fact that though no radio source is a white-noise emitter, the noise power received from such sources can be described as bandwidth-limited white noise so long as the pre-detector bandwidth is chosen to be so small that the spectral density function of the source is nearly constant over the chosen bandwidth. So for various measurements, different bandwidths are chosen. The condition is met in cases (a) to (c) of Fig. 14.3 if bandwidths of a few MHz are used. For the study of neutral hydrogen (Fig. 14.3 (d)) bandwidths of 10 to 20 kHz are usually used. It follows that the characteristics of the radio astronomy noise powers to be measured are those of bandwidth-limited white noise. Such signals are measured in the presence of noise originating in the atmosphere, the antenna, and the whole receiver system. It is adequate to describe all these unwanted noise powers also as bandwidth-limited white noise.

When a radio astronomy signal is received, a method of measuring the available power from the antenna terminals is that of antenna temperature. The power available from the antenna is related to an antenna temperature T_A as

$$\text{Available power} = k\, T_A\, B. \tag{14.4}$$

The methods of using and calibrating a radiometer result in a knowledge of the change of T_A due to the radio astronomy signal. Thus, the concept of antenna temperature is used as a measure of the power delivered to the receiver.

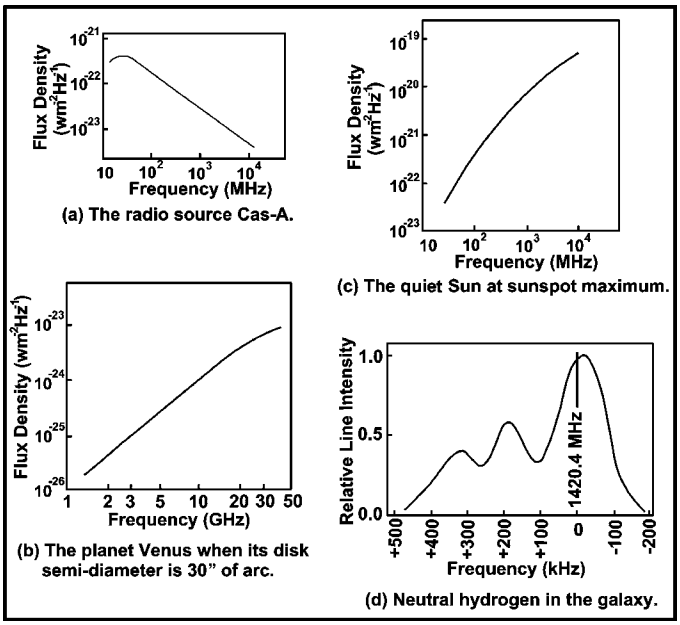


Figure 14.3: Spectra of some typical radio sources.

14.4 DESIRED CHARACTERISTICS OF A RADIOMETER

To detect the smallest signals the radiometer must be sensitive and to measure these signals accurately it must be stable. In fact, some stability is needed even to detect a source, but it of course depends on the techniques that must be used.

(*i*) ***Sensitivity.*** The sensitivity of a radiometer is measured by the smallest change in T_A, which an instrument can detect. Even if the stability were perfect, the total noise (i.e., noise from the radiometer T_e together with antenna noise and atmospheric noise) limits the sensitivity of the instrument. If the total noise in the radiometer system be represented by T_{op}°K, where T_{op} is the operating noise temperature, then the root mean square (rms) fluctuations of the radiometer output signal (ΔT_A), measured in terms of changes of antenna temperature is expressed as,

$$T_A = c\frac{T_{\text{op}}}{\sqrt{\tau B}}. \tag{14.5}$$

In Eq. (14.5), c is a constant of the order of unity and τ is the time constant or integration time used. For high sensitivity T should be reduced, while B and τ should be increased. The reduction of T_{op} by the development of low-noise receiving techniques is one of the main advances in radio astronomical techniques.

(*ii*) ***Stability.*** Changes of the gain, the bandwidth, or the noise temperature of a radiometer are examples of possible unstable behavior. The techniques of improving stability by good design and construction of circuits, the use of switched radiometer systems, and other devices forms the second area of technical advances.

The problem of stabilizing radiometers can be discussed in the light of Eq. (14.5), which shows that ΔT_A, the rms noise fluctuations should have, for a given T_{op}, bandwidth and time constant. It is desirable that observations should be limited only by ΔT_A, but in practice the changes in gain, noise temperature, or bandwidth of the radiometer may be the limiting factor in radio astronomy observations. There are many ways in which the effect of these changes has been minimized, but it should still be recorded that in practice very few radiometers operated on radio telescopes give fluctuations in their outputs as low as those indicated by Eq. (14.5). It is evident that the overall gain of the system must be stable to a high degree. If no means were adopted to remove the effects of the operating noise temperature T_{op}, any gain change ΔG would cause a change at the output corresponding to a ΔT at the input of $T_{\text{op}}\,\Delta G/G$. With earlier receivers, where T_{op} was about 1000° K, this required gain stability of 0.1% to keep ΔT down to 0.1° K, while even with these receivers the rms noise fluctuations (B = 10 MHz, T = 10 sec) could be only 0.1° K. When T_{op} is reduced with the newest receivers there is low-noise fluctuations, and to take advantage of the reduced fluctuation noise still requires an equivalent fractional gain stability. The means by which this stability is achieved are:

(*a*) Total power radiometers.

(*b*) Dicke radiometers.

(*c*) D-C comparison radiometers.

(*d*) Frequency-switched radiometers.

(*e*) Phase-switched radiometers.

(*f*) Correlation radiometers.

14.5 LOW-NOISE RADIOMETERS

The overall noise temperature of a radiometer is determined by the noise generated in the preamplifier and the components between it and the antenna. The expression for the noise temperature of a system of cascaded amplifiers is,

$$T_e = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1 G_2} + \ldots . \tag{14.6}$$

Here, T_e is the overall noise temperature due to amplifier stages with noise temperatures T_{e1}, T_{e2}, etc., and gains G_1, G_2, etc. Eq. (14.6) shows that the noise in the first amplifier appears directly with the signal, but noise in later stages is of less importance since the signal and first-stage noise have been amplified by the earlier stages. Since power gains of about 100 (20 dB) are usually achieved in the first stage of a receiver, the contributions to T_e from later stages become small and can be neglected. Average values for the contribution to the temperature of an antenna due to radiation from the galactic background and from the atmosphere are shown in Fig. 14.4.

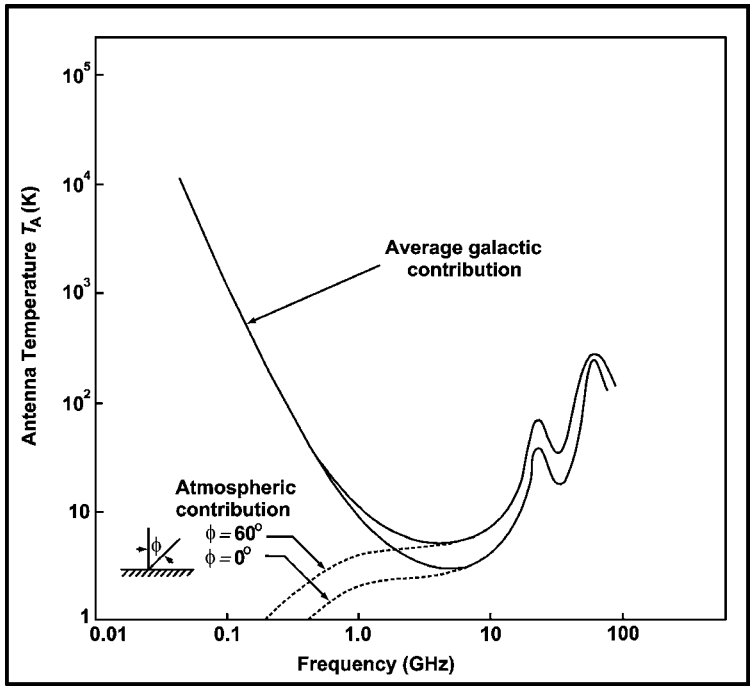


Figure 14.4: Average value for the contribution to the temperature of an antenna due to radiation from the galactic background and from the atmosphere is shown.

14.6 TOTAL-POWER RADIOMETER

In this radiometer, as shown in Fig. 14.5, we must be careful to maintain the gain, T_e, and bandwidth constant by suitable circuit design. The rectified signal at the detector output due to the noise of the receiving system is balanced against a constant voltage and the difference signal amplified at DC before recording. Stable supplies to both heaters and anodes of vacuum tubes, the inclusion of large cathode resistors to a negative supply, temperature control, and the use of temperature-insensitive tuned circuits have all been employed. The results are satisfactory, with gain stabilities of 0.1% maintained over several hours, and reasonable periods existing when the stability is perhaps even five times better. The total-power radiometer observes the source continuously, which is not true for most switched radiometers, so that for a given duration of observing time the noise fluctuation is lower than for the switched system.

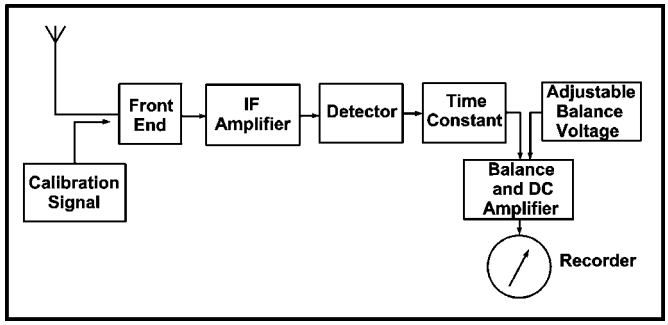


Figure 14.5: The total-power radiometer.

14.7 DICKE RADIOMETERS

This type of radiometer is shown in Fig. 14.6. It is the most widely used radiometer. The input of the receiver is switched between the antenna and a constant comparison source of noise power, usually a resistance at a constant temperature. To a first approximation the receiver noise is the same in both switched positions, so that the output of the detector is a signal corresponding to the constant system noise carrying on it a square wave modulation of peak-to-peak amplitude corresponding to the difference $(T_A - T_c)$ between the antenna temperature and the comparison-noise-source temperature. This signal is applied to the phase-sensitive detector, fed also with a square wave at the switch frequency, and the output of this detector smoothed, amplified, and recorded, is a measure of the difference between the antenna and noise-source temperatures.

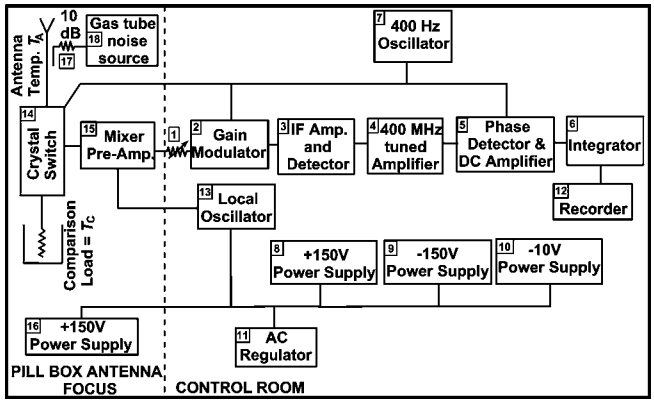


Figure 14.6: A switched-load radiometer.

The radiometer is less affected by gain changes than the total-power radiometer, as the gain changes affect only $(T_A - T_c)$ and not the total system noise. The radiometer is of such value that it needs detailed analysis to show how its performance depends on various instabilities. This can be done by varying the noise-source power to keep the receiver output zero and measuring the antenna temperature by measuring the noise source power. This is a good technique and has not been sufficiently used at the higher frequencies where antenna temperatures are far below room temperature.

Another method by which the noise powers at the two inputs are approximately equalized is by adding noise by a directional coupler into the antenna arm. This is useful only if the resulting increase of system noise is not important. The same effect can be simulated by modulating the gain of the receiver at the switch frequency so that through most of the receiver the noise powers from the antenna in the absence of a radio astronomy signal and from the comparison source are equal. This *gain-modulator* technique has been used in many receivers.

14.7.1 The Instability of a Dicke System

The receiver with the block diagram of Fig. 14.6 is a simple Dicke system, except for the assumption that the gain can be modulated at the switch frequency to give gains G_A and G_c for the *antenna* and *comparison* switch positions. A loss L at temperature $T°$ K is included in the antenna line so that the comparison switch loss can be allowed for. The analysis provides the following conclusions:

(*a*) ***Gain variations.*** A change ΔG in the overall gain G of the system will not be responsible for a change in output provided

$$\frac{G_c}{G_A} = M = \frac{T_A + T_e}{T_o + T_e}. \tag{14.7}$$

In Eq. (14.7) M is the gain modulator ratio and T_e the noise temperature of the receiver after the switch. Thus, without a gain modulator T_A and T_c should therefore be the same but with a gain modulator they can still be different and yet the radiometer can have good stability against small gain changes.

(*b*) ***Changes in T_e.*** A change in receiving noise temperature will show a change in receiver output in all practical conditions unless $T_A = T_c$ and thus $M = 1$. However, this condition cannot be achieved easily in a normal Dicke receiver.

14.8 FREQUENCY-SWITCHED RADIOMETERS FOR LINE OBSERVATIONS

The need for correct measurements of the 1420 MHz radiation from neutral hydrogen in our own galaxy and in external galaxies has led to the development of special radiometers for the purpose. One of the best may be taken as the technique of frequency switching. Fig. 14.7 shows a block diagram of the radiometer in its simplest form. Later developments have added more channels and reduced the bandwidth of the channels.

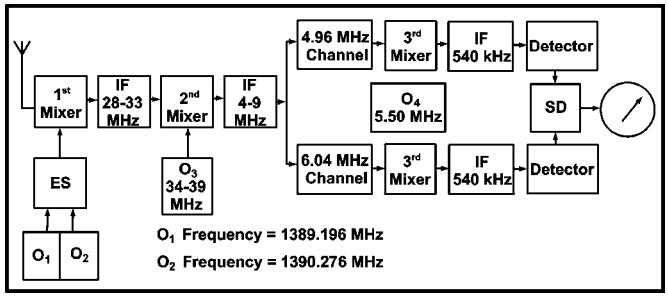


Figure 14.7: The frequency-switched hydrogen-line radiometer.

The first stage of the receiver is a crystal mixer. The frequency switching is accomplished by supplying a local oscillator signal, alternately, 400 times per second through an electronic switch (ES) from the two crystal-controlled sources O_1 and O_2. The frequencies of O_1 and O_2 supplied to the mixer differ by 1.08 MHz. The output of the mixer, after IF amplification at 28 to 33 MHz passes to the second mixer, the local oscillator of which O_3 can be swept in frequency from 34 to 39 MHz. The output after amplification over 4 to 9 MHz is fed to two separate channels; one centered on 4.96 MHz and the other on 6.04 MHz. The difference frequency between these is chosen to be 1.08 MHz—the same as that between O_1 and O_2. These two channels give outputs through a third mixer and 540 kHz amplifiers over a bandwidth of 35 kHz to the detectors and synchronous detector (SD), which

is also supplied with the 400 Hz switching wave. If a 1420 MHz signal is traced through the receiver it will be seen to emerge through the 4.96 MHz channel when O_1 is the LO, and through the 6.04 MHz channel when O_2 is the LO; O_3 is assumed to be set at 35.764 MHz for this to occur.

It is seen that if 1420 MHz is the required frequency it is always available at one of the two channels while the other channel gives a 35 kHz wide band 1.08 MHz above or below 1420 MHz. In this way the synchronous detector output is a measure of the difference between the noise power at 1420 MHz and the mean of the noise power at frequencies equally spaced above and below 1420 MHz. Fig. 14.8 shows how the bands lie with respect to a hydrogen profile to be observed. It is clear that by tuning O_3 across the band, the sensitive band of the receiver may be swept across the hydrogen-line profile.

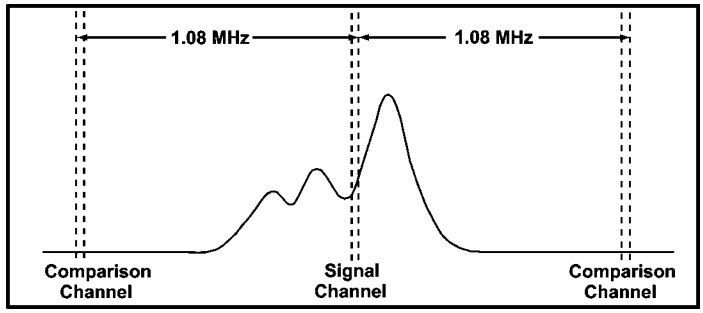


Figure 14.8: The signal and comparison channels in a frequency-switched radiometer.

There have been several improvements in this basic receiver technique. However, the switching technique will improve the overall receiver stability as the effects of receiver noise are eliminated. The switching system also has the advantage that the line is observed in both switch positions so that no sensitivity is lost.

14.8.1 Multi-channel Radiometers

The principle of frequency switching including the sampling of the whole frequency range has been used by several workers. A block diagram of such a receiver built at the NRAO is given in Fig. 14.9. Frequency switching is used to improve stability. The noise power in the signal channel is fed into the bank of 20 filters, each with its own detector, synchronous detector, and time constant. After observation the outputs of all filter channels are sampled and recorded in digital form to give a record of the line profile. The receiver of Fig. 14.9 was designed to study hydrogen in other galaxies, where the hydrogen profile may extend over 2 MHz. Twenty

channels each of 95 kHz wide are used. For galactic hydrogen more channels of narrower bandwidth are necessary.

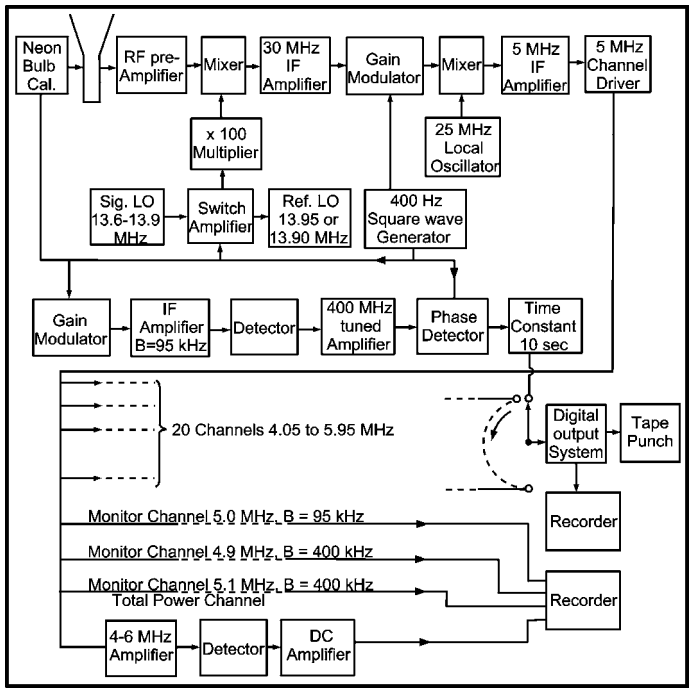


Figure 14.9: A multi-channel hydrogen-line radiometer.

14.9 THE PHASE-SWITCHED RADIOMETER

The use of phase switching is well known and is illustrated in Fig. 14.10 for a two-element interferometer. The phase switch in one of its two switch positions introduces an extra half-wavelength of cable into the line from one antenna to the receiver and thus changing the phase of the signal from that antenna by 180°. The phase-sensitive detector is driven in synchronism with the phase switch. The phase-switching results in a movement on the sky at the switching-frequency of the lobes of the interference pattern as shown in Fig. 14.11. Thus, if a point source is being observed against a uniform sky background only the power from the point source contributes to the recorded fringe pattern. It is also known that for an extended source, the amplitude of the recorded trace is proportional to one term of the Fourier transform of the angular distribution of the intensity across the source. For these two reasons the practice of phase switching has been widely used in

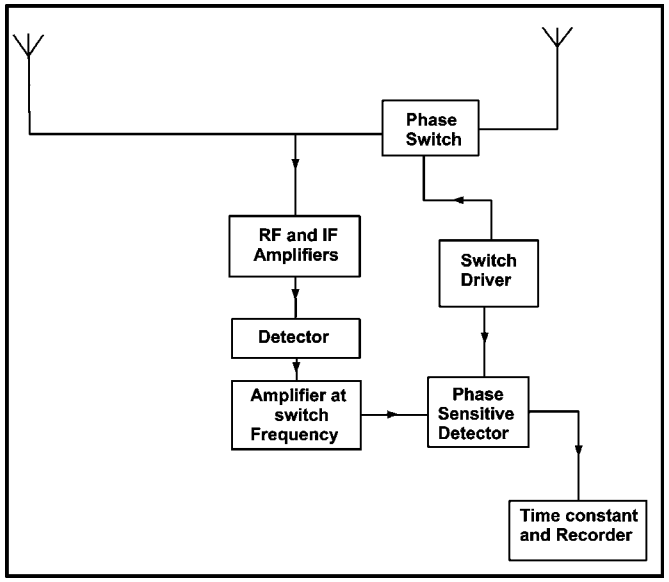


Figure 14.10: The phase-switched interferometer.

radio astronomy. The operation of phase switching removes from the output any contributions from the receiver noise as well as from a uniform sky background, so that the system has the additional stability associated with this characteristic.

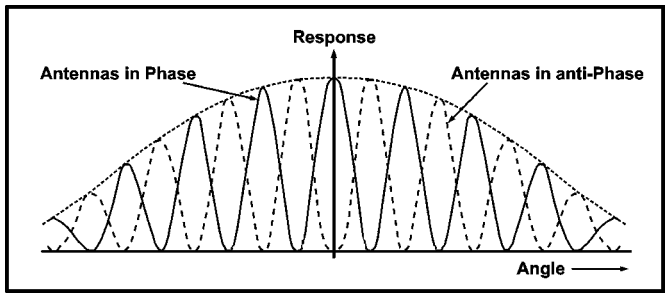


Figure 14.11: The response pattern of two antennas in a phase-switched interferometer.

14.10 CORRELATION RADIOMETERS

The phase-switched interferometer provides a radiometer output proportional to the product of the powers P_1 and P_2 from the two antennas and to the cosine of the phase angle between the phases of the two signals. A similar result is obtained by suitably multiplying the two signals, and in principle this is the technique applied

in correlation receivers. Fig. 14.10 is an example of a two-element interferometer with crystal mixer front-ends used as a correlation radiometer.

Let V_1 and V_2 be the outputs from the two intermediate frequency amplifiers. The correlation coefficient ρ of V_1 and V_2 is defined as

$$\rho = \frac{V_1(t)\, V_2(t)}{\overline{V}^2}. \tag{14.8}$$

Here, $\overline{V}^2$ is the mean value of $[V_1(t)]^2$ and $[V_2(t)]^2$. When voltages $V_1(t)$ and $V_2(t)$ are completely correlated, $\rho = 1$. When they are completely uncorrelated, $\rho = 0$. A multiplier as shown in Fig. 14.12 with appropriate smoothing will give the average value and normalizing will provide a measure of ρ.

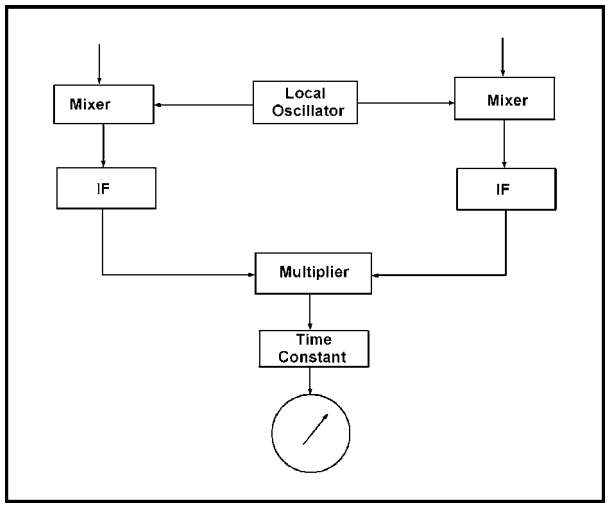


Figure 14.12: The correlation radiometer.

14.11 PARABOLIC REFLECTOR ANTENNAS

The most valuable and versatile antenna in radio astronomy is the parabolic reflector. Such instruments of diameters up to 300 feet have been built on a variety of mountings. In this section we describe these antennas by outlining the choices that lead to the design of such an instrument. Also, we describe some typical antennas, which have already been built and tested.

14.11.1 Choice of Antenna Specifications

Parabolic antennas are general-purpose antennas for radio astronomy and can be used for a wide range of research problems extending over a wide range of

frequencies. In choosing specifications for an antenna, the work to be done with it will generally lead to the choice of the following characteristics, which are to be specified before its design:

1. ***Reflector Diameter (D).*** The choice of reflector diameter is governed mainly by economic considerations. For a given type of antenna mount and for a given upper frequency limit at which the antenna will work the cost of the antenna will vary roughly as D^3 for diameters between perhaps 30 ft and 300 ft.
2. ***Frequency Range.*** The lower frequency at which a parabolic dish will be used is decided largely by the fact that the antenna gain becomes low at long wavelengths. In practice, other types of antenna become more economical. The upper frequency at which the dish will work is set either by the perfection of the parabolic reflector surface or by the size of the holes in the reflector material. The degree of perfection of the paraboloid is determined by the manufacturer of the surface itself, or by the rigidity of the supporting structure. For very large paraboloids the deflection of the structure generally sets the upper frequency limit of the instrument, and the mesh size and the smoothness of the reflector surface are so chosen that they do not themselves reduce this upper frequency limit. The relationship of the gain, beam-width, and side-lobe performance to the imperfections of the paraboloid are of great importance.
3. ***Focal Length (f).*** There is little satisfactory agreement among designers and radio astronomers as to the best method of choosing the focal length for a parabolic dish. The trend is toward an f/D ratio of about 0.3 to 0.4 for the dish with its feed at the primary focus. The Jodrell Bank dish of 250 ft in diameter has its feed in the aperture plane ($f/D = 0.25$). Such deep dishes may give less spill-over but are difficult to feed correctly. They gain also in focal point stability, but may suffer from less satisfactory polarization performance. The Cassegrain antenna is also used as a radio astronomy antenna. The Lincoin Laboratory of Massachusetts Institute of Technology 120 ft Cassegrain antenna (Haystack) was the first large high-precision antenna of this kind used in radio astronomy. These antennas have the advantage of placing the quite considerable weight of electronic instruments required at the focus at or near the vertex of the dish. They suffer from some difficulties in the feed design as the feed must illuminate the sub-reflector quite efficiently and in the blocking effects of the sub-reflector. They have, however, been shown to be very effective as low-noise antennas as the main spill-over effect is caused by radiation from the cold sky entering the feed horn past the edge of the sub-reflector.
4. ***Type of Mounting.*** Almost all parabolic dishes built for radio astronomy have either a polar or an azimuth-elevation (Az/El) mounting. The smaller instruments can be built for about the same cost on either type of mount.

The polar-mounted 85-ft dish with its simple drive and control system is an excellent telescope for radio astronomy. The larger dishes are cheaper if they are on Az/El mounts and cheapest on a simple meridian transit mount. Coordinate converters, which may be mechanical or optical or which may use either analogue or digital computer techniques, have all been used in existing radio telescopes to convert from Az/El to polar coordinates and to permit the telescopes to be moved and pointed in the celestial coordinate system. The drives for such Az/El telescopes must be servo mechanisms of high performance.

5. ***The Drive and Position-indicator System.*** The most important characteristics required of the drive and indicator system are that it can direct the antenna beam to the chosen direction in space and move the beam either to track the apparent motion of a radio source or to scan a selected area of the sky. In radio astronomy the accuracy with which the telescope can be moved has been defined in terms of the telescope and pointing accuracies of one-twentieth of the half-power beam-width are considered to be satisfactory. For an instrument that will be used over a range of frequencies, the smallest beam-width associated with the highest usable frequency sets this pointing and guiding accuracy. Radio astronomers have accepted the fact that the telescopes will deflect under their structural dead load, and so large telescopes have to be calibrated to find the effect of these deflections on the drive and positional accuracy.
6. ***Environment.*** All parabolic dishes used for radio astronomy have been built to withstand the local weather environment. *Haystack* was the first operated within a radome, although many antennas used for radar and communications are so housed. Some savings of cost have been made by requiring precise operation of the radio telescope in winds only up to about 25 mph and thus accepting the fact that observations may have to be suspended in high winds. Survival conditions are chosen to suit the particular telescope location.

14.12 PERFORMANCE OF A PARABOLOID DISH ANTENNA

In this section, we summarize what has been achieved in the performance of parabolic dish antennas.

Antenna theory allows the calculation of the gain of a perfect parabolic antenna if the characteristics of the feed at the focus of the antenna are known. Such calculations overestimate the gain and so are not used in practice to find the gain of radio telescopes. The radio astronomer is concerned with setting adequate gain, satisfactory beam-width, and low spill-over and side-lobes from the antenna.

The greatest gain for a perfect parabolic dish would be achieved with a uniform illumination by the feed horn across the dish aperture and no illumination outside the aperture. Such a feed design cannot be achieved, and most paraboloids have used feed-horn patterns, which diverge from this maximum-gain criterion. Compromises have to be made between the need for gain and the desire to reduce the side-lobes and spill-over in the antenna. The main departure from perfection in parabolic antennas is due to the irregularities of the reflector surface, partial transmission through the surface, and blocking of the total antenna aperture by the feed support legs and the equipment near the focal point of the telescope. All these imperfections lead to a loss of gain of the instrument. The effect due to transmission through the reflector is kept small by making use of closely spaced mesh or solid-surface reflectors. The criterion accepted for transmission is that less than 1% of the energy incident on the reflector surface should be transmitted. This, in turn, has led to choices of solid-surface reflector materials for antenna reflectors designed to work at 10 GHz and of expanded aluminum material for lower frequencies.

The effects of aperture blocking have not been studied, in general, for parabolic radio telescopes. Diffraction theory can be applied to this problem if the obstacles to the wave front are large compared with the wavelength used and if the blocking is assumed to be a geometrical obstruction both of the plane wavefront entering the telescope aperture and of the ray paths from the surface of the dish to the feed. The aperture blocking may then be represented as a reduction to zero of the field over the obstructed part of the aperture and the gain and pattern of the antenna may be calculated for this aperture illumination. A calculation of this type made for the case of an aperture blocked by a tetrapod feed support and a circular electronics canister at the focus showed a loss of gain of 0.5 dB. The antenna pattern was also modified by the blocking in that the first side-lobes were slightly increased in size and the next slightly decreased. It is important to know the effect on the gain of a parabolic antenna of irregularities in the reflector surface. A theoretical analysis of the problem can be made in the case where only small random phase errors are produced in the wave front due to irregularities of the reflecting surface. These irregularities are correlated over finite distances across the surface of the reflector. The theory gives simple expressions for the cases where the correlation interval (c), the distance at which on the average the errors become independent of one another, is much greater or much less than the wavelength used (λ). In large radio astronomy dishes c is always greater than λ and the gain (G) compared to that for a perfect surface (G_0) is for small phase errors:

$$G/G_o = 1 - \overline{\delta}^2. \tag{14.9}$$

Here, $\overline{\delta}^2$ is the mean square phase error in radians squared.

14.12.1 The Antenna Pattern

In radio astronomy it is necessary that the radiation pattern of the parabolic dish be well known, and that it does not show undesirable features. The main requirements with the best dishes are:

1. The main beam should be as narrow as the physical size of the dish permits. However, some slight widening of the main beam is acceptable if this is accompanied by reductions in the level of side-lobes and in the amount of radiation received by the antenna from surrounding objects.
2. The spill-over radiation should be low enough so that the noise contribution added by it is a small fraction of the total system noise of the antenna and receiver.
3. The level of the side-lobes should be low. Those close to the main beam are normally kept to 25 dB below the main beam in power. The distant side-lobes should be about 50 dB below the main beam in power.
4. If the antenna is to be relied on for polarization measurements, the antenna pattern should be closely similar for all planes of polarization of the primary feed horn.

One experimental difficulty encountered is in the measurement of the antenna patterns for large parabolic radio telescopes. To make such a measurement requires the use of a source of radio frequency power, which has an angular size small compared with the telescope beam-width. This radiates sufficient power to be detected even in the relatively insensitive side-lobes of the telescope, which is more than the *Rayleigh distance* from the telescope. This Reyleigh distance requirement states that the antenna pattern is to be illuminated by an essentially plane wavefront. This is normally taken to mean that a spherical wave can be used if the difference between paths from the source to the edge and center of the antenna aperture is not greater than $\lambda/16$. The point source, therefore, must be at the Rayleigh distance R from the antenna, where

$$R \geq \frac{2D^2}{\lambda} \tag{14.10}$$

for this criterion to be satisfied. Phase path errors larger than $\lambda/16$ can be tolerated, but deformed radiation patterns result. R is impractically large for man-made sources to be used to measure patterns of large radio telescopes. The Australian 210 ft and the NRAO 300 ft only move from the zenith to an elevation angle of 30°. The values of R for these instruments near their shortwave limits are both about 50 miles, and at 30° elevation an artificial source would have to be at a height of 25 miles. The use of the stronger radio astronomical sources is possible, provided those of small enough angular diameter are chosen. The Sun is both too large and

too variable in intensity for reliable pattern measurements of large dishes. The most intense of the other radio sources give adequate power for good measurements of main beam shape.

EXERCISES

1. How has the science of radio astronomy developed?
2. Draw the block diagram of a radiometer, and discuss the different blocks.
3. Explain the following parameters of a receiver used for radio astronomical observation:
   (*a*) Central frequency.
   (*b*) Bandwidth.
   (*c*) Time constant.
   (*d*) Effective input noise temperature.
4. What is bandwidth emitted white noise?
5. What is antenna temperature? How is the power available from the antenna related to antenna temperature?
6. Mention the different desired characteristics of a radiometer.
7. Give the names of different radiometers by which the stability is achieved.
8. How is the overall noise temperature of a radiometer determined?
9. Describe a total power radiometer.
10. Describe a switched-load radiometer, and discuss briefly.
11. Obtain an expression for gain variations. What is the gain modulator ratio?
12. What is the utility of frequency-switched radiometers? Discuss the hydrogen line radiometer using a block diagram.
13. Give a block diagram of a multi-channel hydrogen-line radiometer.
14. Illustrate a two element interferometer. What is its importance?
15. Give the principles of correlation radiometers.
16. Describe some typical parabolic reflector antennas. How is the reflector diameter determined?
17. In parabolic reflector antennas, how are the following factors considered?
   (*a*) Frequency range.
   (*b*) Focal length.
   (*c*) Type of mounting.
18. What are the important characteristics required for the drive and indicator system?
19. Describe the performance of a parabolic dish antenna.
20. Mention the main requirements for the best dishes. What are the main experimental difficulties?

Chapter 15

The Sun's Sporadic Radio Emission

15.1 INTRODUCTION

At metric and decimetric wavelengths partly a suitable characteristic is the so-called dynamic spectrum of the radio emission (i.e., the time dependence of the frequency spectrum). The form of the frequency spectrum and its behavior in time provide important information on the generation mechanism of the radio emission.

Wild and McCready (1950) divided the bursts in the metric band into three types and started the classification of the sporadic radio emission by the nature of the dynamic spectra. Two further components, Types IV and V radio emission, have now been added to the Types I, II, and III bursts. Together with the enhanced radio emission owing to the sunspots these five components are the basic group of phenomena observed in the metric band. Enhanced radio emission accompanied by Type I bursts is known as a *noise storm* while Type II bursts are called *outbursts*. The microwave radio emission of the *disturbed* Sun is not divided into its individual components in accordance with the nature of their dynamic spectrum largely. This is due to the lack of spectral observations in this range. On the contrary, it is quite

possible that at centimetric wavelengths and also in the short-wave part of the decimetric band the dynamic spectrum as a whole is not a decisive characteristic for identifying the observed radio emission as it is at metric wavelengths. This is due to the fact that the dynamic spectrum of the microwave radio emission does not have the variety of details that distinguish it at the longer wavelengths.

This chapter discusses the features of the sporadic solar radio emission and the connection between these components and different kinds of geophysical phenomena. We have thus presented a general picture of the phenomenon of sporadic radio emission when a center of activity appears on the Sun.

15.2 CHARACTERISTICS OF RADIO EMISSION FLUX AND SUNSPOTS

The slowly varying component on a recording of the flux of the Sun's radio emission looks like smooth rise in the signal received above the emission level of the quiet Sun with a characteristic duration of the order of tens of days. This component is observed mainly in the range of wavelengths from 3 to 50 cm and occupies a broad spectrum of frequencies of the order of several gigacycles. The relative contribution from the S-component outside the range $\lambda \approx 3$ to 50 cm is very small. Observation of the S-component is complicated because of the masking action of the more intense enhanced radio emission connected with sunspots. The time dependence of the S-component's flux reveals a periodicity with a characteristic time of 27 days, equal to the Sun's period of rotation. This indicates that the source of this component is not only the Sun but individual local regions on its disk. The variations in the enhanced level at centimeter and decimeter wavelengths correlate well with the relative number of sunspots W known as the Wolf number. The corresponding correlation coefficient goes up to 0.8. In relative units the radio emission flux S_w at $\lambda = 10$ cm is connected with w by the approximate relation

$$S_w = 1 + 1.10^{-3}\,\mathrm{W}. \tag{15.1}$$

The intensity of the S-component also correlates with the total area occupied by spots on the Sun's disk. In general, the intensity S of the S-component is directly proportional to the visible area σ of the spots. The coefficient of proportionality depends on wavelength. In the decimetric band the best index of the radio emission when compared with the magnitude of σ is the so-called *complex area of the spots* σ_{comp}. The correlation between S_w and σ_{comp} implies that radio emission continues from a center of activity even after a group of spots has decayed.

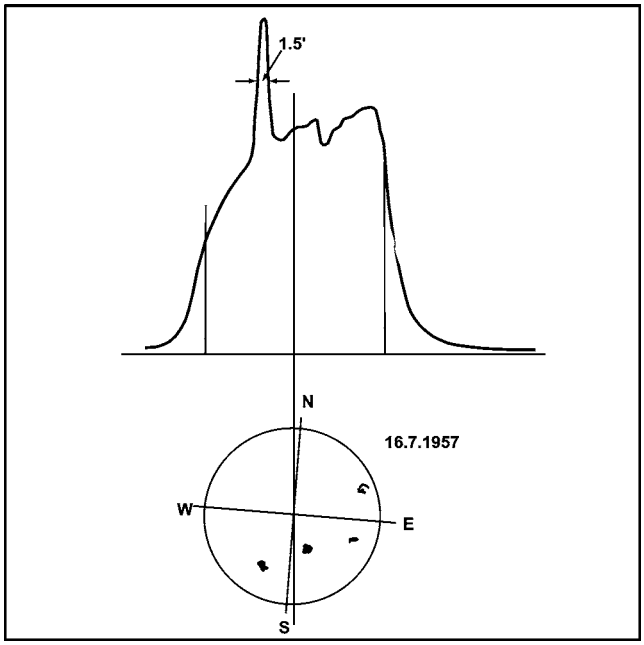


Figure 15.1: One-dimensional distribution of radio brightness over the Sun's disk at $\lambda = 3.2$ cm. The figure also shows the positions of sunspots on the disk during the observations.

15.3 FORM OF LOCAL SOURCES OF RADIO EMISSION

Extensive observations were made to study the connection of photospheric, chromospheric, and coronal formations with local sources of radio emission and to find the position, form, and effective temperature of the sources of the S-component. These observations made it possible to get one-dimensional and two-dimensional distributions of the radio brightness over the disk of the *perturbed* Sun. Typical examples of one-dimensional distributions of the radio brightness over the Sun's disk is shown in Fig. 15.1. At $\lambda = 3.2$ cm the distribution was obtained with an aerial having a *knife-edge* beam about 1 wide. Data on the nature of the local sources of the S-component is obtained if we have two-dimensional distributions of the radio brightness over the Sun's disk.

15.4 FREQUENCY SPECTRUM OF RADIO-EMISSION

The frequency dependence of the intensity I or the effective temperature T_{eff} is important in the theory of the S-component. Reliable data on the frequency

spectrum can be obtained only as a result of simultaneous observations over a wide range of frequencies with narrow-beam aerial arrays. The requirements imposed on the aerials are very rigid as the characteristic size of local sources in the decimetric band is about 5 and at centimetric wavelengths about 1 to 2′, it is clear that the width of the beams should be less than these values or comparable with them. From the value of the aerial temperature T_A the detailed distribution of T_{eff} over the source can be judged. It is also possible to estimate the effective size of the source and the value of the effective temperature averaged over the source $\overline{T}_{eff}$.

Fig. 15.2 reveals total radio emission flux from the bright part of local sources as a function of wavelength: (*a*) observations of April 19, 1958; (*b*) observations of February 15, 1961. The line of dots and dashes indicates the spectrum of a local source located above a group of flocculi (among which there were no spots). To recover the frequency spectrum of the sources if $\overline{T}_{eff}(\lambda)$ or $\overline{I}(\lambda)$ from the known function $S_w(\lambda)$ we need to know the sizes of the sources at different wavelengths. If one considers that at decimetric wavelengths the area of the emitting regions is the same as the area of the flocculi accompanying them, and in the centimetric

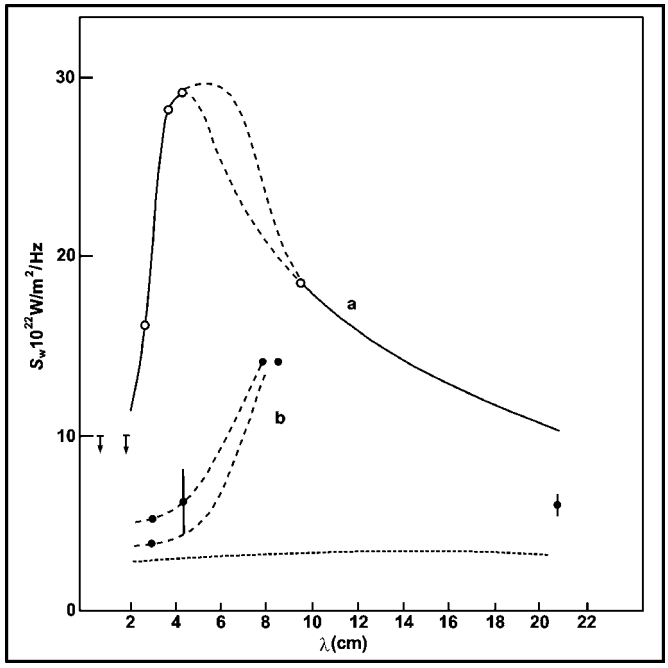


Figure 15.2: Total radio emission flux from the bright part of local sources as a function of wavelength: (a) observations of April 19, 1958; observations of February 15, 1961. (b) The line of dots and dashes indicates the spectrum of a local source located above a group of flocculi.

band is the same as the area of the spots and groups of spots and does not change significantly with wavelengths, then the following conclusions can be drawn. In the decimetric band $\overline{T}_{eff}$ for large sources reaches values of about 10° K, and they are of the same order of magnitude when the wavelength is reduced right down to 5 to 10 cm. In the centimetric band at wavelengths $\lambda < 5$ to 10 cm the effective temperature decreases together with the wavelength more rapidly than λ^2. This can be confirmed by considering $T_{eff} \propto I\,\lambda^2$. For shorter wavelengths the intensity decreases, and T_{eff} drops more rapidly than λ^2. It is possible that the spectrum of $T_{eff}(\lambda)$ is not the same for the bright part of a source and the halo. For the latter it is perhaps close to the spectrum of local sources above flocculi, which are free of spots.

15.5 LOCAL SOURCES ABOVE PHOTOSPHERE: OPTICAL FEATURES OF THE SOLAR CORONA

In maps of radio isophots it can be seen that the sources of the S-component often extend beyond the optical disk of the Sun. This means that they are situated at a considerable altitude above the photosphere. At $\lambda \approx 3$ cm the angular size of a radio emission source is close to the size of the sunspots. This causes the radio emission to appear before the group of spots comes from behind the Sun's disk and to disappear after the group is hidden behind the disk because of the Sun's rotation. At $\lambda \approx 3$ cm the interval between the appearance of a local source and a group of spots is about a day. This corresponds to an altitude of the order of 2×10^9 cm above the level of the photosphere. This interval increases with the wavelength, reaching two or three days at $\lambda = 10$ cm, which can be explained by the increase in the altitude of the local sources at longer wavelengths. It is possible that the effect observed is caused by the extent of the sources increasing in heliographic longitude as the wavelength rises. In the decimetric band the center of a source of emission located near the central meridian coincides with the corresponding plage in the decimetric band and with the nuclei of spots at centimetric wavelengths. The difference in the altitudes at which the local sources and the optical centers of activity are located leads to a difference in the rates of motion over the Sun's disk. The local sources lag behind the optical features connected with them right up to the moment they intersect the central meridian while after passing through the meridian the local sources are ahead of the spots and plages.

Despite the fact that the local sources of the S-component are connected with spots and flocculi, measurements of source altitudes indicate that the generation regions, being located in the higher layers belonging to the inner corona, are not identical with these photospheric and chromospheric formations. From optical observations in continuous light and in the coronal lines it is seen that in the corona

above the spots and flocculi there are regions of enhanced electron concentration, which have a slightly greater kinetic temperature than the ambient medium. The connection between the local sources of radio emission and the features of the solar corona in the λ = 5303 Å line and in continuous light has been investigated by various authors.

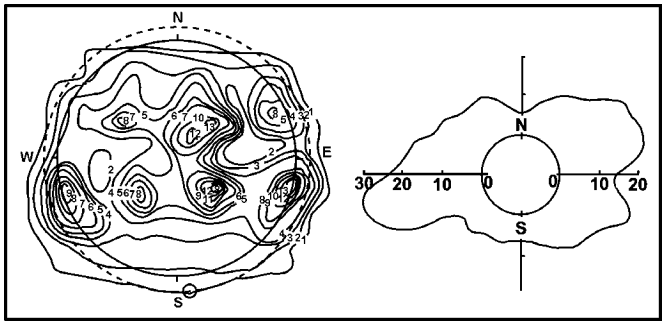


Figure 15.3: Comparison of the Sun's radio isophots at λ = 21 cm with the polar diagram characterizing the intensity of the K-corona at a distance 1.06 R_0 from the center of the Sun's disk. The aperture of the coronagraph is shown by the circle in the left-hand figure.

Some results that allow comparison of the features of the coronal emission in the optical region and the radio frequency band are shown in Figs. 15.3 and 15.4. Fig. 15.3 shows that the radio brightness distribution at 21 cm above the Sun's limb follows the variations in intensity of the K-corona's glow along the line, which is shown by dashes. Fig. 15.4 illustrates the presence of a general correspondence between the radio brightness distribution at λ = 21 cm and the coronal isophots in the 5303 Å line.

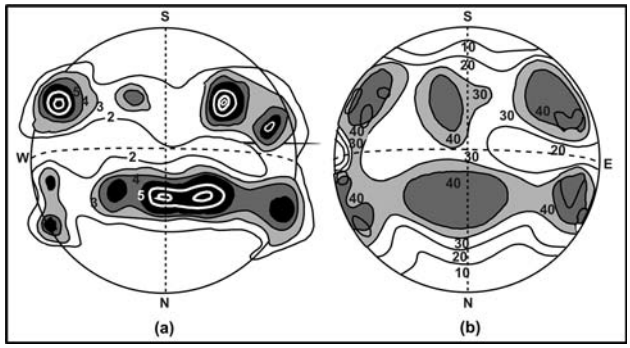


Figure 15.4: Radio isophots of the Sun at 21 cm (a) and isophots of the corona on the 5303 Å line (b). The radio brightness contours are drawn at intervals of 10^{5}° K. The black areas in the right-hand figure are groups spots; wavy lines are boundaries of plages.

15.6 BASIC TYPES OF MICROWAVE BURSTS

In recent years the importance of microwave bursts has been recognized due to their connection with sudden ionospheric disturbances, geomagnetic storms and cosmic ray variations, bursts of gamma radiation, aurora, etc. The limited information about this component of the solar radio emission is felt in the absence of spectrographic observations, which is the usual method of studying the sporadic radio emission at metric wavelengths. It should be pointed out that the term *microwave (centimetric) bursts* only stresses the most characteristic range in which these events occur. They are also observed at millimetric and decimetric wavelengths and closely connected with sporadic phenomena in the metric band. Profiles of typical microwave bursts are shown in Fig. 15.5. Although the bursts frequently appear in groups their component elements can apparently be discussed individually so that the grouping of the majority of centimetric bursts is not their characteristic sign. At rare periods of *irregular activity* numerous relatively weak bursts of comparable intensity (Fig. 15.5 *a*) appear for several hours. The bursts, which have only one maximum, make up the greater majority of all events. Only a tenth have a complex structure with two or three or more maxima. In the clearly defined twin bursts the individual components, which are separated by intervals of the order of a minute, often show very similar characteristics of duration, intensity, and size of the source of emission. The microwave bursts consist chiefly of elementary events of three types: *A*, *B* and *C*.

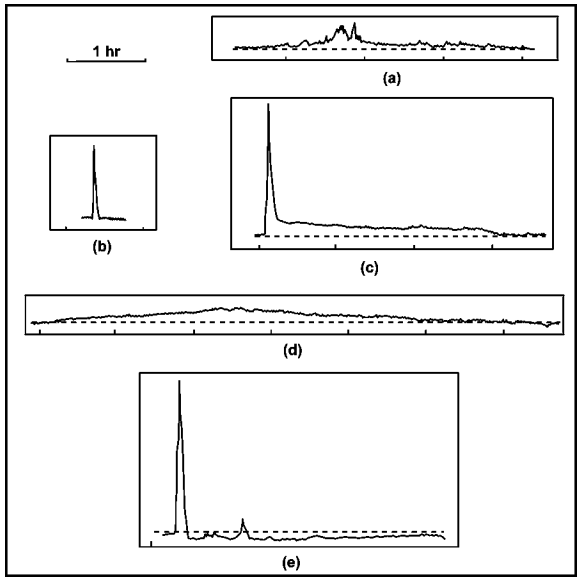


Figure 15.5: Profiles of characteristic bursts of solar radio emission (λ = 10 cm).

Type A bursts (simple bursts as shown in Fig. 15.5 *b*) have the nature of a pulse, being distinguished by a rapid rise in intensity to a maximum value and a slower falling away to the pre-burst level. The ratio of the time elapsing from the beginning of the burst to the maximum to the overall duration of the burst is 0.1 to 0.5 with a mean value of 0.25 while the duration of a burst is 1 to 5 min. Measurements of interference have shown that regions with a diameter of about 1 minute are the source of weak type A bursts with a radio emission flux less than 0.1 from the level of the quiet Sun. In the case of strong bursts the size of the generation increases slightly to 1.6 minute. The corresponding effective radio emission temperature is of the order of $10^{6\circ}$ K for weak bursts and of the order of 10^{7} to $10^{9\circ}$ K for strong bursts. The angular diameter of the source changes in the process of the development of the burst and reaches its minimum value at the time of the burst's maximum phase.

Type B bursts (post-bursts as shown in Fig. 15.5 *c*) follow an individual simple burst or a group busts. Their duration varies from a few minutes to a few hours. The generation region is large in size and the effective temperature of the source of emission is from 10^{5} to $10^{7\circ}$ K. Type C bursts (gradual rise and fall as shown in Fig. 15.5 *d*) are distinguished by a slow rise in intensity to a maximum value and a slow fall to the pre-burst level. The ratio of the duration of the first period to the life of the burst averages 0.4. The duration of these bursts varies from 10 minutes to several hours. The sources of the bursts are localized in regions of small diameter with an effective temperature below $10^{6\circ}$ K. Type A, A–B, and C bursts may be superimposed on each other or exist separately.

15.7 FREQUENCY SPECTRUM OF BURSTS

Fig. 15.6 reveals the first dynamic spectrum of an outburst obtained on a radio spectrograph in the 7.5 to 15 cm band (Kundu and Haddock, 1961). It shows the recording of the burst's intensity while observing at a wavelength of 10.7 cm and the dynamic spectrum of Type III bursts occurred at the same time as metric wavelengths. The microwave outburst consisted of an intense Type A pulse on which another peak of low intensity was superimposed. It was followed by a Type B burst, during which one more Type A simple burst appeared. The spectrum of the Type B radio emission cannot be seen in the figure due to its low intensity. The spectrum of the Type A burst clearly indicates that this emission occupies a broad spectrum of frequencies. Type III bursts at metric wavelengths appeared at the time of the maximum phase of the microwave outburst. However, the nature of the dynamic spectra and the duration of these bursts are different. Type III bursts are thus not a simple continuation of the spectrum of the microwave outbursts into the metric waveband. In certain cases the impression is created that on the radio

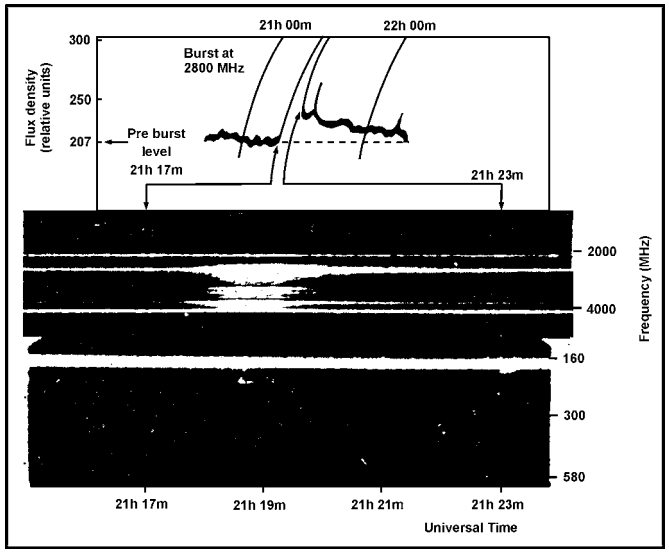


Figure 15.6: Dynamic spectrum of the outburst of radio emission of July 29, 1959 in the 2000 to 4000 MHz band. At the top is the profile of the outburst at a frequency of 2800 MHz.

emission with a spectrum of the type shown in Fig. 15.7 there is superimposed broadband emission with a well-defined maximum of spectral intensity located at

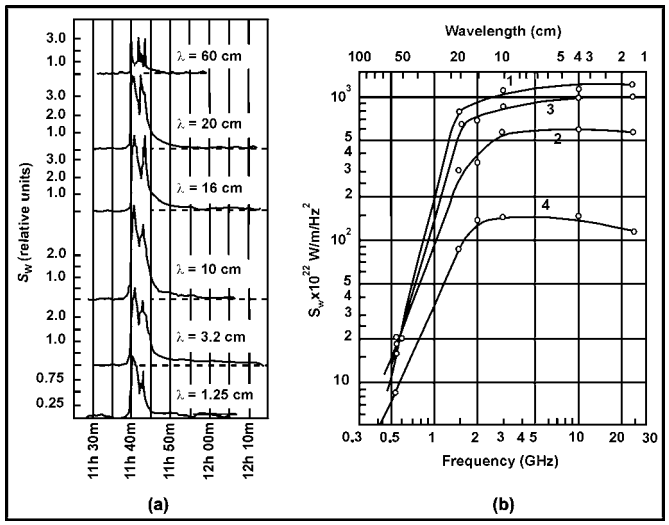


Figure 15.7: Profiles (a) and frequency spectrum (b) of the burst of radio emission of April 18, 1959; at the time of the maximum ($11^h\ 40^m$); at the time of minimum ($11^h\ 41.6^m$); at the time of the maximum ($11^h\ 43.3^m$); during the post-bursts decay ($11^h\ 45^m$).

frequencies where the rise in emission alternates with a region of constant energy (Fig. 15.8). The presence of the maximum of this broadband emission sometimes affects the frequency spectrum in the interval where the intensity increases with the frequency.

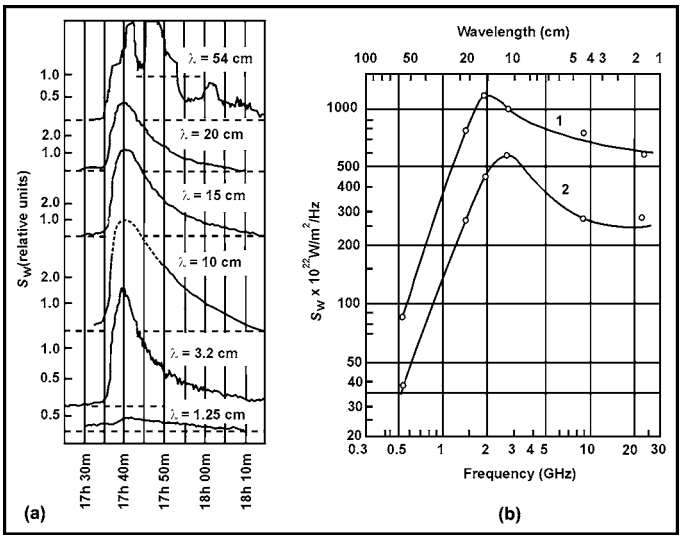


Figure 15.8: The same as in Fig. 14.7 (a) for the burst of June 9, 1959; at the time of the maximum ($17^h\ 40^m$); at the time of attenuation ($17^h\ 50^m$).

15.8 POLARIZATION OF RADIO EMISSION

It was reported that at $\lambda = 3$ cm the emission of two-thirds of all bursts are partially circular or elliptically polarized. The degree of polarization varies from a few percent to 30% and in exceptional cases is as much as 50%. Sometimes quasi-linear polarization was also noted. At $\lambda = 8$ cm with a polarimeter whose band was 10 MHz linear polarization was not found in a single burst among more than a hundred bursts studied. The nature of the polarization of Type B bursts is similar to the nature of the polarization of the preceeding simple bursts. In the majority of cases Type C bursts are partially circularly polarized similar to the S-component. The polarization of outbursts is typical for type A and B events. A study of the radio brightness distribution over the sources of outbursts has shown that the generation area consists of two regions: a comparatively small sector with a low intensity and a long life responsible for the creation of polarized emission, and an extensive region with a high, strongly varying intensity and a short life.

The simultaneous measurements of circular polarization at four wavelengths (3.2, 8, 15, and 30 cm) made by Tanaka and Kakinuma (1959). It was shown that the number of polarized bursts is over 90%. At the wavelengths of 15 and 30 cm polarized bursts appear rarely but even at the shorter wavelengths the degree of circular polarization generally does not exceed 40%. In 70% of polarized bursts the magnitude and sign of the polarization vary with the frequency. This dependence is illustrated by Fig. 15.9, which shows the values of the radio emission flux and the degree of circular polarization ρ_c at different wavelengths during the development of the burst. It is seen that the degree of polarization at the wavelengths of 8 and 15 cm is small and the sign of the polarization is different. Any change of sign outside this range is rarely observed. Sometimes the polarization changes in the range 3 to 30 cm for twice. A change in the sign of the rotation with time is observed far more rarely than a change with frequency at the longer wavelengths (15 and 30 cm).

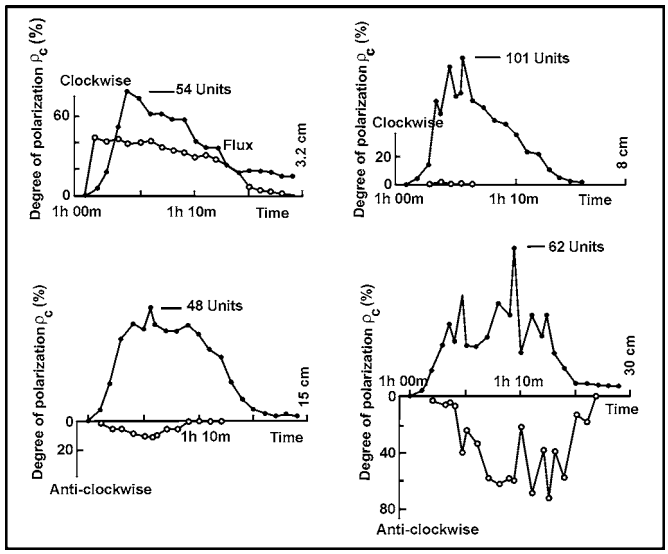


Figure 15.9: Radio emission flux and degree of polarization of burst of July 15, 1957 at various frequencies.

15.9 MICROWAVE BURSTS AND CHROMOSPHERIC FLARES

The sources of the microwave bursts are localized in the regions of the Sun's disk from which the S-component of the radio emission originates. A connection is also known between the intensity of this component and the probability of the appearance of centimetric bursts.

Microwave outbursts are closely associated with chromospheric flares. It can be stated that all outbursts are accompanied by flares and that solar flares whose strength is greater than 2 (i.e., 3 and 3+) are accompanied by bursts in 75% of the cases. This connection becomes weaker for less intense bursts and flares. The beginning of the flare that can be seen in the H_α line almost coincides with the beginning of the burst at $\lambda \approx 10$ cm. Furthermore, the time of the flare's maximum brightness occurs slightly later than the maximum intensity of the burst at $\lambda \approx 10$ cm. For certain flares in which the brightness increases very rapidly, the times of the maximum phases almost coincide with the corresponding times for the bursts. The events in the radio band whose maximum phase lags behind the period of a flare's maximum brightness are about 5% of all Type C bursts. The time that Type A bursts end comes at a period near the maximum of the H_α radiation. Type A bursts occur at the time that the flares are increasing in brightness. Complex bursts frequently continue after the flare maximum as well. The duration of Type B and C bursts is equal to or greater than the duration of the flares connected. The variations in the optical brightness of flares are weakly connected with the variations in intensity of the Type A bursts. A better agreement is sometimes noticed between the curve of flare brightness and the profile of a Type C burst. As an example, Fig. 15.10 shows the functions of the radio emission intensity at $\lambda = 9.4$ cm and the brightness of the flares in the H_α line for two bursts: burst (*a*) consists of Type A phenomena and burst (*b*) can be included among Type C events. It is seen from Fig. 15.10 that there

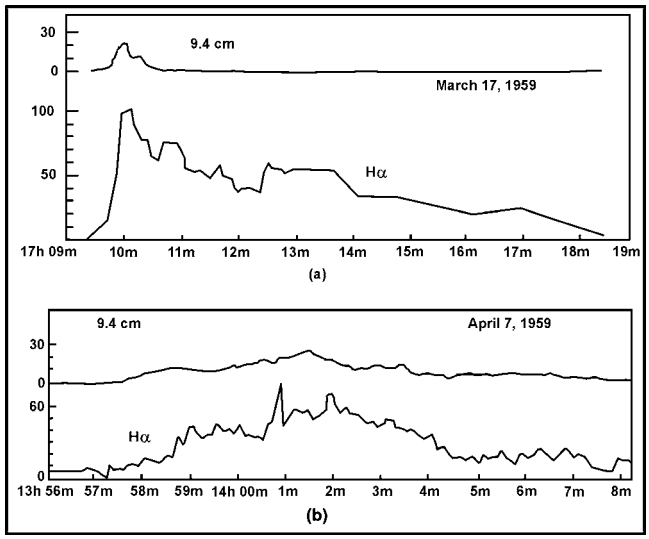


Figure 15.10: Radio emission intensity at $\lambda = 9.4$ cm and brightness in the H_α line for two bursts and the flares connected with them (in percentages of the level of the radio emission and brightness existing before these events). The strong fluctuations in the H_α line in the second figure were probably caused by the effect of the Earth's atmosphere.

is a very clear connection between the radio emission flux in a microwave burst and the width of the H_α emission line in the flare. A close analogy between the profile of microwave bursts and the nature of the variation in the width exists for the great majority of events studied.

15.10 NOISE STORMS

The term *noise storm* combines two closely interconnected events: (*i*) slow rises in radio emission intensity of the order of hours and days appearing at frequencies of the order of hundreds of megacycles, and (*ii*) brief bursts of the order of seconds and fractions of a second following rapidly upon one another. It should be noted that there are bursts of two different types (I and III) among surges of this kind. Noise storms containing Type III bursts are more rare than Type I storms.

15.11 TYPE-I BURSTS

Type I bursts are the most frequent sporadic phenomenon in the radio band. Each burst appears and stays for the whole of its *life* in a narrow frequency range whose width may be much less than the radio emission frequency. The duration of individual bursts can be obtained from recordings of the radio emission intensity with high-speed devices having a small time constant. An example of one of these recordings is shown in Fig. 15.11. Type I surges exhibit a noticeable tendency to collect in groups of two to three bursts. The interval between two successive bursts in a group is about 0.5 seconds on the average.

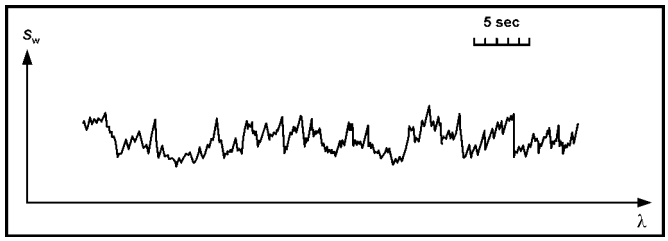


Figure 15.11: High-speed recording of solar radio emission flux during a noise storm.

Noise storms are mainly observed at metric wavelengths; in the decimetric band they are rarer and less intense. The intensity of the enhanced radio emission generally has a maximum at wavelengths of 2 to 4 m and the range of frequencies occupied rarely exceeds 250 MHz. The radio emission spectrum does not remain constant as a noise storm develops but changes noticeably in a few tens of minutes.

When a noise storm starts, the fluxes from the solar radio-emission within the metric band increase by a factor of 10^2 to 10^3 or more.

15.12 TYPE-II BURSTS

Type II bursts are rarest events in the Sun's radio emission. Even at a period of maximum solar activity in every 50 to 100 hours of observations on the average only one burst of this type is noticed. On the contrary, the frequency of appearance of Type I and III bursts is one burst per minute. The number of Type II bursts observed shows considerable variations over a solar cycle and decreasing near its minimum phase. Type II bursts, in general, appear at the time of chromospheric flares near the maximum of their radiation in the H_α line or slightly later, and generally stop earlier than the flare itself. In some cases, the appearance of bursts is not accompanied by any noticeable effects in the optical region. Flares of importance Type I or over have been observed in approximately 60% to 80% of the cases when Type II bursts have appeared. However, the probability of the appearance of a Type II burst when a flare arises on the Sun's disk rises rapidly with the class of flare—from 2% for a Type I flare to 30% for a strong Type III flare. The eruptive activity of the flares (i.e., ejection of material into the corona) is not apparently a sufficient condition for the generation of Type II bursts since the ejection of prominences from the flares is a more frequent phenomenon than bursts with slow frequency drift. During Type II bursts a rise in the radio emission flux at metric wavelengths to 10^{-19} W m^{-2} Hz^{-1} may not be considered out of the way. Cases are known when the radio emission intensity was even higher. The data on the nature of the polarization of the Type II bursts are very incomplete but some say that Type II bursts are either not polarized or are polarized very weakly. The most outstanding feature of Type II bursts (and also Type III) found when investigating the dynamic spectra of bursts is the so-called second harmonic. In the dynamic spectra of the Type II bursts the radio emission occupies two bands several tens of megacycles wide whose contours repeat each other in general outline and in many details. Since the structure of one band at the frequency is repeated in the other band approximately at a frequency 2ω, this served as a basis for calling the first band of the radio emission the fundamental frequency and the second band the second harmonic.

EXERCISES

1. What is meant by dynamic spectrum of the radio emission?
2. What are three different types of bursts in the metric band?

3. What are 'outbursts'?
4. Describe the characteristics of radio emission flux and sunspots.
5. What are the forms of local sources of radio emission?
6. With a suitable diagram, explain the frequency spectrum of radio emission.
7. Describe the local sources above the photosphere.
8. Explain the optical features of the solar corona.
9. What are microwave bursts? Discuss the basic type of microwave bursts.
10. Show the first dynamic spectrum of an outburst obtained on a radio spectrograph, and discuss the recording of the burst's intensity.
11. Briefly discuss the polarization of a radio emission.
12. What are the sources of microwave bursts?
13. "Microwave outbursts are closely associated with chromospheric flares"—explain.
14. The term 'noise-storm' combines two closely interconnected events. What are they?
15. Discuss Type II bursts in the Sun's radio emission.

Chapter 16

Some Unsolved Problems of Solar Meterwave Radio Bursts and Some International Solar Radio Instruments

16.1 INTRODUCTION

The entire existing terrestrial energy resources owe their origin directly or indirectly to the Sun. However, the most vivid demonstrations of the terrestrial response to solar transient phenomena come from magnetic storms, Short Wave Fade out (SWF), ozone variability, weather connection, etc.

16.2 INFLUENCE OF SOLAR RADIATIONS ON THE EARTH'S ATMOSPHERE

The Sun's influence on the Earth's atmosphere manifests itself in various ways such as radiation (UV, IR and x-rays) and gravitational tides. Variations in

solar activity may be transients (e.g., solar proton events, solar magnetic sector boundary passages, solar flares), long period variations (e.g., the 27-day synodic rotation, 11- or 22-year solar cycle), and short period variations. Several workers have been taken these variations into account in solar activity and correlated them with atmospheric phenomena like atmospheric electricity, vorticity, ozone content, and rainfall. Satisfactory physical mechanisms connecting solar activity with weather has not yet been evolved and the subject remains highly controversial. The vital problem is that the energy brought into the atmosphere by processes associated with variable solar activity is much less than the energy of tropospheric circulation. The physical mechanisms involved have been exhaustively examined on 'Response of the atmosphere to solar activity variations' with an emphasis on the role of atmospheric electricity in Sun weather relations. The solar controlled conductivity variations in the lower stratosphere over thunderstorms control the current flow in the global circuit. This may ultimately influence cloud physical processes and thereby the atmospheric energy. The solar modulation of the Earth's electric field takes place through regulation of ionizing radiation. The lower mesosphere and stratosphere appear to be regions where the solar signal interacts with the lower atmosphere.

16.3 GLOBAL ELECTRICAL CIRCUIT

According to the electrical global circuit concept, worldwide thunderstorm activity is in effect a DC degenerator resulting to a flow of current through a global circuit and maintaining the Earth's electric field. At about 60 km an equipotential surface is considered to exist since at that height the conductivity is so large that potential gradients are very small. This equipotential surface, the so-called *ionosphere* is considered to be the outer shell of a spherical capacitor, while the inner conductor being the Earth's surface and in between these two conductors, the ionosphere behaves as a leaky dielectric.

16.4 TYPES OF SOLAR RADIATIONS

Solar radiations can be broadly divided into three categories: direct, diffuse, and global. Diffuse solar radiation also known as sky radiation represents the short wave energy of solar origin, scattered and diffused downwards by (*i*) gas molecules, (*ii*) dust particles, (*iii*) water vapor, and (*iv*) clouds within the atmosphere. Clouds are visible symbols of atmospheric activity and these are the seats of atmospheric electricity. No detailed analysis has yet been made between solar diffuse radiation and cloud discharges. In fact, the phenomena of scattered radiation in the

atmosphere are very complex and varied. Numerous factors determining the quantity of scattered radiation change continuously in time and space thereby making computation difficult. The amount of diffuse radiation received for a particular geophysical location on the Earth's surface depends mainly on the solar altitude, turbidity in the atmosphere, and cloudiness. For a proper understanding of the radiation climatology of a region, apart from information on direct and global solar radiation, a knowledge regarding the distribution of diffuse or sky radiation in the region is necessary. An assessment of the radiation received from the sky assumes great importance in problems concerning the utilization of solar energy, in illumination studies, and in architectural research, especially in the tropics where the diffuse radiation reaches very high levels. Diffuse solar radiation is related to electrical discharge from clouds.

16.4.1 Solar-ionizing Radiation

It is now well established that extra ionizing radiations emitted from the flare region of the Sun induce sudden ionospheric disturbances, which affect a radio communication system in distinctive ways at different frequencies. In fact, during solar flares and microwave bursts there is an enhancement of x-ray flux emitted from the Sun, which causes extra ionization in the lower ionospheric regions, thereby producing Sudden Ionospheric Disturbances (SIDs) of different kinds. Due to these extra ionizations the height gradient of electron density (dn/dh) of the upper region increases, causing an increase in the reflectivity of the region, while at the same time the electron density (N) of the lower region increases, producing an increase in the absorption of the radio signal. Thus, the solar activity considerably influences the terrestrial atmosphere and produces disturbances in the ionospheric medium. These disturbances can be observed indirectly through influence on the propagation of the radio wave. At the level of the D layer, they can be studied by means of long waves ($\lambda \approx 0$ km) and among other methods by observing the atmospheric noise emitted mainly by distant tropical storms. The solar activity results in a sudden enhancement of the level of distant atmospherics and a relation can be found between the slope at onset on the recording and some physical solar phenomena, which are not yet well established.

16.4.2 Solar Radiation and D-Region of the Ionosphere

It is evident that the D-region of the ionosphere, at the altitude between 60 Km and 90 Km, reflects VLF and LF radio waves, whereas it absorbs high frequency radio waves during daytime. The D-region is formed mainly due to ionization of nitric oxide (NO) by Lyman-α radiation and from an ionization of molecular oxygen (O_2) and nitrogen (N_2) by x-rays of wavelengths smaller than 10 Å. It is well established that the increases in intensity of Lyman-α and x-radiations are highly

correlated with the solar microwave radiations at different frequencies. Thus, solar radiations, which control the electron density in the D-region of the ionosphere are related to 10.7 cm solar flux, which is the ionospheric index. The geomagnetic field plays a remarkable role in controlling the electron D-region of the ionosphere. It has been reported that the ionospheric region below 85 to 90 km holds extra ionization during the period of geomagnetic storms. The signal transmission is affected by this change. Magnetic fields also exist on the Sun. The correlation between the signal amplitude and the mean solar magnetic field is also a subject of great interest. Since the Sun is rotating about its axis, its magnetic field strength observed from the Earth's surface is likely to change gradually with certain periodicity. Mention may be made here that the slowly varying component of the solar radio emission, abbreviated as S-component, emanates from the active region surrounding the sunspots on the solar disk. Correlations of radio signal with the solar microwave radiation, geomagnetic activity, and solar magnetic field should be studied in greater detail to examine many questions yet unanswered.

16.5 SOLAR MODULATION OF ATMOSPHERIC ELECTRIFICATION

Electromagnetic radiation, originating in lightening discharges and associated with thunderclouds, propagates via the ionospheric layers. An explanation of how thunderstorm and lightning activity response to the ionosphere is presented here, which may provide important clues to understand how the Sun's variations modulate weather. By solar modulation we mean a modulation owing to the varying intensity of the ionizing radiation originating from the Sun and affecting the lower ionospheric characteristics. If thunderstorm frequency or intensity is changed, additional meteorological consequences are expected. Thunderstorms are one of the more energetic natural phenomena; an average storm releases energy of about 10^8 kWh. Release of latent heat when water vapor condenses is the major source of energy for driving atmospheric circulations.

To explain solar modulation of atmospheric electricity the ionosphere is taken as the outer shell of a spherical capacitor whose inner conductor is the Earth's surface. The atmosphere as a leaky dielectric exists in between them. As mentioned, according to this approach, world-wide thunderstorm activity is in effect a DC degenerator (vertical dipole) responsible for the current flow through the circuit and maintaining the Earth's electric field. An equipotential surface is assumed to exist at an altitude of about 60 km. Due to this altitude, the conductivity is so great that potential gradients become very small. Above the thunderstorm generator (the charging resistor) conduction current flows between cloud top and the ionosphere. This charging current maintains the ionospheric potential

at about +250 kV relative to the Earth. In the *return* path of the circuit (i.e., over the fair-weather portions of the globe), a conduction current flows through the *load resistor* part of the circuit. Fig. 16.1 reveals how estimates of resistances in the global circuit elements are made. An order of magnitude approximation for the thunderstorm generator is 1500 storms at l A per storm for a total current of (1.5×10^3 A). The thundercloud tops are at potentials of 10^8 to 10^9 v relative to Earth and ionosphere, which gives a resistance of 10^5 to 10^6 ohm for the charging resistor over the thunderstorm generator (all thunderstorms in parallel). In the return flow over most of the Earth's surface, the load resistor is the ionospheric potential divided by the total current giving a value of about 165 ohm. It has been estimated (Markson, 1978) that the resistance under the thunderstorm generator is 10^4 to 10^5 ohm. Therefore, the resistance over the thunderstorm generator contains 90% of the total circuit resistance and the one under it contains 10%. The top load resistance is 0.001% and the bottom load resistance is 0.01% of the total circuit resistance. The actual values of the resistances so obtained are also indicated in Fig. 16.1.

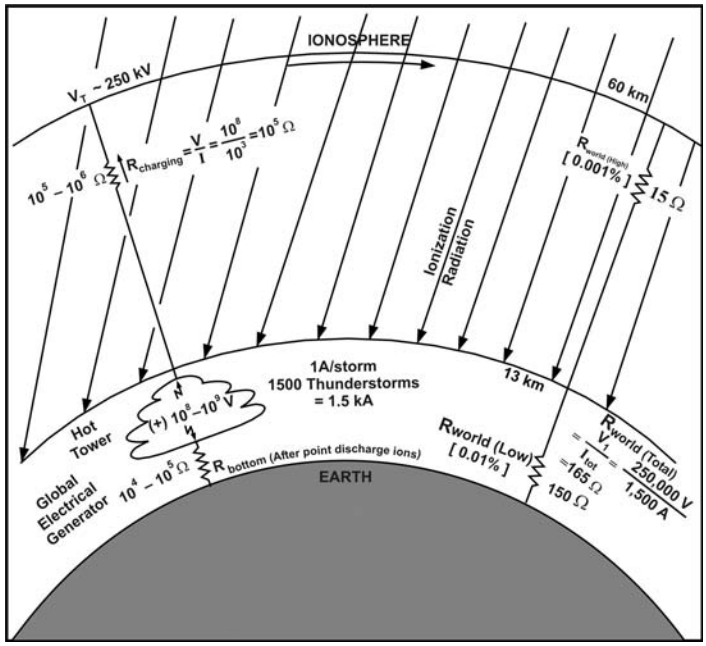


Figure 16.1: The electrical global circuit of the atmosphere showing how solar modulation of thunderstorm electrification might influence the atmospheric circulation. The thunderstorm depicted represents the totality of all global thunderstorms. Estimated resistances in the global circuit are also shown.

16.6 DIFFUSE SOLAR RADIATION AND ASSOCIATED PARAMETERS

For a proper understanding of the radiation climatology of a region apart from the information on direct and global solar radiation, knowledge of the diffuse or sky radiation is required. Diffuse solar radiation represents the shortwave energy of solar origin scattered downward by gas molecules, aerosols (or suspended particulate matter), water vapor, and clouds in the atmosphere. The variability in the amount and the type of the cloud has a dominant influence on the value of diffuse radiation. An assessment of the radiation received from the sky assumes great importance in problems concerning the utilization of solar energy, in illumination studies, and in architectural research, particularly in the tropics where the diffuse radiation reaches very high level. The paper presents results of measurements of diffuse solar radiation in India as recorded at the four metropolises situated at the four corners of the country. The relationships between solar activity and meteorological phenomena has formed an interesting subject of study during the last few decades but the data obtained so far are not enough. Clouds are visible symbols of atmospheric activity and these are the seats of atmospheric electricity. No detailed analysis has yet been made between the solar diffuse radiation, turbidity, and cloud discharges. We were thus prompted to study the variation of diffuse radiation with (*i*) turbidity and (*ii*) cloud discharges appearing in the form of atmospherics over the tropics.

Solar radiation data including the diffuse radiation, global radiation, shortwave and long-wave balances, net radiation and Sunshine hours, recorded by meteorological observatories in India have been collected and properly processed. These solar data have been extensively analyzed separately in conjunction with both turbidity and atmospherics data in the very low frequency (VLF) band.

16.6.1 Instrumentation

The equipment used for the continuous registration of diffuse solar radiation is the same as that for global solar radiation, a Moll-Gorczynski pyranometer and a Cambridge recording milli-voltmeter with the addition of a Schuepp shading ring to shade the thermopile element with its two hemispherical glass domes from the direct Sun. The shading ring arrangement, the installation of the instrumental equipment, the correction to be applied for the sky inevitably obstructed by the ring, the standardization of the instruments, and the evaluation records have been carefully done throughout the work. Regular daily observations with Angstrom pyrgeometers have been taken at the four stations with some limitations. They can only be used at night when there is no precipitation and the winds are light. Precautions have to be taken to avoid the effects of the temperature differences. The instrument has to be calibrated regularly and despite various practical

difficulties regular nightly measurements have to be taken at all the stations. Three Tuned Radio Frequency (TRF) receivers at 10, 20, and 30 kHz were used to record the atmospherics round-the-clock. The receivers used were designed suitably to ensure a large dynamic range with a view of handling a fairly wide range of field intensities.

16.6.2 Factors Controlling Net Radiation

The radiation balance of the whole Earth-atmosphere system during the month of July is shown in Fig. 16.2. There is an energy surplus at this time of absorbed solar radiation over emitted long-wave radiation everywhere north of 10°s. Maxima of energy occur over the relatively cloudless northern subtropical ocean and a deficit over North African and Arabian deserts. The deficit is due to high values of the albedo 30–40 percent and high surface temperatures. Over the entire southern hemisphere the radiation balance is found be negative.

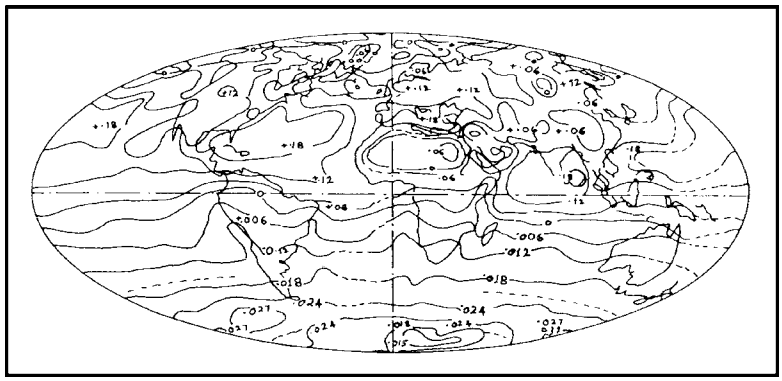


Figure 16.2: Radiation balance of the Earth-atmosphere system for July over the entire path. The figure indicates the net radiation flux ($\times 4.186 \times 10^4$) at the top of the atmosphere in J m^{-2} min^{-1}.

The global average values of albedo, absorbed solar, outgoing long-wave and net radiation are of the order as shown below in Table 16.1:

Table 16.1: Global average values.

Global albedo	29–30 percent
Absorbed radiation	1.423×10^4 J m^{-2} min^{-1}
Outgoing radiation	1.444×10^4 J m^{-2} min^{-1}
Net radiation for July	-0.012×10^4 to 0.029×10^4 J m^{-2} min^{-1}

The global distribution of the albedo, the outgoing long-wave radiation and the radiation balances agree fairly well with the ground-based and air-borne measurements of the same parameters.

Net radiation Q is the difference between the total incoming and outgoing radiation fluxes, both shortwave and long-wave and is expressed as,

$$Q = R_s\uparrow - R_s\downarrow + R_L\uparrow - R_L\downarrow \tag{16.1}$$

Here, R_s and R_L are the shortwave and long-wave radiation fluxes. All fluxes toward the surface are considered to be positive and all fluxes away from the surface as negative. $R_s\downarrow$ is the downward flux of solar radiation at the surface or global solar radiation T. This is affected mainly by cloud amount and also partly by atmospheric aerosols and atmospheric absorption. The nature of the surface itself has no direct influence on T though it has a profound influence on the reflected solar radiation $R_s\uparrow$ and hence on shortwave radiation balance $R_s\downarrow - R_s\uparrow$. The downward terrestrial radiation $R_L\downarrow$ depends on air temperature, on cloud amount and temperature and on the water content in the atmosphere, particularly when the layers are relatively low. The upward terrestrial radiation $R_L\uparrow$ depends on both temperature and emissivity of the Earth's surface. Hence the net radiation, the difference between the amount of shortwave radiation absorbed by the Earth's surface ($R_s\downarrow - R_s\uparrow$) and the amount of long-wave radiation ($R_L\downarrow - R_L\uparrow$) is rather complex of interacting variables, with no clear separation of 'cause' and 'effect'. For a given regime of T, the net radiation and its day will depend on the nature of the surface.

16.6.3 Distribution of Sunshine hours, Normalized D and IFIA

The distribution of the mean daily hours of bright sunshine for each month over the Indian subcontinent is shown in Fig. 16.3. from a consideration of 40 stations. The duration of sunshine shows a maximum over central and northwest India, with an average value between 8–10 hours per day. During the cold season it is highest over central India while the northern regions receives less sunshine due to the passage of winter disturbances from West and decreases in the length of the day with latitude and the south due to cloudiness associated with disturbances in the Bay of Bengal and Arabian sea. As the year advances, till the onset of the monsoon during June, the region of clear skies and maximum sunshine shifts to the North and West Pakistan. In the arid and semi-arid zones of central and northwest India, the skies are clear for over 8–10 hours a day in general.

The mean actual hours of sunshine over Kolkata (Calcutta) is plotted in Fig. 16.4. A normalized curve for the diffuse radiation obtained from the mean values is also plotted month-wise in the figure. For the purpose mean monthly value is first calculated for different months wherefrom normalized values of diffuse

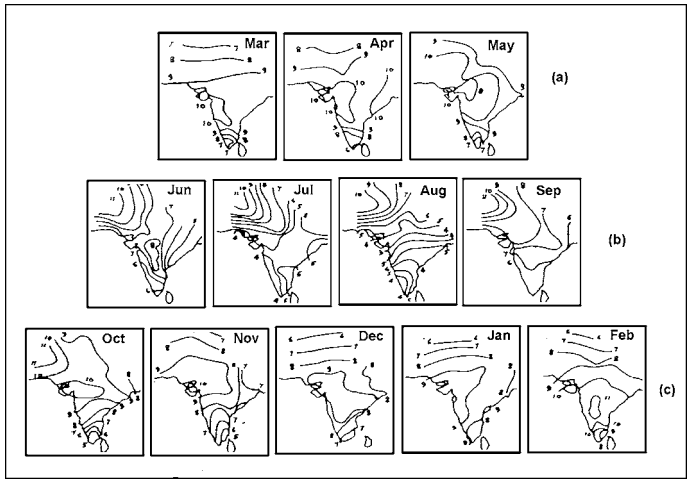


Figure 16.3: Distribution of sunshine hours during (a) pre-monsoon (b) monsoon and (c) post-monsoon seasons over India.

sky radiation is found by dividing the individual values by the maximum value obtained. Similar normalized curve for IFIA is further superimposed by considering the median noise level obtained from the means at 10, 20 and 30 kHz. The figure reveals that the trends for the two normalized curves are similar with a little more fluctuation for the D values. It is also interesting to note that the curve for sunshine hours is almost a mirror image of the other two curves with a minimum during the cloudy monsoon season and a maximum during the remaining clear months.

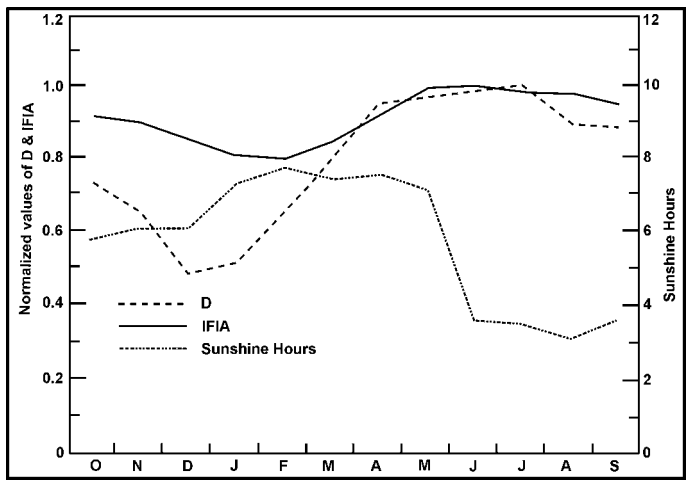


Figure 16.4: Curves showing the sunshine hours, and normalized values of (i) diffuse radiation and (ii) IFIA for Calcutta.

16.7 EFFECTS OF GEOMAGNETIC DISTURBANCES OF SOLAR ORIGIN

Fading of radio waves due to drifting irregularities in the ionosphere has long been a subject of great interest. It has been shown that the period of such fading would increase with scale-size of irregularities, while the period decreases with increasing velocity of the drift involved. Naturally, if drifting irregularities of different scale-size ranges exist at the relaxation level, one would expect distinctive periods of fading. Geomagnetic storms are known to be followed by long-lasting storm after-effects in the ionospheric D-region at high altitudes. Research has shown strong enhancement in low frequency ionospheric absorption during and after very strong magnetic disturbances in the mid-latitudes. The duration of the storm after-effects were found to correlate with the enhancement of high energy (< 40 keV) particles. VLF and LF data of such absorptions at low latitude may be examined to see how they are dependent on season, sunspot activity, and geomagnetic disturbances. The propagation characteristics will help to suggest an origin of the fading and the role of atmospheric gravity waves at such times.

16.8 SPRITE FLASHES

Sprites last only a few thousandths of a second and extend from above storm clouds up to about 60 miles high, reaching the bottom of the ionosphere. Radio noise in the VLF/LF, MF/HF, and VHF/UHF bands associated with the sprites flashes makes a distinct *pop* when played back through a speaker producing a sound that differs from normal lightening discharge signals. The *sprites* appearing as blood red flashes with blush tendrils, appear to have some association with cosmic radiations including the role of solar radiations. As a result sprites have become a subject of interest to astrophysicists and space-physicists. A sprite flash can be one of the strong impulsive radio noise sources that pollute radio-astronomical observations and it may also create problems in space missions by affecting the *space weather*.

16.8.1 Visual Reports

Pilots and others have reported seeing colored columns of light above tropospheric thunderclouds for years (Fig. 16.5). To the eye, they resemble material ejected from a very high explosive source, something like the tracks of atomic particles or rays in a cloud chamber. The flashes may sometimes be described as carrot-, turnip- or jelly-fish shaped, while at other times they have an unusual appearance of dazzling arrays of fireworks, which seem to dance for milliseconds sufficiently above the cloud top. The unusual flashes that occur over intense thunderstorms appear

in two distinctly different forms: (*a*) blood red flashes may be called sprites and (*b*) blue jets appearing in narrow beams may also be called sprites. The flashes last only a few thousandths of a second and extend up to about 90 km in height (i.e., at the bottom of the ionosphere).

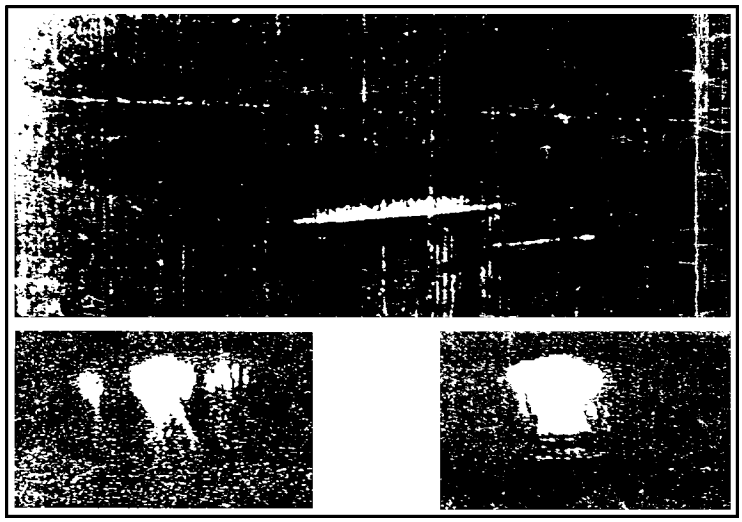


Figure 16.5: Photograph of some typical unusual flashes as reported both in the stratosphere and mesosphere, much above the tropospheric thunder clouds.

Recently, the space shuttle in one of its flights recorded 18 such flashes over Australia, Africa, South and North America. To capture the position and altitude of the flashes, special low-light-level cameras were used in aircraft and were recorded on color video.

16.8.2 Observational Reports

1. *Red sprites* are mesospheric striated glows that typically recur at intervals of several minutes over mature and dissipating organized convections. They have been documented over the stratiform regions of mid-latitude Mesoscale Convective Systems (MCS).
2. Sprites tend to occur in clusters, sometimes appearing to follow the horizontal progression of in-cloud lightening. Compared to the typical MCS lightening flash rates of 40 to 50 per minute, sprites are infrequent, although not rare, occurrences.
3. In MCSs, Cloud-to-Ground (CG) lightening occurs in a *horizontal dipole pattern*, with negative strokes occurring in the areas of active convection and positive strokes in the stratiform regions. These positive CG strokes are infrequent

but frequently exhibit large peak return stroke currents and may lower tens to hundreds of coulombs of charge to ground. In most of the cases they are associated with horizontal extensive *spider lightening* discharges, long dendritic channels that may linger for many tens of kilometers along cloud base.

4. Positive CG strokes are disproportionately associated with electromagnetic *Q-burst* events in the ELF Schumann Resonance (SR) band, the large amplitude ringing of the entire Earth-ionosphere cavity. The SR *continuous* spectra are believed to be driven by the integrated effects of all global lightening. The Q-bursts, separated in time by minutes to hours, are assumed to be excited by the largest lightening events on the planet.
5. The combined data of sprites and ELF transients provide confirmation of their overall coincidence. For example, in Fig. 16.6, the time of the first video detection of a sprite is coincident with the transient's onset to within a few milliseconds. The time of an associated positive ground stroke, having a peak current of 327 KA, is also shown. The coincidence of sprites with ELF transients implying extraordinarily large charge transfer favors an electrostatic over an electromagnetic trigger mechanism.

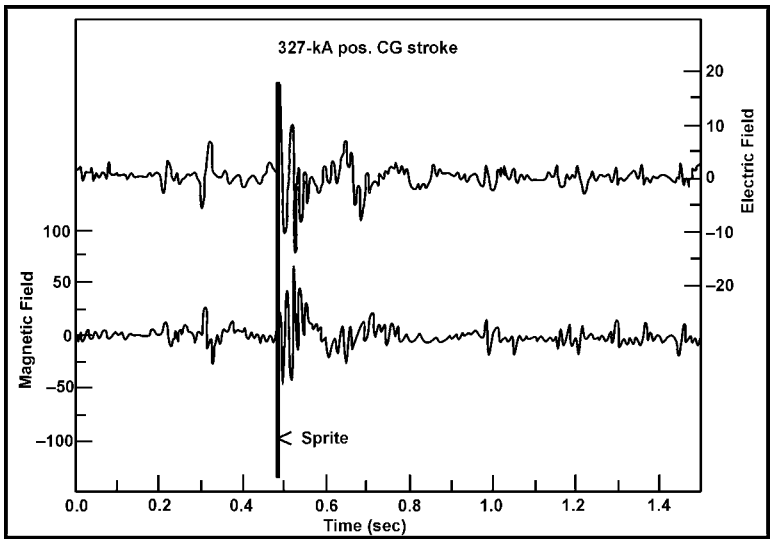


Figure 16.6: Record showing sprite related ELF transient. The electric and magnetic fields are represented by thin and bold markings, respectively. The times of positive CG strokes and sprites are overlaid.

6. Studies have suggested that sprites are caused by the radiated electromagnetic fields associated with charge acceleration in intra-cloud flashes or return strokes (Fig. 16.7).

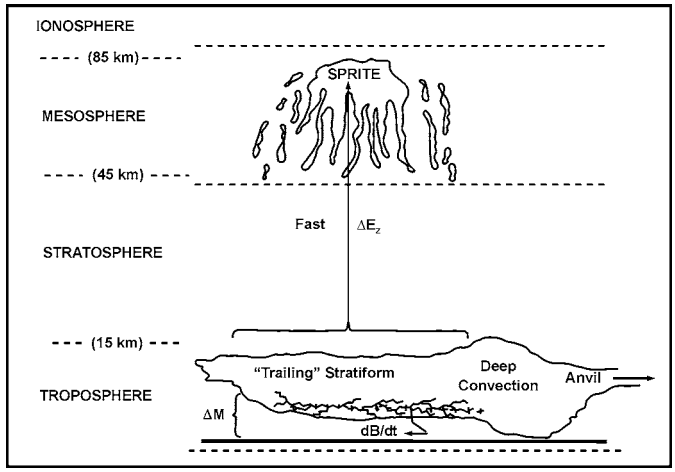


Figure 16.7: Schematic diagram showing the proposed connection among sprites, positive CG strokes, and Q-bursts. The positive CG stroke assumed to be the electrostatic source for the sprites and electromagnetic source for the Q-burst.

16.9 THE DYNAMIC SPECTROGRAPH

The spectrograph usually consists of three parts: (*i*) an array of four Log Periodic Dipole Array (LPDA)[13] antenna in the frequency band of 40 to 1000 MHz, (*ii*) a spectrum analyzer with a 30 dB preamplifier, and (*iii*) a PC-based data acquisition system. Each of the LPDAs essentially consists of two booms and the elements are attached alternately with the two booms. The separation as well as the length of the elements decrease in the logarithmic scale. The receiver of the spectrograph consists of a preamplifier having a noise figure of 3 dB, and a spectrum analyzer. The spectrum analyzer should have a facility of being operated at various IF bandwidths, integration time, and scan rate. The frequency range may be chosen as 40 to 215 MHz, and the spectrum may be available in 400 or 600 data points on a GPIB port, which is to be connected to a PC having an HPIB card for data acquisition. The computer takes approximately 10 seconds to access the data through GPIB port and to write in a data file on the hard disk.

The dynamic radio spectrum resulting from the solar bursts consists of three important parameters: (*i*) time, (*ii*) frequency, and (*iii*) intensity of the spectrum. These can be intensity mapped on a grayscale and plotted as functions of time

[13] The LPDA is a wide-band antenna, which provides more or less uniform gain over a wideband of frequency. These are very useful for capturing meter-wave radio bursts.

and frequency.[14] The data browsing software GRAYLEV is used for this purpose and divides the screen into an array of cells consisting of nine points. The intensity is quantized into $n + 1 = 10$ ($0 \leq n \leq 9$) levels and accordingly any particular cell is weighted in gray. The steady interferences like TV, radio, wireless, etc., appear as lines parallel to the time axis. The solar bursts may be roughly visualized as curved, diagonal, or randomly varying lines across the plot. The software provides flexibility in intensity scaling, which aids in enhancing the visibility of the trace of a solar-burst on the plot.

16.10 STUDY OF SOLAR METER-WAVE RADIO BURSTS

The study of solar meter-wave radio bursts dates back nearly half a century. During this period the impressive series of advances in radio-astronomical techniques as well as in theoretical concepts in plasma physics has led to a better understanding of the different aspects of the solar radio emission in meter-wave regions, designated primarily as Type I, II, III, IV, and V bursts. There are numerous unresolved issues about the aforesaid radio bursts, which draw the attention of various groups of uncompromising researchers in this field. Moreover, due to the advent of different kinds of space observatories, particularly in the x-ray and γ-ray bands, the whole scenario about these bursts is likely to be changed. The solar mission Yohkoh with soft x-ray and hard x-ray imaging instruments launched in 1992, the SOHO (Solar and Heliospheric Observations) space mission with the white-light/UV coronagraph, and the Compton Gamma Ray observatory will be useful for comparing the data available from these missions with those of metric bursts that convey information about the coronal features. Before going into the different kinds of unsolved problems about metric bursts, let us summarize the important aspects of these bursts, which are quite well known to us.

1. ***Type I.*** These are long-lasting (a few hours to several days) radio emissions, consisting of thousands of short-lived spikes superimposed on a slowly varying continuum.

   The emission mechanism is thought to be a fundamental plasma emission and is due to coalescence of Langmuir waves with low-frequency waves (e.g., ion-sound waves); the exciting agency for the Langmuir waves is a population of energetic particles trapped in a closed magnetic structure over an active region.

[14] The time comes along x-axis, frequency along y-axis, and intensity can be called grayness of a particular cell.

2. ***Type II.*** These bursts drift slowly from high to low frequencies, with a drift rate less than 1 MHz/sec. They often give rise to a double structure, signifying fundamental and second harmonic emission.

   The drift is thought to be due to the outward motion of a magneto-hydrodynamic shock wave at the leading edge of material ejected as a result of catastrophic explosion of a flare. The emission is due to the electrons accelerated during the propagation of shock front.

3. ***Type III.*** These bursts are intense and highly drifting from high to low frequencies, with a drift rate of the order of 100 MHz/sec, which decreases with the decrease of frequency. These bursts occur over a frequency range of 1 GHz $\geq f \geq$ 10 kHz and are generally repeated at the second harmonic.

   They are caused by streams of electrons moving out with a velocity 0.33 c, and their passage sets up plasma oscillation—Langmuir waves—which then radiate at their characteristic frequency.

4. ***Type IV.*** These are long-lasting, non-spiky bursts, which may originate in stationary or moving sources. The stationary type Type IV are broadband continua with fine structures, and moving type Type IV are broadband and slowly drifting with smooth continua.

   The evolution of such bursts can be explained with the help of the Gyro-synchrotron emission mechanism.

5. ***Type V.*** On dynamic spectrograms they appear as smooth and short-lived continua following certain Type-III bursts. They are pronounced in the long-meter wave region and have longer durations at the lower frequencies.

   They are generated due to plasma radiation from electrons trapped in high coronal loops adjacent to open field lines traced by Type-III electrons.

6. ***Type U and Type J.*** In Type U, the observed radiation drifts first downward and then upward in frequency as time proceeds. U-bursts come from the leading edge of the CME, which comprises an ejected (or rising magnetic loop) in which those energetic electrons travel with velocity 0.1 to 0.2 c that gives rise to radio emission.

   Sometimes the bursts turn over (i.e., the drift rate goes to zero) but there is no emission on the descending legs, and they are called J-bursts.

Apart from the above mentioned types, there are other forms of radio bursts, such as stria bursts, reverse drift pairs, etc., which have been studied little.

16.10.1 Some Unresolved Issues

There are several theories regarding the generation of the various types of bursts, but none of the theories can fully explain the observed characteristics of bursts.

In order to clarify issues arising out of different theories the following points in regard to the observations are to be emphasized.

Type I

(*a*) Quantitative data on the dynamic spectra of storm bursts obtained from high-resolution radio spectrography of good sensitivity are needed. The time profile, frequency, and frequency drift should be measured accurately by the spectrographs operating in some selected frequency bands and at suitable locations.

(*b*) Simultaneous observations in optical, ultraviolet, x-ray, and radio bands are needed for getting information on the coronal structures, which are the sources of noise storms.

(*c*) As the physical processes involved in stationery Type-IV continuum and Type-I noise storms are thought to be same, the existence of the link between the two phenomena should be clarified from observational point-of-view.

(*d*) The inter relationship between metric and decametric storms should be studied.

(*e*) Clarification of the degree to which scattering influences observations of Type-I emissions is needed.

Type II

(*a*) The details of acceleration mechanism for electrons at the shock front are not known till today.

(*b*) It is still not clear whether the emission comes from both the front and back of the shock.

(*c*) The inter relationship between the Type-II exciter and the outwardly moving mass ejected in the coronal transient is unclear.

(*d*) It is not clear why some shocks in the interplanetary medium lead to Type-II emissions, while others do not.

Type III

(*a*) Quasi-linear acceleration of the electron distribution under the influence of the Langmuir waves would cause the beam to stop in a very short distance (1 km) and time (nanosecond). But these values are much too small to account for the observed Type-III behavior.

Type V

(*a*) The hypothesis that Type-V emission is due to plasma radiation from electrons trapped in high coronal loops adjacent to the open field lines traced by Type-III electron, has been questioned mainly because Type-V radiation extends to frequencies of 10 MHz or lower.

(*b*) It is not clear why Type-V source positions in many cases differ from those of the preceding Type-III.

(*c*) It is not clear why Type-V radiation is prolonged.

(*d*) The cause of the polarization reversal from Type-III to Type-V is not clear.

Type U and Type J

(*a*) One of the unresolved issues regarding Type-U bursts is that it is still unknown whether U-bursts are F (fundamental) or H (harmonic).

16.11 A FEW INTERNATIONAL RADIO-SPECTROGRAPHS

In the following sections we describe some of the radio-spectrographs of today. At the end we list most of the current solar arrays, radio-spectrographs and radio-heliographs in Table 16.2.

Table 16.2: Some current solar arrays, radio-spectrographs, and radio-heliographs.

S.N.	Name and Country	Frequency of operation	Spectral channels	Time Resolution
1.	Solarly Radio Observatory Potsdam Tremsdorf Germany	40 to 170 MHz 200 to 800 MHz	256	0.1 s
2.	Ichon Radio-spectrograph (IR) Korea	30 MHz to 2.5 GHz	≈ 4033	≈ 0.5 s
3.	Green Bank Solarly Radio Burst Spectrometer USA	20 to 70 MHz (plan:1050MHz)		1.0 s
4.	Siberian Solarly Radio Telescope (SSRT) Russia	5.73 GHz		30 s to 180s
5.	Digitally Spectro-polarimeter (DSP) Graz, Austria/Kharkov, Ukraine	10 MHz to 30 MHz	1000	0.002 s

6.	ARTEMIS-IV Multichannel Radio-spectrograph, Greece	20 to 650 MHz	758	0.1 & 0.01 s
7.	IZMIRAN Radio-spectrograph Russia	25 to 270 MHz	> 1225	0.04 s
8.	Solarly Broadband Radio-spectrograph (SBRS) China	0.7 to 7.6 GHz	1292	0.001 s to 0.01 s
9.	USAF Radio Solarly Telescope Network (RSTN) USA, Australia, Italy	25 to 180 MHz	826	3 s
10.	Solarly Radio Burst Locator (SRBL) USA	18 MHz to 1GHz	120	4.8 s
11.	Nobeyama Radio-heliograph (NoRH) Japan	17 & 34 GHz		1 s & 0.1 s
12.	Metsähovi Radio Telescope Finland	22.6 & 36.8 GHz		600 s
13.	Frequency Agile Solarly Radiotelescope (FASR) USA	30 MHz to 30 GHz		0.01 to 0.1 s
14.	Ondrejov Radio-spectrograph Czechoslovakia	0.8 to 4.5 GHz	512	0.1 s
15.	Gauribidanur Digital Spectrograph (GDS), Bangalore, India	30 to 80 MHz		3 s
16.	Culgoora Radio-spectrograph (CULG) Australia	18 MHz to 1.8 GHz	2044	3 s
17.	Nancay Radio-heliograph (NRH) France	151, 164, 237, 327, 411 & 432 MHz		0.125 s
18.	Solarly Submillimetric Telescope (SST) Argentina	405 MHz & 212 GHz		300 s
19.	Phoenix-2 Radio-spectrograph Switzerland	0.1 to 4 GHz	2000	1 s
20.	Hiraiso Radio-spectrograph (HiRAS) Japan	25 MHz to 2.5 GHz	501	3 s

21.	Bruny Iceland Radio Spectrometer (BURST) Australia	3 to 50 MHz	128	3 s
22.	Trieste Solarly Radio System (TSRS) Italy	0.1 to 10 GHz	6	0.001 s
23.	Gauribidanur Radio-heliograph (GRH) India	40 to 150 MHz		
24.	Owens Valley Solarly Array (OVSA) USA	1 to 18 GHz	45	8 s
25.	Nancay Dekameter Array France	20 to 70 MHz	> 1000	1.25 s

Hiraiso Radio-spectrograph (HiRAS)

The HiRAS consists of three antennas known as HiRAS 1, HiRAS 2, and HiRAS 3 covering the frequency ranges 25 to 70 MHz, 70 to 500 MHz, and 500 to 2500 MHz, respectively. HiRAS 1 is made of spatially crossed 13 elements log-periodic antenna; HiRAS 2 has a 10 m parabolic dish with a crossed 20 element log-periodic feed; and HiRAS 3 is a 5 m parabolic dish with a crossed 23 element log-periodic antenna-feed. Both the LCP and RCP signals coming from the antenna are amplified with the help of LNAs (Low Noise Amplifiers) and are then fed to spectrum analyzers with a time resolution of 2 to 4 sec. The spectrum data are collected by a system control computer (HP9000/R362) through GPIB and are transferred to a workstation (HP 9000/710). Fig. 16.8 shows the various antennas of this spectrograph.

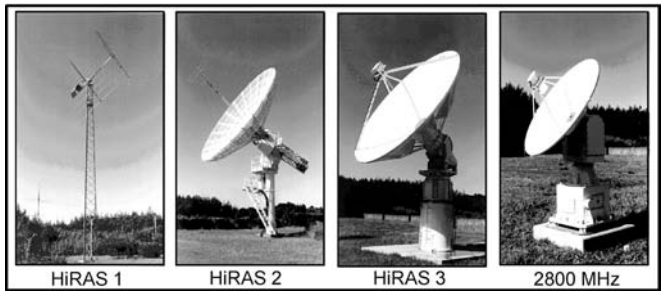


Figure 16.8: Various antennas of the HiRAS radio-spectrograph.

Culgoora Radio-spectrograph

The Culgoora radio-spectrograph is being operated in the four ranges of frequency 18 to 57 MHz, 180 to 570 MHz, and 570 to 1800 MHz designated as

A-, B-, C-, and D-bands, respectively. The antenna for the A-band is a dual crossed log-periodic array with two horizontal and vertical feeds. In the B-band, an open 13 m parabolic dish, with a crossed log-periodic feed, which gives horizontal as well as vertical components over this band. For the C-band, a solid 5 m parabolic dish, with a crossed log-periodic feed provides both horizontal and vertical components. The signals are amplified by using Anzac AMC-162 amplifiers in the A-band, Anzac AMC-117 amplifiers in the B-band, and Anzac AMC-178 amplifiers in the C and D bands and are then fed to respective spectrum analyzers (Tektronix 2710) meant for the aforesaid bands. The spectrum analyzers are controlled by a PC through GPIB. The control PC initiates a sweep of each analyzer in sequence where one sweep consists of 512 samples, with an 8-bit amplitude resolution, and takes 0.75 seconds. In this way the full frequency range is sampled once in every 3 seconds. After some processing, the signals are displayed on the main observatory computer, which is a Sun Microsystems Sparc workstation fitted with an 8-bit color display card. In order to track the Sun across the sky the antenna drives are controlled by utilizing an XT PC, which runs a program providing the right ascension and declination of the Sun. Fig. 16.9 shows some of the antennas of this spectrograph.

Figure 16.9: Some antennas of the Culgoora radio-spectrograph.

Zurich ETH Spectrograph

The new spectrometer known as *Phoenix* consists of a parabolic dish of 7 m diameter with a crossed log-periodic feed in the primary focus, which covers the range of frequencies 0.1 to 4 GHz with a gain of about 7 dB. In order to convert the linearly observed polarization into circular polarization the front end contains two

90-degree hybrids for the ranges above and below 1 GHz, a preamplifier with a noise figure less than 4.6 dB, and several PIN diodes and mechanical switches controlled by the receiver. In the receiver, the signal is amplified and then up-converted by a 1000 channel direct synthesizer having the range of frequency 11.2 to 12.2 GHz and the switching speed of 0.45 μs. In order to suppress the image frequencies in the four GHz bands, four 60 MHz band-pass filters with center frequencies of 8.17, 9.17, 10.17, and 11.17 GHz are used and one of four corresponding quartz-stabilized oscillators at 8.1, 9.1, 10.1, and 11.1 8Hz is then selected to mix the signal down to the second intermediate frequency at 70 MHz, where it is amplified and reduced to 1, 3, or 10 MHz bandwidth. Fig. 16.10 shows some antennas of this spectrograph.

The duty cycle of 500 μs is divided into four integrating phases of 106 μs each and several switching times are used to select frequency, band-width, and signal source in the front end and A/D conversion. Thus, a scan in frequency containing 1 to 500 observing channels is obtained with a time resolution of 0.5 to 250 μs at one frequency. The timing of the spectrometer is controlled by an atomic clock.

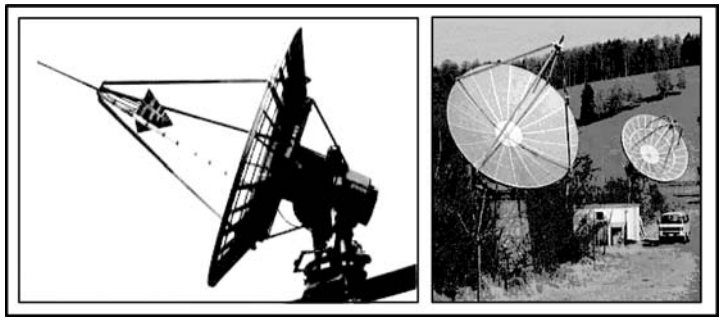


Figure 16.10: Some antennas of the Zurich spectrograph.

Nancay Radio-spectrograph

This is a multi-channel radio-spectrograph operating in the frequency range of 110 to 470 MHz, which has been subdivided into 36 bands of 10 MHz band-width. Each of the 10 MHz band-width amplifiers feeds a set of 10 mixers, with local oscillator frequencies differing from each other in steps of 1 MHz. The amplifiers with 1 MHz band-width are connected with 120 mixers out of the 360 signal outputs which give rise to 120 points in spectrum with 1 MHz resolution. In order to get higher frequency resolution a further division of frequency can be made by the simultaneous use of 100 more amplifiers of 200 kHz band-width.

The radio-heliograph presently consists of a cross shaped multi-antenna array. It has a branch running along the East/West line consisting of 19 antennas. The

baseline is 3200 m long. It has a branch along the North/South line consisting of 24 antennas over a length of 1250 m. Fig. 16.11 shows some of the antennas of this spectrograph.

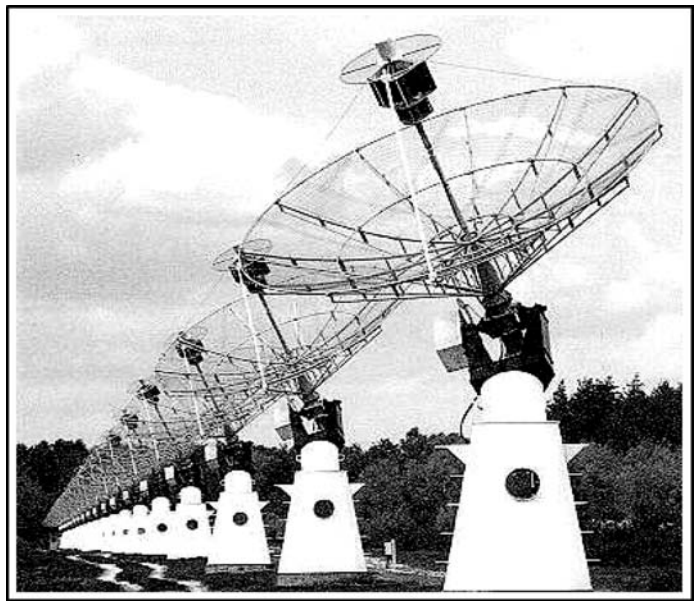


Figure 16.11: Some of the antennas of the Nancay radio-spectrograph.

Weissenau Spectrograph

This spectrograph consists of four parts: (*a*) log-periodic antennas and a parabolic dish, (*b*) structural steel bridge, (*c*) mechanical drive mount system, and (*d*) electronic equipment. The antennas system is operated in the frequency ranges 30 to 46, 46 to 86, 86 to 160, and 160 to 1000 MHz with the help of 2 LPDAs, 4 LPDAs, 8 LPDAs, and 7 m parabolic dish with an LP-feeder made of a pyramidal LP plate structure, respectively. The antennas are mounted on a steel bridge, supported by 4 legs running on 2 profile wheels on a circular steel bridge. The radiometer is a swept radiometer. Six radiometer trains operate in parallel over bands identical with the antenna intervals.

Considering the different frequency scales one realizes that the Weissenau record does show only the red features of the Tremsdorf record, while the green features and part of the yellow features do not appear on the WEIS record. Indeed, the band 40 to 90 MHz seems to reproduce weaker than the corresponding band on the WEIS record contrary to the band 100 to 170 MHz. So the fundamental band is patchy on the color plot.

Gauribidanur Radio-spectrograph

This instrument operates in the band of 40 to 150 MHz. It consists of a *T* shaped array stretching 1.2 km along the East/West line and 0.45 km along the South. The signals from individual antennas are high pass filtered with a cut-off at 40 MHz and fed to broadband amplifiers of a 30 dB gain and a noise figure of 3 dB. There are 128 log-periodic antennas appearing in 16 groups of 8 antennas along the East/West direction. The internal spacing between the antennas is kept as 10 m. The Southern arm consists of 64 antennas arranged in 16 groups each consisting of 4 antennas. The internal spacing in this case is 7 m. All the 32 outputs from the antennas (16 from the East/West and 16 from the South) are taken into the receiver room where it is down converted to 10.7 MHz having 1 MHz bandwidth. This is then digitized and correlated in a 1024 channel correlator system. The resulting output is stored in digital media. The spatial resolution is 5 × 8 arc-minute at 150 MHz and the sensitivity is nearly 200 jansky for a 1 MHz bandwidth with 5 sec integration. Fig. 16.12 shows the Western arm antennas of this spectrograph.

Figure 16.12: The western arm antennas of the Gauribidanur radio-spectrograph.

EXERCISES

1. Explain how the Earth's atmosphere is influenced by solar radiation.
2. Outline the concept of a global electrical circuit.
3. What are the three different types of solar radiations? How do they contribute to the Earth's atmosphere?
4. What is diffuse radiation? What are the important factors controlling this radiation?

5. Explain how the ionizing radiation emitted from the flare region affect the ionosphere and the radio communication system.
6. Discuss how the D-region of the ionosphere is affected by solar radiation.
7. Explain with a suitable diagram how the atmospheric electrification is modulated by solar radiation, pointing out the influence of ionizing radiation.
8. Briefly discuss the various associated parameters of diffuse radiation.
9. Describe a Moll-Gorczynski pyranometer for continuous registration of diffuse solar radiation.
10. What are the different factors controlling the net radiation? Explain them briefly.
11. What is meant by facing of radio waves? Discuss the effects of geomagnetic disturbances of solar origin.
12. What are sprite flashes? Discuss various visual reports of sprite flashes.
13. Summarize the observational reports of different type of sprites.
14. Draw a schematic diagram showing the connection among sprites, positive CG stokes, and Q-bursts and explain.
15. Summarize the important aspects of meter-wave radio bursts.
16. Why does the HiRAS consist of three antennas? Explain the role of each antenna.
17. Describe the operation of the new Culgoora radio-spectrograph and the Zurich ETH spectrograph.
18. Explain the multi-channel radio-spectrograph of Nancay.
19. What are the four different parts of the Weissenau spectrograph? Discuss.
20. Discuss the radio-spectrograph of Gauribidanur.

Chapter 17

The Life of Stars in the Main Sequence

17.1 INTRODUCTION

From the Hertzsprung-Russell diagram in Chapter 5, it is seen that the majority of stars are found in the main sequence within the temperature/spectral classification ranging from Type O to Type M. It is, therefore, understood that stars spend a large fraction of their time of existence as a main sequence star. Our Sun, a very ordinary G2 star, is one of them. The Sun must have maintained its luminosity more or less constant until today. Otherwise, the existence of life on Earth would not have been possible since the temperature of water on Earth has to be maintained within limits for evolution of life. The formation of the stars has also been described in Chapter 5. With the knowledge we have gathered mainly from the Sun and other stars, we shall examine the physical properties of a main sequence star. Once we understand the processes that dictate the structure of main sequence stars, it will be easier for us to investigate more about their evolution.

17.2 STELLAR STRUCTURES

The structure of a star is highly dependent on its hydrostatic equilibrium, radiation transport mechanism, and nuclear energy generation. These again depend on the mass of the star and the constituent elements. In 1926, Henry Norris Russell and Heinrich Vogt derived an interesting result, known as the Russell-Vogt theorem. It states that "the structure of a star is uniquely determined by the star's total mass and by its chemical composition." In other words, given only these two factors all other physical aspects of a star can be calculated. To put it in another way, we may say that a certain mass of stellar material composition can only reach a stable configuration. The areas of most uncertainty are the behavior of a star's opacity and convection. These are some of the major areas of scientific research.

17.2.1 Stellar Models

Let us look into physical conditions such as temperature, pressure, density etc., existing in a typical star and the way they vary within the star. The directly measurable parameters, as in the case of our Sun, are only for the outer layers such as the photosphere, chromosphere, and the corona. But their variation below the photosphere can't be studied through direct observation since those regions are inaccessible. Thus, we have to depend on indirect information like the results of helioseismology. But for the stars other than the Sun, the results obtained from asteroseismology provide still lesser information as the discs of stars can't be resolved. Still for understanding the process of evolution of different stars, it is not necessary to have an understanding of the interior conditions since these properties get reflected on the nature, rate, and the extent of the nuclear processes of a star. A set of equations known as the equations of the stellar structure are constructed, which help in predicting the various physical properties like temperature, density, and pressure distributed within the star. These are simple equations but the associated theory and the techniques adopted for their solution are complex. But by some simplified assumptions, it is possible to derive some limited quantitative information about the stellar structure and also study the basic properties of stars.

A star loses a fraction of its energy by radiation from its surface, which indicates that the temperature should increase with depth into a star. The pressure should also increase accordingly due to the increasing amount of overlaying materials with an increase in depth. The elemental composition of a star and its mass determines the thermal conductivity, opacity, pressure gradient, and the rate of energy generation. They are all interrelated and it is one of the major challenges of stellar physics to determine these relationships accurately.

17.2.2 Equations of Hydrostatic Equilibrium Inside a Star and Temperature of the Core

Fig. 17.1 assumes that the star is composed of ideal hot gases above the core and is spherically symmetric. Due to high temperature, the gases are expelled outward, which is counteracted by the self-gravitational pull within the star, which is dependent on its mass. Consider a shell of very small thickness. Let Q be the flux (energy per unit area per unit time) flowing out radially. To achieve this condition there must be a temperature difference between the inner surface and outer surface of the shell. Let the shell's outer surface temperature be T_2, pressure P_2, and the inner surface temperature be T_1 and pressure P_1. Obviously, T_1 should be greater than T_2 to maintain an outward flow of energy and the temperature increases with an increase in depth. Let the thickness of the shell be dr and the mass of the shell be dM_r.

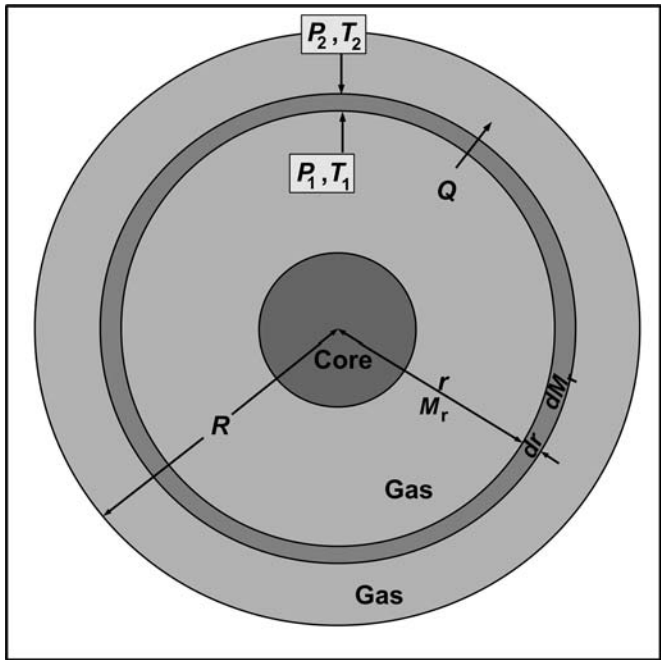


Figure 17.1: A simple diagram for deriving the equations for hydrostatic equilibrium in a star. The star consists of a small core (where the nuclear reactions take place), which is surrounded by a sphere of hot gases. The core is much smaller than shown in the diagram.

Since in reality the gas is ionized on account of high temperature, and the ionized particles and electrons behave as ideal particles, we can use simple equations. Thus, the gas pressure P can expressed based on the ideal gas law as

$$P = \frac{k\rho T}{m}. \tag{17.1}$$

The quantity ρ is the gas density, m is the mass, T is the temperature, and k is the Boltzmann constant. Let us now find the variation of the pressure at various distance r from the center of the star. The balance between the gravitational pull and the pressure for ejecting the gas outward is called the *hydrostatic equilibrium*. Consider a star of mass M and radius R composed of ideal gas. If P_r is the pressure at a distance r from the center of the star, and dr is a small elemental length of r as shown in Fig. 17.1, then the change of pressure with distance from the center of the star can be expressed as

$$\frac{dP_r}{dr} = -\frac{GM_r \rho_r}{r^2}. \tag{17.2}$$

The quantity M_r is the mass inside the radius r, G is the gravitational constant, and ρ_r is the matter density at distance r. Also, the equation of mass continuity can be expressed as

$$\frac{dM_r}{dr} = 4\pi r^2 \rho_r. \tag{17.3}$$

It is required to complement the above equations with an equation of state, which expresses a relationship between pressure, temperature, and density. Since the gas is assumed to be ideal, the pressure can be expressed as

$$P = nkT = \frac{\rho}{\mu m_H} kT. \tag{17.4}$$

The quantity μ is the mean molecular weight of the ionized gas, which is the average atomic mass per particle, n is the number of particles per unit volume ($n = N/V$, where N is the total number of particles in a volume V), and m_H is the mass of a neutral hydrogen atom. A hydrogen atom shall yield one electron and one proton (two particles). It can thus be written

$$n_H = 2\frac{\rho_H}{m_H} = 2\frac{\rho X}{m_H}, \tag{17.5}$$

where X is the mass fraction of hydrogen. Similarly, a helium atom shall yield one atomic nucleus and two electrons (three particles). The helium atom has a mass of $4m_H$ and thus we get

$$n_{He} = \frac{3}{4}\frac{\rho_H}{m_H} = \frac{3}{4}\frac{\rho Y}{m_H}, \tag{17.6}$$

where Y is the mass fraction of helium. Finally, all the elements heavier than helium are considered metals by the astronomers. In general, a metallic atom shall give off A electrons as it is ionized and its mass is $2\,Am_H$ roughly. Thus, we obtain

$$n_m = A\frac{\rho_H}{2Am_H} = \frac{1}{2}\frac{\rho Z}{m_H}. \tag{17.7}$$

The total pressure can be now expressed as

$$\left.\begin{aligned} P &= n_H kT + n_{He}kT + n_m kT \\ &= \left(2X + \frac{3}{4}Y + \frac{1}{2}Z\right)\frac{\rho kT}{m_H} = \frac{\rho kT}{\mu m_H} \end{aligned}\right\}, \tag{17.8}$$

where the molecular weight is written as

$$\frac{1}{\mu} = 2X + \frac{3}{4}Y + \frac{1}{2}Z. \tag{17.9}$$

The typical values of the mass fractions are $X = 0.7$, $Y = 0.28$, and $Z = 0.02$. However, the value of Z can vary significantly depending on the class of the star. The velocity distribution of the particles is generally described by Maxwell-Boltzmann distribution. In certain stars, the temperature is high enough so that the random motion of the photons can significantly contribute to pressure. Assuming that the radiation from a star follows the black-body spectrum, the radiation pressure P_{rad} can be expressed as

$$P_{rad} = \frac{1}{3}\alpha T^4. \tag{17.10}$$

The quantity α is a constant and is equal to 7.55×10^{-16} N m^{-2} K^{-4}. Since the core of the star is much smaller compared to its entire volume, its temperature can be expressed as

$$T = \frac{Gm_{av}\,M}{3\,kR}. \tag{17.11}$$

The quantity m_{av} represents the total mass M of the star divided by the total number of particles N in the star ($m_{av} = M/N$). If we substitute the values for the Sun in Eq. (17.11) we find that the temperature of the core of the Sun is $T \approx 4 \times 10^6$ K. This is only an approximate value. A more detailed analysis shows that the temperature of the core of the Sun is $T \approx 1.56 \times 10^7$ K.

17.2.3 The Energy Transport Mechanism Inside a Star

The nuclear reaction takes place at the core of a star, where the temperature is highest and is the source of energy. The size of the core varies with the mass of the

star. The energy thus generated is transported outward by means of convection, conduction, and radiation. For a main sequence star the convection and radiation dominate in energy transport. The conduction is relatively unimportant in the stellar interiors. The efficiency of the radiation process depends on the opacity of the material that the radiation must traverse. Opacity depends on the temperature, density, and elemental composition of the gas in a complicated way. The convection depends on the viscosity of the gas and any forces like gravity, which tend to resist the motion of convection. Convection takes over when the opacity is high thereby preventing the radiation type of transfer. Much detail about the efficiency of convection has not been understood until today and is quite possibly still a major uncertainty in the study of stellar structure.

The main sequence stars have zones at different radii, which are convective and zones, which are radiative. Their location depends on the opacity and other properties of the stars. Stars in the main sequence with a mass less than $\approx 1.5\, M_\Theta$ are called *lower main sequence stars*. The low mass stars ($\approx 1\, M_\Theta$ or less) are convective in their outer layers and radiative in their cores. Their core is non-convective. Just outside the core, the temperature gradient is very small and the convection is negligible. Therefore, the transfer of energy takes place through radiation near the core. As we move outward, there exists a region where the convection takes place. For stars with very low mass, this convective envelope begins deeper inside. If the mass is less than $\approx 0.3\, M_\Theta$, the convective envelope approaches right into the center.

Stars in the main sequence with masses greater than $\approx 1.5\, M_\Theta$ are called *upper main sequence stars*. Since the temperature of the core is sufficiently high for a different set of nuclear reactions to predominate, the energy release at the core is sufficiently concentrated to trigger a convective instability. Thus, the center of the star is convective. The convective zone may even exceed the core. Intermediate mass stars ($\approx 3.5\, M_\Theta$) may be radiative throughout. Massive stars having several M_Θ are convective deep in their cores and are radiative in their outer layers. Fig. 17.2 shows the comparison between the locations of the convective zones within stars of $1\, M_\Theta$ and $5\, M_\Theta$ masses.

If a hot corona exists like that in the Sun, it could be associated with active convection in the outer layers. The depth of the convective layer determines the extent to which material from the deep interior of the star is mixed into the outer layers. Since the interior material is most likely to have undergone nuclear reactions, which also change the elemental abundances, it affects the abundances in the star's atmosphere. This can also affect the composition of the interstellar gas and may be observed in the stellar spectra. It can also get ejected out of the star through stellar wind.

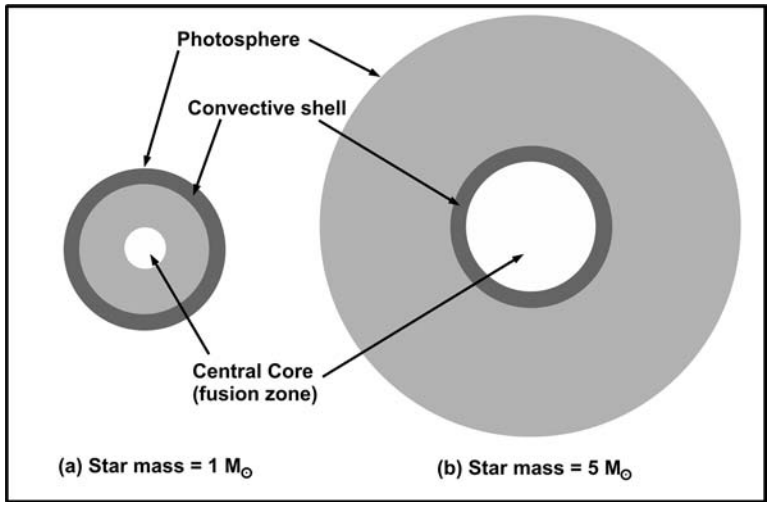


Figure 17.2: The cross-sections of stars having different masses. (a) A star like our Sun. (b) A star of five solar masses.

17.2.4 The Main Sequence and Lifespan of Stars

All stars in the main sequence burn hydrogen into helium in their cores through nuclear fusion. The fusion creates heavier elements from lighter elements like helium from hydrogen. It also releases energy (photons), which travel outward and escape from the surface. The photons in the process of traveling outward exert a pressure on the stellar material, which is called *radiation pressure*. The radiation pressure pushing outward exactly balances the gravitational pressure pulling inward and a hydrostatic equilibrium is maintained as explained previously. Stars spend 90% of their lifetime in the main sequence. This absolute main sequence lifetime of a star is proportional to $M^{-2.5}$, where M is the mass of the star. Thus, more massive stars die fast. The reason being massive stars have much higher central core temperatures due to higher pressure and thus burn their fuel faster. Hence, they are more luminous. For a star like our Sun having a mass of 1 M_Θ, the main-sequence stage lasts about 10,000,000,000 years, whereas a star having a mass of 10 M_Θ will be 1,000 to 10,000 times brighter but will last only about 20,000,000 years. A star having a mass of 0.1 M_Θ may only be 1/1,000th to 1/10,000th of its brightness, but will last about 1,000,000,000,000 years. If the brightness of a main sequence star has to increase twice, then it has to burn its fuel 10 times faster. Table 17.1 lists the approximate luminosities and approximate life periods of stars with different mass/radius in the main sequence. Please also see Eqs. (5.8) and (5.9).

Table 17.1: The life periods of main sequence stars having different masses/radii.

Mass (M_Θ)	Luminosity (L_Θ)	Effective temperature (K)	Radius (R_Θ)	Main sequence Life span (year)
0.10	3×10^{-3}	2,900	0.16	2×10^{12}
0.50	0.03	3,800	0.6	2×10^{11}
0.75	0.3	5,000	0.8	3×10^{10}
1.0	1	6,000	1.0	1×10^{10}
1.5	5	7,000	1.4	2×10^{9}
3.0	60	11,000	2.5	2×10^{8}
5.0	600	17,000	3.8	7×10^{7}
10.0	10,000	22,000	5.6	2×10^{7}
15.0	17,000	28,000	6.8	1×10^{7}
25.0	80,000	35,000	8.7	7×10^{6}
60.0	790,000	44,000	15	3.4×10^{6}

17.3 NUCLEAR REACTIONS INSIDE A MAIN SEQUENCE STAR

It is believed that after the Big Bang matter consisted mainly of hydrogen and helium. These elements were converted into different elements in the nuclear factories of the stars by a process called *nucleosynthesis*. Unlike the nuclear plants we have on Earth, stars do not produce energy by splitting atoms through nuclear fission. On the contrary, they join the atoms together through nuclear fusion. The requirement for the fusion process to start is an extremely high temperature in millions of K. In the cores of freshly born stars, the hydrogen atoms fuse to form the next heaviest element, helium. After the hydrogen gets exhausted the helium atoms fuse together to form carbon. And when the helium runs out the carbon atoms fuse to form oxygen. This manufacturing process of heavier and heavier elements continues until the star finally dies. Our Sun is likely to halt during the carbon-oxygen stage of fusion. Eventually, the stars that are a few times bigger than the Sun will have cores made entirely of iron. Elements heavier than iron are created in supernova explosions.

Let us now look more deeply into the nuclear reactions taking place in a main sequence star like our Sun. There are two basic reactions taking place, namely the Proton-Proton reaction and the Carbon-Nitrogen-Oxygen cycle as shown in Fig. 17.3.

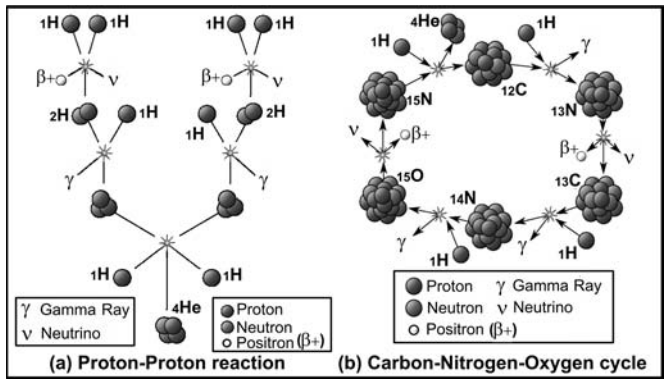


Figure 17.3: The temperature-dependent basic reactions taking place in a star. (a) Proton-Proton reaction. (b) Carbon-Nitrogen-Oxygen cycle.

17.3.1 The Proton-Proton (PP) Reaction

The PP reaction is the primary energy producer in stars like our Sun. Inside the core of the Sun where the density is about 100 times of water and the temperatures are about 1.5×10^7 K, the hydrogen is ionized into a state of plasma consisting of free electrons and protons. The heat provides sufficient energy to the hydrogen ions to collide among themselves with enough force to overcome the mutual repulsions due to like charges, and fuse. When two protons (1H) fuse, one of them decays into a neutron through β+ decay forming a deuteron (2H). The decay also releases an anti-electron (positron) and a neutrino (ν). The positron will collide later with an electron and annihilate each other thereby releasing two gamma rays (high energy photons). The neutrino interacts very weakly with any matter and will therefore pass out of the Sun. The freshly formed deuteron may collide with another hydrogen nucleus creating a helium isotope (3He) and releasing a gamma ray (γ). When two of the 3He isotopes collide, two of their protons are released thereby creating 4He. The sequence of the reactions can be written in three equations as shown below.

$$_1\mathrm{H} + {}_1\mathrm{H} \longrightarrow {}_2\mathrm{H} + \text{positron } (\beta+) + \text{neutrino } (\nu) \qquad (17.12\,a)$$

$$_2\mathrm{H} + {}_1\mathrm{H} \longrightarrow {}_3\mathrm{He} + \text{gamma-ray } (\gamma) \qquad (17.12\,b)$$

$$_3\mathrm{He} + {}_3\mathrm{He} \longrightarrow {}_4\mathrm{He} + {}_1\mathrm{H} + {}_1\mathrm{H} \qquad (17.12\,c)$$

It must be noted that Eqs. (17.12 *a*) and (17.12 *b*) must be done twice for every (17.12 *c*). 3He and 4He are stable isotopes. To fuse 4He higher energies are required since the repulsion between two pairs of protons in helium is even more than the repulsion between only two hydrogen nuclei.

17.3.2 The Carbon-Nitrogen-Oxygen (CNO) Cycle

This is another cycle of energy production, which dominates in massive stars since the temperature is much higher (greater than 1.6×10^7 K). It ultimately results in conversion of the hydrogen into helium. Some heavier elements are also present in the star in small quantities apart from the major elements hydrogen and helium. If carbon (C), nitrogen (N), and oxygen (O) ions exist, they may get into releasing energy within the stars with the reactions shown below.

$$_{12}C + _1H \longrightarrow _{13}N + \gamma \qquad (17.13\ a)$$

$$_{13}N \text{ (through } \beta + \text{ decay)} \longrightarrow _{13}C + \text{positron } (\beta+) + \text{neutrino } (\nu) \qquad (17.13\ b)$$

$$_{13}C + _1H \longrightarrow _{14}N + \gamma \qquad (17.13\ c)$$

$$_{14}N + _1H \longrightarrow _{15}O + \gamma \qquad (17.13\ d)$$

$$_{15}O \text{ (through } \beta+ \text{ decay)} \longrightarrow _{15}N + (\beta+) + \nu \qquad (17.13\ e)$$

$$_{15}N + _1H \longrightarrow _{12}C + _4He \qquad (17.13\ f)$$

The carbon acts as the catalyst since it initiates the chain of reactions but does not get consumed. Nitrogen and oxygen can also start similar reactions. According to the current estimate, the Sun produces 98% to 99% of its energy by means of the proton-proton reaction and 1% to 2% from the CNO cycle. It is presumed that CNO would be the major energy producer if our Sun was 50% more massive.

As said earlier, neutrinos interact weakly with matter and therefore escape out of the star. Detection of neutrinos would allow us to visualize a portion of the processes taking placing within the core of a star. Until today, the neutrinos detected by experiments are much less than predicted in theory. This only suggests that our understanding of Sun is incomplete. One possibility is that there could be three type of neutrinos and as they leave the core they may possibly change into a type that is undetectable using today's instruments. But definite proof is yet to come.

It can be seen from Fig. 17.4 that the relative amount of energy release from the two types of reactions described are dependent on the temperature of the core of the star. For temperatures below 1.8×10^7 K, the proton-proton chain reaction releases relatively more amount of energy than the CNO cycle. At a temperature of 1.8×10^7 K, both reactions release an equal amount of energy. Beyond this

temperature, the CNO cycle generates a relatively more amount of energy. In other words, the CNO cycle becomes the chief source of energy in stars of 1.5 solar masses or higher.

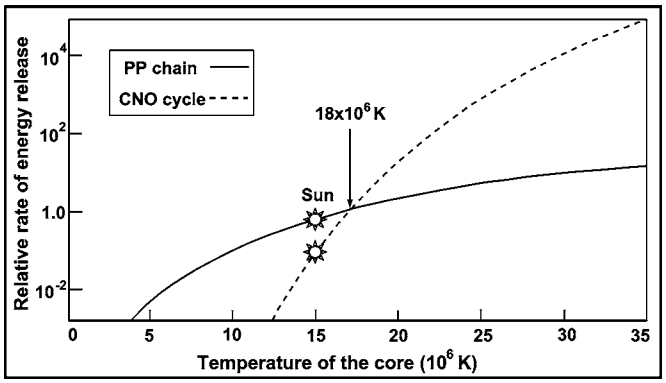


Figure 17.4: In a main sequence star, the proton-proton chain reaction generates a relatively more energy for core temperatures below 18×10^6 K. Above this temperature the carbon-nitrogen-oxygen cycle dominates in releasing energy. Thus, the major energy release mechanisms of the massive stars are the CNO cycles.

EXERCISES

1. State the Russell-Vogt theorem about the structure of a star and explain.
2. Outline the physical conditions existing in a typical star.
3. What is provided by the equations of the stellar structure?
4. Assuming that the star is composed of ideal hot gases, establish the equations of hydrostatic equilibrium in a star.
5. Explain the energy transport mechanism inside a star.
6. What is radiation pressure? Discuss briefly the main sequence and life span of the stars, with a special emphasis on the nuclear reactions inside a main sequence star.
7. Explain the temperature dependent basic reactions in a star considering:
   (*a*) Proton-proton chain reaction.
   (*b*) Carbon-Nitrogen-Oxygen cycle.

Chapter 18

The Post-Main Sequence Period of the Stars

18.1 INTRODUCTION

In this chapter, we will study what happens to a star immediately after its main sequence life comes to an end (i.e., immediately after the hydrogen fuel comes to an end). Some part of it may be scientifically guessed assuming that nuclear reactions were suddenly suspended. It takes about 1,000,000 years for the energy of a photon to reach the surface of the Sun. In its way, the original gamma ray photon loses its energy through interaction with the gases and is finally emitted at the surface as many ultraviolet and optical photons. Thus, if the nuclear furnace is suspended today, the luminosity of the Sun would probably remain more or less constant over a long time by human standards. But that is not the complete story. The star continues to undergo its next life sequence called the post-main sequence. This process of evolution is also very much mass dependent. We will learn about them in the following sections.

18.2 THE RED GIANT PHASE

Immediately after the hydrogen in the core gets converted to helium, the main-sequence life period of a star ends and the first part of the post-main sequence begins through the red giant phase. During a star's life, *ashes* of one nuclear-burning phase provide fuel for next burning phase. The star expands to a extremely large diameter and may possibly encroach the nearby orbiting planets. The temperature of the outer layer falls and the color of the star tend toward red. The star evolves into the spectral class K and M and their position in the H-R (Hertzsprung-Russell) diagram changes. This red-giant life of a star mainly consists of (*a*) a red giant branch phase, (*b*) horizontal branch phase, and (*c*) a asymptotic giant branch phase. During these processes, the helium in the core of the star and the surrounding hydrogen shell layers get into fusion with different combinations depending on the mass, pressure, and temperature, which characterize the temporal behavior of the star. The surface luminosity, temperature, and diameter of the star fluctuate. New elements are produced, some of which are thrown into space resulting in mass loss of the star. We shall study them in detail in the following sections.

18.2.1 The Red Giant Branch (RGB)

When the hydrogen in the core of a star is exhausted, the core contains only the heavier helium. The surrounding portion of the core still contains hydrogen. Since the nuclear furnace of the core has stopped, the outer gaseous regions start contracting under the self gravity of the star. The gravitational energy gets converted into heat. As we have seen in the earlier chapters the temperature within a star increases with depth and the hydrogen near the vicinity of the core gets heated up sufficiently to start the nuclear fusion again outside the core. This is also called shell burning of hydrogen and is shown in Fig. 18.1. This heats up its surrounding envelope and causes it to expand thereby increasing the diameter of the star. Due to large diameter, the surface of the star cools and in turn increases the opacity. Thus, heat travels outward through convection and reaches the star's envelope. The convective outflow causes the star's structure to change. The radius of the star increases and its surface temperature decreases. The decrease in temperature only increases the opacity so that at some point, most of the energy reaches the surface by convection thereby increasing the luminosity to a great extent. When plotted on the H-R diagram, the position of a star like our Sun will be in the region of the giants (see the RGB or red giant branch in Fig. 18.3). The star is now a red giant. It is red because it is cooler than it was in the main sequence star stage and it is a giant because the outer shell has expanded outward. Note that, when our Sun becomes a red giant its radius will be approximately 0.5 AU; that is, about 100 times its current size. It may even possibly occupy the Earth's orbit within its radius. A good example of a red giant is Alderban in the constellation of Taurus.

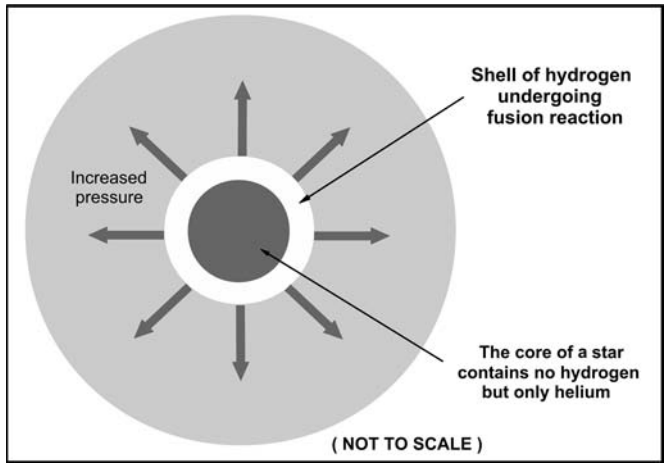


Figure 18.1: Post-main sequence hydrogen fusion.

Most of the mass of a red giant is concentrated in a dense core having a radius equal to few Earth's radii at temperatures of 50×10^6 K. The core is so dense that the electrons inside it turn to a degenerate gas.[15] The pressure of a degenerate gas depends on its density and is independent of temperature. It is this pressure that stands against the gravitational collapse.

While the hydrogen keeps burning around the core, the latter continues to contract and more gravitational energy is converted to thermal energy. The temperature continues to rise up in the core and when it reaches 100×10^6 K, the helium in the core starts a new range of nuclear fusion reactions called the triple alpha process. The heat spreads in the core very rapidly through conduction. The rest of the core quickly ignites. If the core were an ordinary gas, the explosive ignition would expand it due to the rapid increase of temperature and pressure. But the core is a degenerate gas and so the increased temperature does not increase the pressure. The increased temperature increases the rate of the triple-alpha process generating more energy and continues to further increase the temperature. This out of control process in the core is called the helium flash (see it on the H-R diagram in Fig. 18.4) and is abbreviated as HF. When the temperature of the core reaches about 350×10^6 K, the electrons become non-degenerate. The core expands and cools.

[15] Gases having a density below 100,000 g/cm^3 are perfect gases since the number density of electrons compared to possible energy states is very low allowing electrons to readily gain and lose energy. Gases with still higher density behave as degenerate electron gases since the number density of electrons compared to possible energy states is approximately the same and electrons cannot readily gain or lose energy since the Pauli exclusion principle prevents two identical electrons from occupying the same energy state.

18.2.2 The Triple-alpha Process

The reaction in a triple-alpha process is shown in Fig. 18.2. Two alpha particles (helium nuclei) fuse to form a ^{8}Be nucleus. But a ^{8}Be is very unstable and almost decays back to two ^{4}He as shown in Eq. (18.1 *a*). However, the high density and temperature of the core can cause another ^{4}He to strike the ^{8}Be (before it decays) resulting in the formation of a ^{12}C nucleus and emitting two γ photons as shown in Eq. (18.1 *b*). Since in this reaction three alpha particles get converted to one ^{12}C it is called the triple-alpha process.

$$_{4}\mathrm{He} + {}_{4}\mathrm{He} \Leftrightarrow {}_{8}\mathrm{Be} \qquad (18.1\ a)$$

$$_{8}\mathrm{Be} + {}_{4}\mathrm{He} \longrightarrow {}_{12}\mathrm{C} + 2\gamma \qquad (18.1\ b)$$

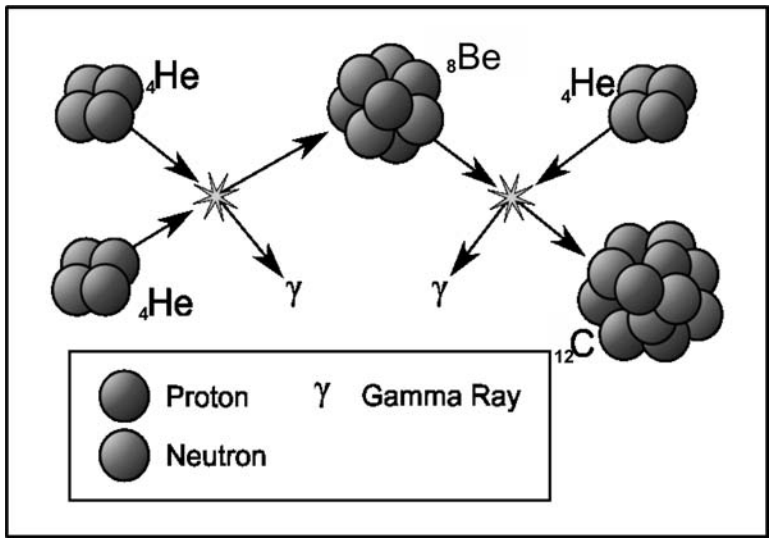


Figure 18.2: The triple-alpha process.

18.2.3 The Horizontal Branch (HB)

After the helium flash, the luminosity and the radius of the star starts decreasing and its position on the H-R diagram begins to shift, which is known as the Horizontal Branch (HB). After this phase, when plotted in the H-R diagram, the position of the star will be slightly below and left to its previous position. During this period, both helium and hydrogen burn, respectively, in the core and the shell enclosing the core. This phase of core-helium-burning is an analog of the main sequence phase of core-hydrogen-burning. The central region of the core gets converted to carbon faster by means of the triple-alpha process.

18.2.4 The Asymptotic Giant Branch (AGB)

Ultimately, all the helium in the core gets fused into carbon and oxygen and thus the core contracts again. Since the carbon and oxygen nuclei contain more protons than helium, the coulombic repulsion is larger. The temperature required to fuse these into heavier nuclei is larger than 100×10^6 K (required to fuse helium). In stars of mass less than 8 M_Θ, the gravitational energy is insufficient to compress and thereby raise the temperature for the next range of fusion. Thus, no more fusion takes place. However, the core contraction generates a sufficient temperature for fusing the surrounding layers of helium. Thus, helium-shell burning starts. The energy thus generated from shell-helium-burning heats up the surrounding fresh hydrogen, which also starts burning as a shell surrounding the helium-burning shell. This is shown in Fig. 18.3.

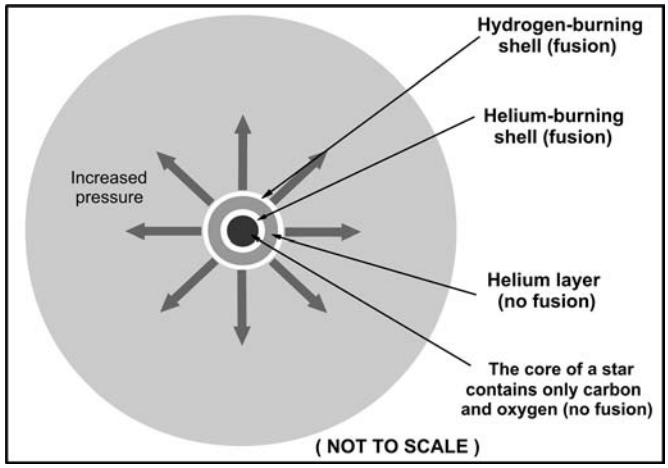


Figure 18.3: The giant star expands once again. This time the envelope may possibly increase up to 1.5 AU, equivalent to the orbit of Mars. The star is now an asymptotic giant branch star (AGB), sitting on the upper right portion of the H-R diagram (see its phases in Fig. 18.4). The luminosity of a 1 M_Θ AGB is about 10,000 times that of our Sun. An example of an AGM star is Mira, also known as *o Ceti*.

18.2.5 Thermal Pulses in AGB Stars

The outer layers of AGB stars are only weakly held by gravity. The helium-burning shell makes the star unstable since the rate of the triple-alpha process is extremely sensitive to temperature fluctuations. If the star contracts a bit, the temperature in the layer increases thus enabling more energy release through the triple-alpha process. This will increase the pressure more than that required for compensating the gravitational pull. Therefore, the outer parts of the star expand, which results

in a decrease of temperature and pressure, which in turn reduces the energy release rate. Thus, the star again contracts and the cycle repeats. This process of expansion and contraction of the stars through bursts of thermonuclear explosions are called *thermal pulses*. This phase is marked in the H-R diagram of Fig. 18.4. During this phase the nuclei within the helium-burning shell can be synthesized into heavier nuclei through the capture of neutrons and radioactive beta decay. This so-called the *s-process* (s for slow, in comparison with the rapid and the *r-process* that occurs in more massive stars) produces even heavier elements like bismuth with 83 protons. The frequency of the thermonuclear explosions is about one in ten thousand to a hundred thousand years. The effects can be observed on the luminosity of the star, which dips rapidly by 20% to 50% in several years time. Each blast of energy consists of bubbling gases containing elements fused in each explosion and moves outward through convection.

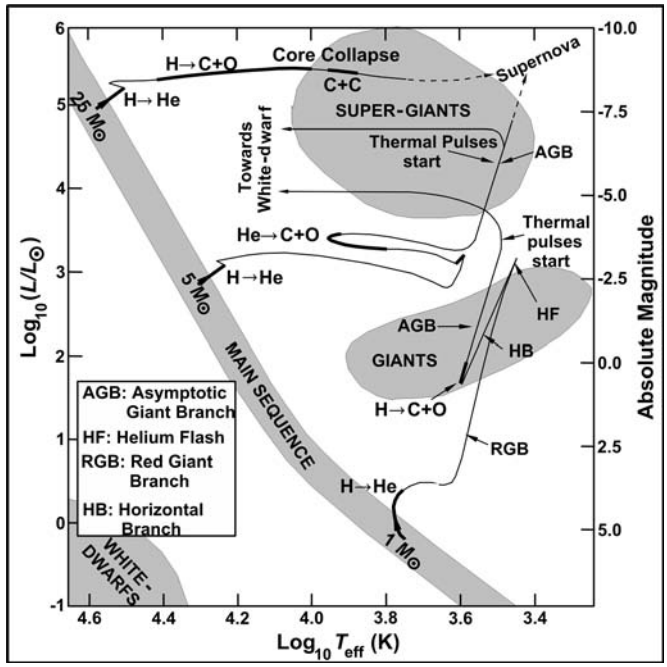


Figure 18.4: A theoretical picture of red giant phases of 1 M_Θ, 2 M_Θ, and 5 M_Θ stars. The dark lines indicate major phases of fusion. The dashed lines indicate uncertain phases. (Adapted from a diagram by I. Iben, Jr.)

18.2.6 Mass Loss from AGB Stars

Mass loss in the form of molecular gas and dust is one of the key characteristics of AGB stars. Huge convection currents inside AGB stars carry the material produced

in the thin helium-burning shell up to the surface of the star. These heavier nuclei can be detected in the star's spectrum, which provides insight into the undergoing processes deep within the star. The outer layer of an AGB star loses a mass of about $10^{-4}\ M_{\odot}$ per year due to increase in pulsation (this is about 100 times more than in the case of RGB stars). Mass loss is an important mechanism that enriches the interstellar medium with matter processed in the interior of AGB stars. This is responsible for the chemical evolution of galaxies and the Universe as a whole. The production of stardust is the starting point of the formation of all solid bodies ranging from planetesimals to asteroids and planets. The ejected material is comprised of mixtures of carbon, oxygen, and other elements. The carbon rich molecules from the stardust tend to shroud the star in the form of clouds. As the dust clouds expand, they absorb ultraviolet, visual, and near-infra-red radiation very efficiently and re-emit them at longer wavelengths (mainly in the range of 5 to 100 μm). Thus, AGB stars are more luminous in the infra-red than at visible wavelengths. The expanding cloud can also be observed at radio wavebands.

18.3 ALL STARS PASS THROUGH THE RED GIANT PHASE

Do all stars become a red giant? The answer is likely to be yes. Up to the red giant phase, stars with any mass evolve in a more or less similar fashion. The only

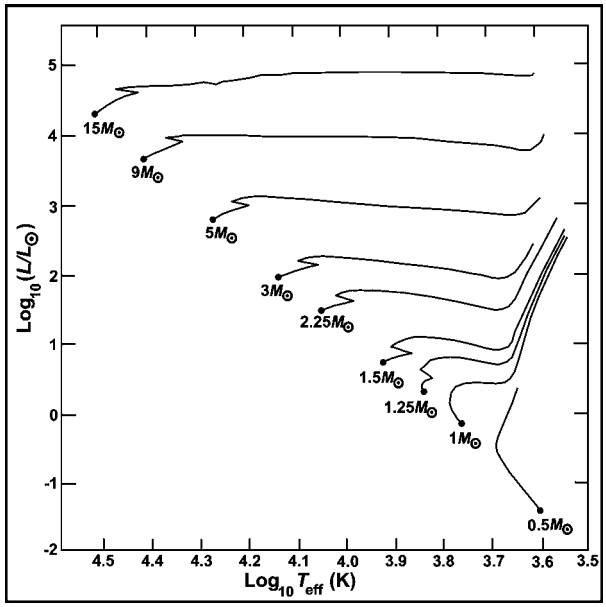


Figure 18.5: Stars of different masses reach the red giant (or red super-giant) phase in a similar fashion. (Adapted from a diagram by I. Iben, Jr.)

difference is the size of the giant. In the case of a massive star it evolves as a red super-giant instead of a red giant. It may be noticed in Fig. 18.4, a star with 1 M_Θ enters into the giant area, whereas stars with 5 M_Θ or more enter into the super-giant area of the H-R diagram. Fig. 18.5 shows the phases of stars having different masses reaching the RGB in a similar fashion. The basic physics is same. However, the HB phase and AGB phase may differ to some extent. The fate of the star after the completion of the red giant phase could be a white-dwarf, neutron star, or black hole depending on its mass.

Let us categorize the stars into three groups based on their Mass: small, medium, and massive, and summarize their process of post-main sequence evolution toward the red giant stage. We shall also see some of their properties in the following sub-sections. Table 18.1 lists the elements involved/produced through thermo-nuclear burning phases inside stars of different masses.

Table 18.1: Thermo-nuclear burning inside a star based on main sequence masses.

Mass of stars in main sequence	Phases of thermo-nuclear burning inside stars
$<< M_\Theta$	Hydrogen
$< M_\Theta$	Hydrogen and possibly helium
$= M_\Theta$	Hydrogen and helium
$> M_\Theta$	Hydrogen, Helium, Carbon and possibly Oxygen, Neon, and Silicon
$>> M_\Theta$	Hydrogen, Helium, Carbon, Oxygen, Neon, and Silicon

18.3.1 Small Stars

In stars with a mass less than 2.3 M_Θ, after the hydrogen gets exhausted in the shell, the star starts contracting due to gravitational pull thereby increasing the temperature of the shell of hydrogen surrounding the core. As the temperature rises the fusion process starts in the shell causing the star to expand. The surface of the star cools and emits in the red region. The core keeps contracting and its temperature rises further, which ignites the helium within the core. This increases the temperature of core thereby accelerating the helium-burning process, which is known as helium flash. The reason is that the core of the star contains a degenerate electron gas, which fails to increase the pressure as the increase in

temperature. Thus, the core cannot cool. As more and more helium fuses, the more the temperature rises in the core causing more helium to ignite. Within a few hours, the temperature shoots up to tremendous amount resulting in the generation of luminosity of 100 billion stars. But even this energy is not sufficient to blow the star. The outer shell of hydrogen surrounding the helium also burns. The fusion of helium leads to the production of carbon. About 5% to 10% of the lifetime of the star is spent in burning the helium and hydrogen out of the main sequence.

After several 100,000 years, the low mass stars cease the helium-burning process and expel the hydrogen rich envelope resulting to a birth of nebula. These are also called *planetary nebula* since when viewed through telescopes their size appears to resemble the size of planets like Jupiter. Some of the expelled matter is chemically evolved, particularly carbon. Remnants of the stars contract and get heated up to 100,000 K, which radiate ultraviolet photons that are absorbed by the surrounding gaseous shell.

18.3.2 Medium Mass Stars

Medium mass stars are those which have masses ranging roughly between 2.3 M_Θ and 9 M_Θ. In medium mass stars, with a mass less than 4 M_Θ as they undergo the red giant phases, their inner cores reach a temperature of around 200 million K when the helium atoms gets fused into carbon atoms and other helium elements. These new nuclear fusions generate a large amount of energy thereby expelling the outer leftover hydrogen layers and heat and illuminate them. This makes the star appear in beautiful, intricate shapes called *planetary nebulae*. After the helium in the core is exhausted, the core shrinks once again due to gravity and becomes very small and extremely hot. The star is then called a *white-dwarf*. The mass of a white-dwarf cannot exceed the Chandrasekhar limit of 1.4 M_Θ. The illumination fades with time as it becomes cooler and finally becomes invisible. It remains forever as a small, very dense cinder called a *black-dwarf*. If the Earth is reduced to within a diameter of few hundred miles, the resulting density would be comparable to that of matter in a black-dwarf. The materials in the nebula keep expanding with time and diffuse into the interstellar space. These gases enriched with important elements, especially carbon and nitrogen, form new clouds, which eventually get recycled when new stars are formed. The most massive stars among the intermediate mass stars also produce some iron. Most of the carbon and nitrogen in the Universe comes from planetary nebulae.

18.3.3 Massive Stars

Stars with masses 10 M_Θ to 100 M_Θ or above are considered massive stars. After they become red giants, they burn helium into carbon like in intermediate stars,

and subsequently burn carbon to form oxygen, nitrogen, and other heavy elements down to iron. Since iron cannot be fused to produce still heavier elements since the process absorbs energy instead of releasing, the nuclear burning comes to an end. Thus, there is no other force left out to counteract gravitational pull and the enormous mass collapses toward the core. The nucleus collapses, the intermediate layers burn out instantaneously, high heat is generated in a short period, and a catastrophic explosion called a *supernova* takes place. Some supernova remnants of stars are listed in Table. 18.2. The gases in the supernova attain temperatures in billions of K. The explosion also produces very rare heavy materials like gold, lead, uranium, etc. The star becomes exceptionally bright and is visible for days or weeks. The ejected materials from the outer layers of the star continue to travel and expand injecting the chemical elements (which it produced in its lifetime of few million years) into the interstellar medium. New clouds are formed from which new stars and planets may be born.

Table 18.2: Some known remnants from a supernova.

Remnant's Name	Age (years)	Distance (light years)	Diameter (light years)
Cassiopeia A	300	10,000	15
Kepler's SN	370	20,000	13
Brahe's SN	400	9,800	20
Crab Nebula	925	6,500	8
Lupus	975	4,000	35
Puppis A	4,000	7,200	55
Vela X	10,000	1,600	130
Cygnus Loop	20,000	2,500	120
IC 443	60,000	5,000	70

EXERCISES

1. What happens to a star immediately after its main sequence life comes to an end?
2. Discuss the Red Giant Phase of the stars.

3. Explain the Red Giant Branch with an illustration.
4. Show the reaction in a triple alpha process and explain with a diagram.
5. Considering the asymptotic giant branch, explain how the helium shell burning starts?
6. What do you mean by 'thermal pulses'? Define the 's-process' and 'r-process' in this regard.
7. "Mass loss is one of the key characteristics of AGB stars"—Explain.
8. "All stars pass through the red giant phase"—Justify.
9. What is helium flash? Make a comparative study of massive stars, medium mass stars, and small stars.

Chapter 19

Brilliant Phenomena: Black Holes, Pulsars, Quasars, Dark Matters, Galaxies, and Clusters

19.1 INTRODUCTION

The Universe consists of mysterious objects like white-dwarfs, neutron stars, black holes, pulsars, quasars, dark matter, and the galaxies themselves. Many of the rare phenomena occur such as the supernova when a star explodes. Much of the existing knowledge makes us believe that the stars are the major driving force in the nature and the activities of the Universe. The majority of the visible Universe is due to stars. Stars are the factories which generate ingredients from basic hydrogen and helium. The stars also support life in the Universe like our Sun, which supports life on Earth. Billions of stars form galaxies, clusters, etc. The remnants of the stars form white-dwarfs, pulsars, quasars and black holes, etc., which play important roles in the Universe. But a large portion of the Universe also consists of invisible *dark matter*, which has not been clearly understood until today. In this chapter we will begin with the discussion on the objects produced out of the remnants of the stars and then study galaxies and dark matter.

See all of the color images in this chapter on the CD-ROM.

19.2 SUPERNOVA

One of the most brilliant phenomena observed in the sky is the *supernova*, which is an explosion producing brilliant light visible for weeks in the sky. It may be thought as astrophysical fireworks. Supernova is broadly classified into two categories: Type I and Type II. Type I occurs in a binary system of stars where one is a red giant and the other is a white-dwarf.[16] The latter accretes hydrogen and helium from the former, generates heat by gravitational compression, and starts a rapid fusion resulting in an explosion. Type II supernovas occur when a red giant or red super-giant explodes. The supernova types are further sub-categorized based on the elements found in the explosion through spectral observation. In general, it has been observed that in a Type-I supernovas the hydrogen Balmer lines are absent, whereas in a Type-II supernovas they are present. These are further sub-classified based on the presence of other spectral lines due to different chemical constituents that appear in the spectra.

19.2.1 Type-I Supernova (Nova)

The Type-I supernova is also known as a nova, which is a nuclear explosion on the surface of a white-dwarf caused by accretion of hydrogen. This happens when a white-dwarf has a nearby binary companion star, which overflows its Roche lobe.[17] The gases from the latter's outer atmosphere will constantly accrete around the white-dwarf. The companion could be a main sequence star or an old star expanding into a red giant. The gases consisting of mainly hydrogen and helium get compressed as they reach near the surface of the white-dwarf due to the latter's intense gravity and get heated to high temperatures. As more and more materials are drawn in, the temperature increases steadily. The interior of a white-dwarf consists of degenerate matter, which is not affected by heat. But the accreted hydrogen is sensitive to heat. Thus, after the temperature rises sufficiently, a stable hydrogen fusion can start on the surface of the white-dwarf

[16] We shall read about white-dwarfs later in this chapter, but for the time being assume that a white-dwarf is an extremely small but dense hot ball whose gravity can pull the gases from nearby and develop a highly pressurized atmosphere around itself. It is the remnant of a star like our Sun after all its fuel has been consumed.

[17] The region of space around a star in a binary system within which orbiting material is gravitationally bound to that star. If the star expands past its Roche lobe, then the material outside of the lobe will fall into the other star.

through a CNO (carbon-nitrogen-oxygen) cycle. When the temperature near the surface of the white-dwarf reaches about 20×10^6 K (due to compression of gases), the conversion of hydrogen into heavier elements gets highly accelerated and an enormous amount of energy is generated within a short period of time, which blows the remaining gases away from the white-dwarf's surface and produces an extremely bright outburst of light. The brightness intensity of the nova reaches its peak and then falls steadily. This rise could be very rapid and can be of the order of $50 \times 10^3 L_\Theta$ to $100 \times 10^3 L_\Theta$. The novas are classified based on the time taken by its peak intensity to fall by 2 or 3 magnitudes. Typically, a fast nova would decay by two magnitudes in 25 days, whereas a slow nova would take more than 80 days. The amount of material ejected by a nova in space is less than the order of $10^{-4} M_\Theta$. The power for the outburst is generated from the fusion of 5% of the accreted matter. Through spectroscopic studies, it has been found that the ejected material from the novas, which form nebula, contain elements like helium, carbon, nitrogen, oxygen, neon, and magnesium. But the overall quantity of elements contributed by a nova in the interstellar medium is much less compared to that from the supernovas from red giants or red super-giants.

A white-dwarf can generate multiple novas during its lifetime if it receives sufficient hydrogen from any source. A good example is RS Ophiuchi, which is known to have produced novas in 1898, 1933, 1958, 1967, 1985, and 2006. If the nova is bright enough and close to us it may be seen with naked eye. A good example is Nova Cygni 1975, which appeared on August 29, 1975 in the constellation of Cygnus. It acquired a magnitude of 2. It has been estimated by astronomers that the number of novas per year in our Galaxy would be between 20 and 60 annually. However, the number of novas actually detected per year is much below expectations probably due to their long distance, lower intensity, and inadequate sensitivity of the instruments used for their search. Table 19.1 lists some of the novas detected from 1970 onwards.

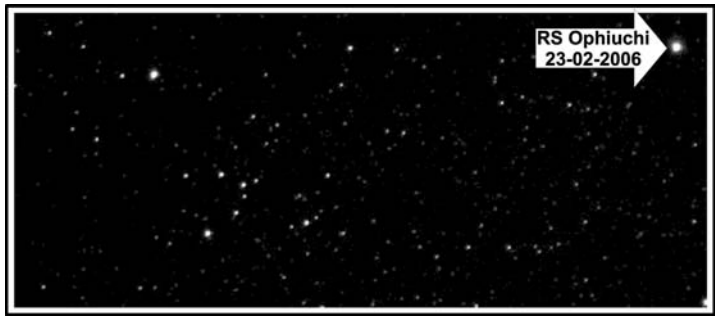


Figure 19.1: A re-occur of the nova on RS Ophiuchi as seen on February 23, 2006 from Mt. Laguna, California.

Table 19.1: Some novas detected from 1970 onwards.

S.N.	Year	Name of the Nova	Peak Magnitude
1.	1970	FH Serpentis	4.0
2.	1975	V1500 Cygni	2.0
3.	1975	V373 Scuti	6.0
4.	1976	NQ Vulpeculae	6.0
5.	1978	V1668 Cygni	6.0
6.	1984	QU Vulpeculae	5.2
7.	1986	V842 Centauri	4.6
8.	1991	V838 Herculis	5.0
9.	1992	V1974 Cygni	4.2
10.	1999	V1494 Aquilae	5.03
11.	1999	V382 Velorum	2.6
12.	2006	RS Ophiuchi	4.5

19.2.2 Type-II Supernova

The Type-II supernova is sometimes called simply supernova, which is one of the most energetic explosive events that results from the explosion of a star after its main sequence life is over (when its nuclear fuel is exhausted), usually in the red super-giant phase while moving in the AGB (asymptotic giant branch; see Chapter 18). This occurs only if the star has mass greater than 5 M_Θ so that when its core collapses, huge amount of gravitational energy is converted to heat. This in turn generates a powerful blast or shock wave which ejects the star's envelope into the interstellar space. It produces an extremely bright object made of plasma that remains visible over weeks or months. It outshines its entire host galaxy. The amount of energy release is terrific. Possibly, it would take 10 billion years for our Sun to deliver the same energy output that an ordinary Type-II supernova would produce. After the explosion, the remnant of the star could be a non-rotating neutron star, generally a *pulsar*, or could turn out to be a black hole. Fig. 19.2 shows an x-ray image of the remnant of Kepler's Supernova, 1604.

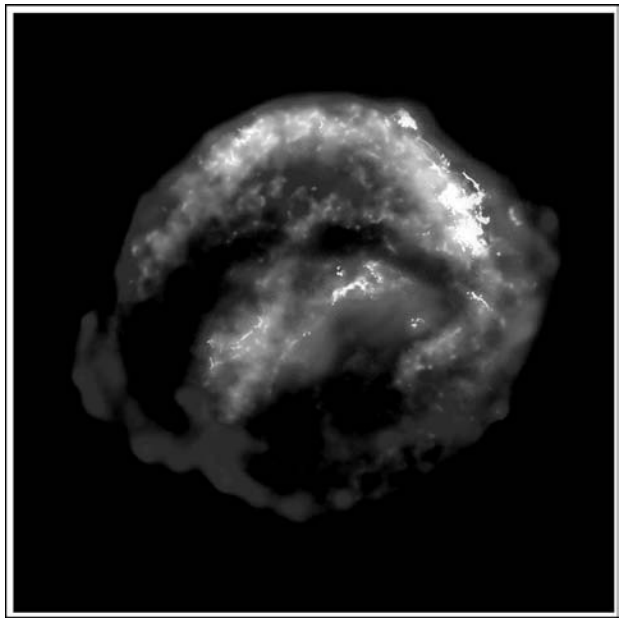


Figure 19.2: Multi-wavelength x-ray image of the remnant of Kepler's Supernova, 1604. (Chandra x-ray Observatory.)

See the color photos on the CD-ROM.

19.3 COMPACT OBJECTS OF THE SKY

White-dwarfs, neutron stars, and black holes come under the category of compact objects in the sky, which are formed from stars after they have passed through the red giant phase. The white-dwarf is a hot ball that remains when a star uses all its nuclear fuel. The mass of the white-dwarf is close to that of the Sun and in any case can't exceed the Chandrasekhar limit of 1.4 M_{Θ}. It is a compressed object containing high density material. However, a neutron star is the more compact with its density about 10^9 of that of a white-dwarf, and a much smaller volume radius of around 12 km. The neutron star could be a rotating or non-rotating type. If the mass of the compact object exceeds more than 2.5 M_{Θ}, invariably a gravitational collapse takes place and the star ends up as a black hole.

19.3.1 White-Dwarfs

After the triple-alpha process in red giant stars is complete for those which evolved with a mass less than 4 M_{Θ}, not enough energy is left to start the carbon fusion

process. The cores of these stars finally contain mainly carbon and oxygen. The core shall gradually radiate away its energy and cool down. No energy is left to hold the gaseous structure against the gravitational collapse. The star thus becomes an extremely dense object (typically 10^9 kg/m^3) in the sky with a typical mass of that of the Sun contained within a volume about equal to that of the Earth, which is called a *white-dwarf*. Typically, one teaspoon from a white-dwarf would weigh 5 tons. The higher the mass of the white-dwarf the smaller its size or more densely packed. In the process of the collapse the star gets heated up and glows brilliantly. The only support against further compression is the electron degeneracy pressure. The mass of a white-dwarf cannot exceed the Chandrasekhar limit of 1.4 M_Θ. It would take hundreds of billion years before the white-dwarfs would cool to temperatures at which they would no longer be visible. Our old Sun is expected to follow the same pattern to becoming a white-dwarf.

Since white-dwarfs are extremely small they are very hard to detect. However, in a certain case of the binary system of star, the detection could be eased. There are other methods also to locate white-dwarfs. The first white-dwarf discovered was Sirius B (Fig. 19.1). In 1844 Friedrich Bessel noticed that the original Sirius (Sirius A) had a small back and forth motion, which is more likely if being orbited by an

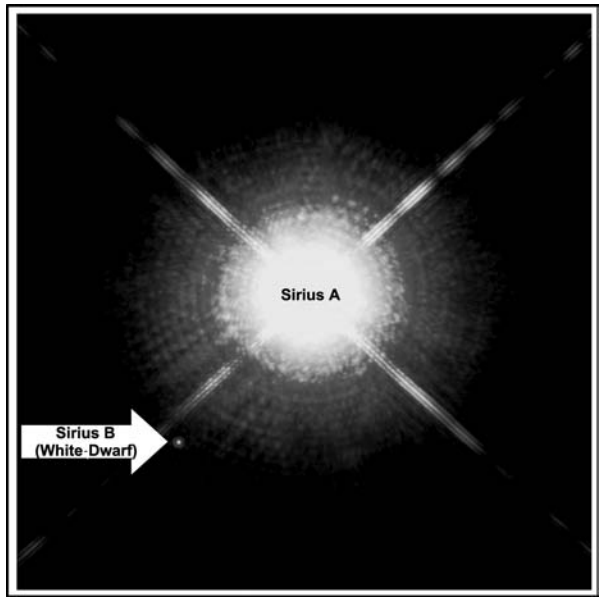


Figure 19.3: An example of a white-dwarf (picture taken from Hubble space telescope, NASA). The Sirius A and Sirius B form a binary system.

See the color photo on the CD-ROM.

unseen gravitating object. In 1863, Alvan Clark an optician, would resolve this mysterious object, which was a very small illuminating object compared to Sirius A, which was none other than a white-dwarf. Thus, it was named as Sirius B. The orbital period of this binary system is about 50 years. The Hubble space telescope has detected more than 75 faint (due to the long distance) white-dwarfs in the globular cluster M4 situated in the constellation Scorpius. Another white-dwarf, HZ 43, was observed using the x-ray satellite ROSAT. Many of the planetary nebulae that can be seen using a moderate-sized telescope contain white-dwarfs near the center. Some of the bright dwarfs are listed in Table 19.2.

Table 19.2: Some of the brightest white-dwarfs.

S.N.	Name of the star	Right Asc.	Dec.	Mag.	Distance (light yr)	Diameter (km)
1.	EG180	04° 31.2′	+59.0°	12.4	18	16,000
2.	Van Maanen's star	00° 49.2′	+05.4	12.4	14	17,000
3.	EG15	02° 08.8′	+25.2°	13.2	100	18,000
4.	40 ERI B	04° 15.4′	-07.7°	9.5	15.7	17,000
5.	Procyon B	07° 39.3′	+05.2°	10.7	11.4	17,500
6.	EG368	16° 48.4′	+59.1°	12.2	40	17,000
7.	W1346	20° 34.4′	+25.1°	11.5	45	17,000
8.	Sirius B	06° 45.1′	-16.7°	8.3	8.6	10,300
9.	W1346	20° 34.4′	+25.1°	11.5	45	17,000
10.	He3 (EG50)	06° 47.6′	+37.5°	12.0	60	13,000

19.3.2 Neutron Stars

After the discovery of the neutron in 1932 by James Chadwick, in 1933 Walter Baade and Fritz Zwicky proposed the existence of stars composed of neutrons. They proposed that the neutron star is formed in a supernova as the mass in bulk is annihilated. The discovery came in 1967 when Jocelyn Bell and Antony Hewish discovered the pulsar, which was later interpreted as a rotating neutron star.

A neutron star is formed from the collapsed remnant of a massive star after a Type II or some special category of Type-I supernova. The matter is extremely

dense. The electrons combine with the protons to form neutrons under influence of terrific gravitational pressure due to the mass of the star. Typically, the mass of a neutron star lies between 1.4 M_Θ to 2.1 M_Θ and its radius lies between 20 km to 10 km. In comparison to the size of the Sun, it is 30,000 to 70,000 times smaller. The typical density of neutron stars varies between 8×10^{13} gm/cm^3 to 2×10^{15} gm/cm^3. Since the size of a neutron star reduces drastically from the size it had when it was in the main sequence, its rotation increases by a terrific amount due to high reduction in the moment of inertia provided it was rotating in the main sequence period of its life. Typically, the rotation could vary from 1/1000th of a second to 30 seconds. The escape velocity from a neutron star is roughly half the velocity of light. Neutron stars not only emit like a black body but also have additional ways of electromagnetic emission. One variety of neutron star is the pulsar, which is a rapidly rotating neutron star with an extremely high magnetic field with an axis inclined away from the axis of rotation. Some of these are responsible for sending pulsed electromagnetic radiations to Earth.

The structural understanding of the neutron star is purely based on existing mathematical models, which may undergo revisions in the near future. We really don't know the true interior structure. Fig. 19.4 shows the interior structure of a neutron star based on these models. The atmosphere of the star is about 1 m thick below which lies a layer of crystalline solid iron crust of about 1 km to 2 km thickness. As we proceed below we find nuclei with ever-increasing numbers of neutrons under tremendous pressures, which prevent them from decaying. This is called in the figure, *heavy liquid*. Further deeper arrives a point called neutron

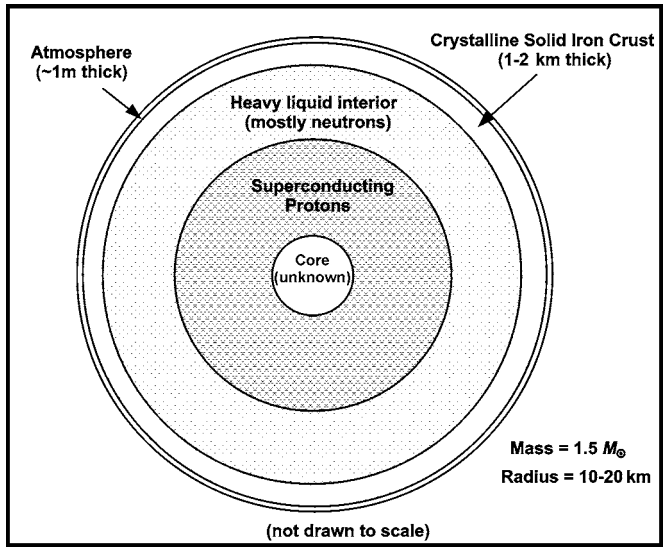


Figure 19.4: A model of the interior of a neutron star of 1.5 solar mass and 10 to 20 km radius.

drip from where onwards there are nuclei, free electrons, and free neutrons until the core is reached. The material in the super-dense core is not well understood. Possibilities are that the core material could be a super-fluid mixture of neutrons with few protons and electrons, or it could contain high-energy particles like pions and kaons in addition to neutrons, or it could even be composed of some strange matter incorporating quarks heavier than its up and down varieties, or it could even be quark matter not bound into hadrons.

19.3.3 Pulsars (A Special Case of Neutron Stars)

As we know, in large stars sudden catastrophic gravitational collapse results in a supernova Type-II explosion. The Crab Nebula is an example of the remnants from such an explosion. After the explosion, all that is left is the core of the original star consisting of mainly neutrons and thus called a *neutron star*. As we have seen, a typical neutron star contains up to 1.5 times of matter as that in the Sun but within a spherical volume defined typically by radius of 10 km. This highly dense matter creates a gravitational pull so powerful that it can bend light. They also exhibit extremely large magnetic fields of the order of 10^{12} times that of Earth's, which is responsible for the electromagnetic wave beacon. Most of the energy escaping from the neutron star by radiation is directed along its magnetic axis as shown in Fig. 19.5. The neutron star may spin physically along an axis, which is likely to be different from the magnetic axis. If the Earth happens to fall in the path of the beams, an electromagnetic pulse is seen each time a beam sweeps over the Earth.

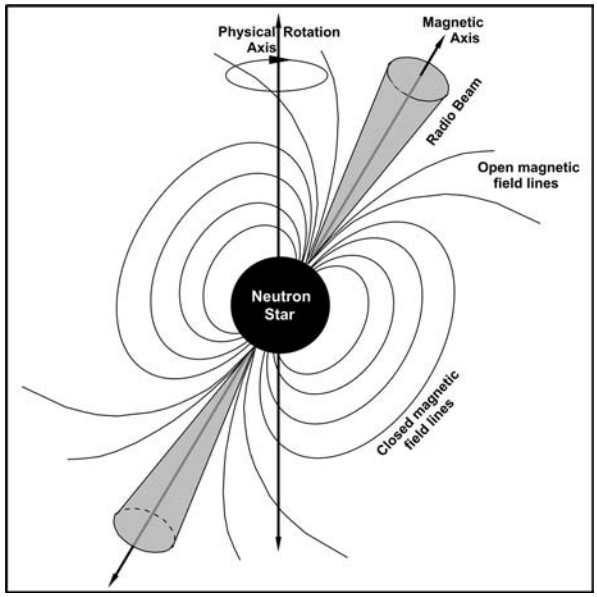


Figure 19.5: The basic structural model of a pulsar.

As shown in Fig. 19.5, the rotation of the strong magnetic field in a neutron star converts it to an electricity generator. The electric field thus generated is strong enough to rip the charged particles such as electrons away from the surface of the star. These particles move along the magnetic field lines and reach the magnetic poles at accelerated velocities, which produce narrow but intense beams of radiation, pointing away from the two magnetic poles. When one of these beams sweeps our Earth, we receive an electromagnetic pulse, very much like we see the light from a flashlight only when it is pointing toward us. This is sometimes called the *lighthouse effect*. Neutron stars whose electromagnetic beams of radiation happen to sweep across us are called pulsars. A good example is the pulsar situated at the center of the Crab Nebula, which rotates 33 times per second. This is shown in Fig. 19.6.

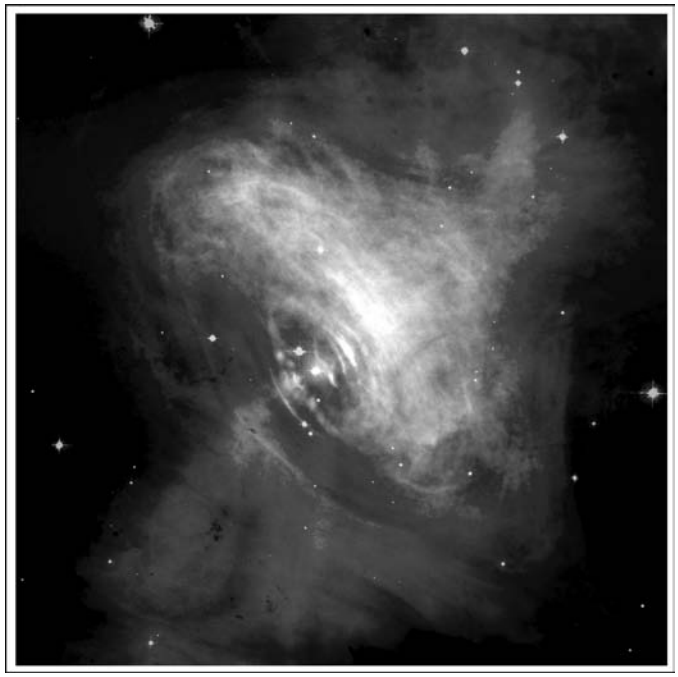


Figure 19.6: This image of Crab Nebula has been prepared with the superposition of x-ray (blue) and optical (red) images. The surrounding nebular gases are being stirred by the pulsar's magnetic field and radiation. (Image created by NASA.)

19.3.4 Black Holes

The black hole is an extremely small-sized theoretical object, which can't be seen but its existence can be felt due to an extremely high gravitational field. The central

region of any galaxy is thought to contain a black hole. It is defined to be a region of space-time from which the escape to the external Universe is impossible. It is also defined by the escape velocity that would have to be attained to escape from the gravitational pull exerted upon an object. The boundary surface of this region is called the *event horizon*, which is an imaginary boundary and nothing can move from inside the event horizon to outside. However, black holes do seem to radiate energy due to quantum mechanical effect known as Hawking radiation. The astronomical observation in x-rays supports the existence of black holes in the Universe. In rotating black holes space-time also rotates and wraps around it. Here, we only discuss the basic properties and structure of these objects from the available theory.

Formation of a Black Hole

The most commonly thought black hole is produced out of dying stars having a mass greater than 20 times that of our Sun. In a main sequence star, there is a constant tug-of-war between gravity pulling in the gaseous matter and pressure pushing them out. The equilibrium between the gravity and pressure is maintained until the star remains stable. The support for this comes from the nuclear furnace of the star. But when the nuclear fuel comes to an end, gravity gains the upper hand and the material in the core is compressed even further. The more massive the core of the star, the more the force of gravity. Thus, the star collapses under its own weight to a very small volume. The future of the star depends on its mass. For small stars like the white-dwarf, after the nuclear fuel is burnt out, the repulsive forces among electrons within the star eventually create enough pressure to halt further gravitational collapse. The star then cools and dies peacefully. But in the case of a very massive star (more than 20 M_{Θ}), after its nuclear fuel has exhausted, it explodes as a supernova. The outer regions of the star are expelled violently into space, whereas the core, which if more massive than 2.5 M_{Θ}, completely collapses under its weight and no known repulsive force within the star can push back hard enough to prevent gravity from completely collapsing the core into a black hole. The core then compacts into a mathematical point with virtually zero volume and is said to posses an infinite density. The escape velocity from a black hole is greater than the speed of light. Since no object, not even light can attain this speed, it is said that nothing can escape from a black hole. The distance from the black hole at which the escape velocity is just equal to the speed of light is called the event horizon. Anything including light, which passes across the event horizon into the black hole, is trapped forever. The theoretical radius collapse of a star with a mass of 20 M_{Θ} as it becomes a black hole is shown in Fig. 19.1. As soon as the radius reduces below 59 km, no radiation can escape from the star.

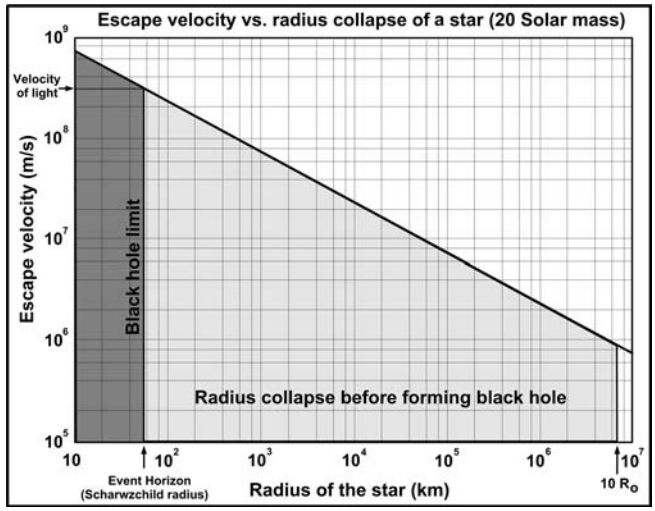


Figure 19.7: Theoretical radius collapse of a star having 20 times the mass of the Sun. As soon as the radius reduces below 59 km, even radiation cannot escape from the star, and the object becomes invisible.

The event horizon, also known as the *Schwarzchild radius*, r_s, is a characteristic radius associated with any non-rotating mass and is given as

$$r_s = \frac{2Gm}{c^2}. \tag{19.1}$$

Here, G is the gravitational constant, m is the mass of the gravitating object, and c is the velocity of light. If m is in kilograms, $r_s = m \times 1.48 \times 10^{-27}$ meters. A non-rotating object, which is smaller than its Schwarzchild radius is a black hole. The Schwarzchild radius of the Sun is only 3 km. Thus, if the radius of the Sun is somehow brought below 3 km it would turn out to be a black hole. The surface at the Schwarzchild acts as an event horizon in a non-rotating body. Fig. 19.8 shows a simulated picture of a black hole.

Classification of Black Holes by Schwarzchild Radius

Depending on the Schwazchild radius, black holes are categorized into mainly three types: super-massive black hole, stellar black hole, and primordial black hole.

Super-massive Black Hole: If $1.5 \times 10^8 M_\Theta$ of matter of normal density (about 1000 kg/m^3) is accumulated, it will fall inside its own Schwarzschild radius, thus forming a super-massive black hole on it own. Super-massive black holes are thought to exist at the centers of most galaxies, including our own

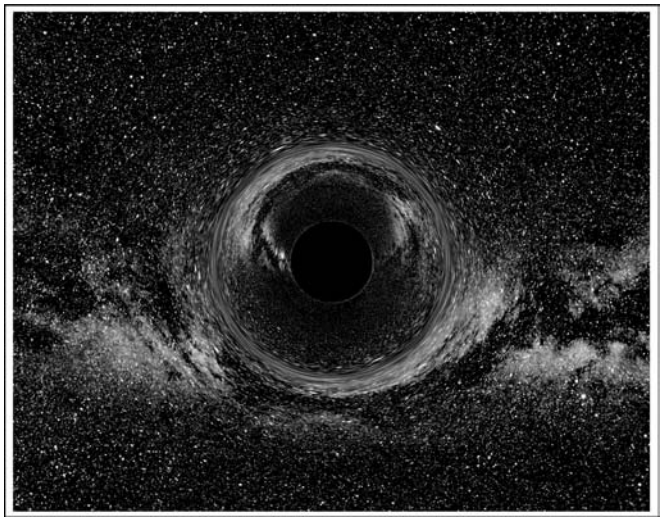


Figure 19.8: A simulated picture of a black hole possessing ten solar masses as would be seen from a distance of 600 km with the Milkyway in the background.

galaxy. Their mass could be equivalent to that of billions of Suns. Our own super-massive black hole is about 2.5 million solar masses. These are thought to have formed gradually and not in one collapse of a cluster of stars. The beginning could be with a stellar-sized black hole, and they grow larger by accretion of matter and other black holes. As evident in the outer parts of galaxies, the distances between the stars are huge, whereas in the central region of the galaxies, the stars are very closely packed. Since everything in the central region of a galaxy is tightly packed, a black hole in the center of a galaxy becomes more and more massive as stars orbiting the event horizon are ultimately captured by gravitational pull and added to the mass of the black hole. The measured velocity of the orbiting stars close to the center of the galaxy suggests the presence of a super-massive black hole. Sometimes there are two jets of hot gases perpendicular to the accretion disk of a super-massive black hole. The length of these jets could be millions of light years. These are probably caused by the interaction of gas particles with strong, rotating magnetic fields surrounding the black hole. The Hubble space telescope provided the best evidence to date that super-massive black holes exist.

Stellar Black Hole: If 3 M_{Θ} of matter of nuclear density (about 10^{18} kg/m^3) is accumulated, it will fall inside its own Schwarzschild radius, thus forming a stellar black hole on its own. This happens in the case of massive stars of more than 20 M_{Θ}. When such a star runs out of nuclear fuel, it can no longer sustain its own weight and begins to collapse. The star heats up due to the collapse and some fraction of its outer layer, which may still contain some

fresh nuclear fuel, and activates the nuclear reaction once again and explodes. This is called a supernova. The remaining innermost portion of the star (i.e., the core) continues to collapse. If the mass of the core exceeds 2.5 M_{Θ}, a stellar black hole is formed for it would end up a singularity with theoretical infinite density.

Primordial Black Hole: When the mass is small, the formations of miniature black holes require extremely high density material so as to collapse within the Schwarzchild radius. The exact mechanisms that result in this type of black hole have not been precisely identified, but a number of hypotheses have been proposed. It is thought that miniature black holes might have been formed shortly after the Big Bang when densities were extremely high. The rapid expansion of some matter might have compressed some slower-moving matter enough to contract into black holes. Some hypothesize that black holes can theoretically *evaporate* and explode. The mass of the black hole determines the time required for evaporation. In the case of massive black holes, the time needed is longer than the current accepted age of the Universe. For a black hole formed at the time of the Big Bang to evaporate today its mass must be about 10^{15} g. During its final phase of the evaporation, such a black hole would explode with a force of several trillion times that of our most powerful nuclear weapon. So far there has been no observational evidence for miniature black holes.

19.3.5 Quasars

Quasars can be thought of as an engineering application of black holes to build powerful ultra wide-band transmitters in the Universe. In other words, black holes are the dynamos of quasars. In 1960, Thomas Matthews and Allan Sandage discovered a faint star-like object of about 16^{th} magnitude, which was named quasar (quasi-stellar). In the Third Cambridge Catalog, this object is recorded as *3C 48* and is usually used for various calibrations of radio telescopes. Some quasars emit extremely broad spectrum starting from radio to x-rays, but many of them do not emit in radio wavelengths. These objects show a high amount of red-shift,[18] which indicates they are extremely distant objects. These are extremely

[18] Redshift is defined as an increase in the wavelength of electromagnetic radiation received by a detector compared with the wavelength emitted by the source, which usually occurs when there is a relative motion between the source and the observer with increasing distance of separation with time like Doppler shift. The opposite of red-shift is blue-shift where the received wavelength is less than the emitted wavelength. The distance between the source object and the observation can be determined from the measured red-shift or blue-shift. The increase or decrease of wavelengths can also occur due to the presence of high gravitational fields, which are called gravitational red-shifts and gravitational blue-shifts.

bright objects producing luminosity of the order of 10^{40} watts. The radiation is usually thought to be generated by a giant black hole devouring its surrounding matter.

The study of quasars has provided some clues to scientists about what happened to the end of Big Bang's re-ionization. The oldest quasars (largest red-shift) show absorption regions in front of them, which suggests that during that time the intergalactic medium was a neutral gas, whereas the younger quasars show no absorption region but on the contrary their spectra contain a spiky area known as the Lyman-alpha forest, which suggests that the intergalactic medium has undergone re-ionization and formed plasma, and that neutral gas exists only in small clouds. Another interesting characteristic of quasars is that they show evidence of elements heavier than helium, which suggests that galaxies underwent a massive phase of star formation between the time of the Big Bang and the first observed quasars.

Mechanism for Emission

Most of the quasar's continuous spectrum is generated by synchrotron emissions,[19] which are high-speed electrons gyrating in some magnetic field. But as we know, electrons gradually lose energy through emission and slow down with time, therefore a continuous supply of high-speed electrons must be maintained to run the show. The best explanation is that quasars are powered by super-massive black holes by consuming mass equivalent to 1000 M_Θ annually. As we have seen with the Type-I supernova, in a binary system high gravity of white-dwarf pulls in the materials from its partner star, and the same mechanism can work with the black hole replacing the white-dwarf. In this case a super-massive black hole is situated in a dense galactic nucleus, which is fuelled by the tidal disruption of the passing-by stars. The stellar material forms a rotating disk-like structure around the event horizon of the black hole perpendicular to its rotational axis known as the accretion disk. A large portion of these materials on their way to entering the black hole are ionized and get trapped in the magnetic field lines around the black hole. These charged particles flow out at a very rapid rate perpendicular to the accretion disk through a phenomenon called *relativistic* jet as shown in Fig. 19.9. Scientists believe that the twisting of magnetic fields in the accretion disk collimates the outflow along the rotation axis of the black hole, so when conditions are suitable, a jet will emerge from each face of the accretion disk. If the jet is oriented along the line of

[19] In the presence of a magnetic field, accelerated electrons near the speed of light move along helical paths and as a result, they emit radio waves. It is called synchrotron radiation since these waves are observed in laboratories when physicists send high energy electrons around in circles using magnetic fields in particle accelerators call synchrotrons.

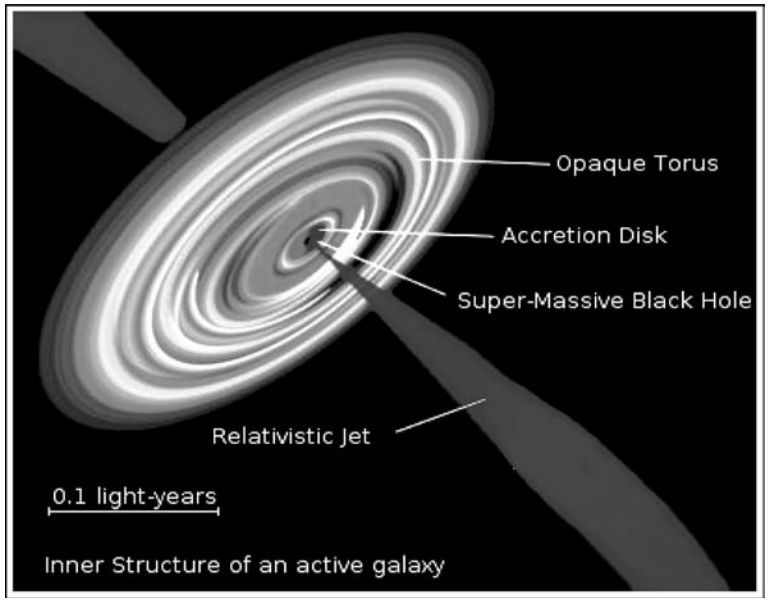


Figure 19.9: The huge disk in the picture represents an active galaxy. The central region contains the super-massive black hole. A small accretion disk is shown. The relativistic jet is shown to propagate out of the center in directions perpendicular to the accretion disk.

sight to Earth, relativistic beaming will change its apparent brightness. Since the calculated speed of these jets is about 99% of the velocity of light, these are called relativistic jets. It is hypothesized that these jets carry as much mass as in Jupiter. An actual picture taken by NASA's Hubble space telescope of a relativistic jet from the quasar situated in the M87 radio galaxy is shown in Fig. 19.10.

19.4 EVIDENCES OF DARK MATTER

Many stars along with planets, gases, and dust live together in a gravitationally-bound system called a galaxy. For example, our solar system along with hundreds of billions of stars and many other forms of matter make the Milkyway galaxy. It has been observed from extremely large distances that even galaxies reside in groups forming a gravitating system called the cluster. The clusters contain the galaxies as well as many other types of material, which are within the intergalactic space. All the matter in the Universe is held together in groups by gravitational forces, which are generated by the matter themselves. The intracluster [20] medium space

[20] Intracluster medium (ICM) is the superheated gas present at the center of any galaxy cluster.

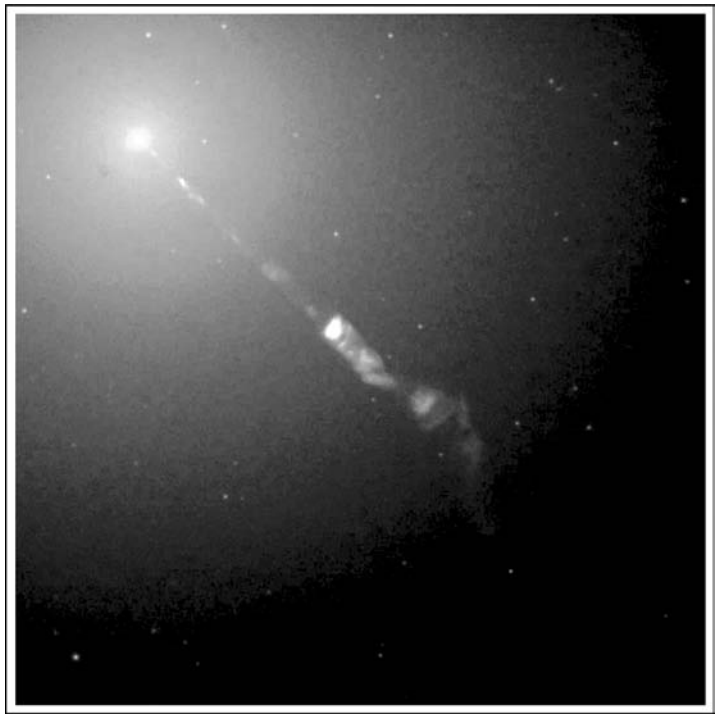


Figure 19.10: Quasar in the M87 radio galaxy emitting a relativistic jet of electrons and other sub-atomic particles traveling at nearly the speed of light from its center. The blue of the jet contrasts with the yellow glow from the combined light of billions of unseen stars and the yellow, point-like globular clusters that make up this galaxy. (Image: NASA Hubble Space Telescope.)

within the clusters is filled with a hot gas that shines in x-rays instead of visible light due to extremely high temperatures of the order of 10^7 K to 10^8 K. From the temperature distribution, it is possible to estimate how much the gases are being squeezed by the force of gravity from all the material within the cluster. From these studies the quantity of matter in that part of space can be estimated.

The results are quite astonishing. It turns out that there are five times more material in clusters of galaxies than we would expect from the galaxies and hot gas we can see. The reason is quite logical. All matter does not emit light and so we can't see them. But their presence is felt from the indirect measurement. This invisible matter contributing five times more gravity is called *dark matter*. Evidence for the existence of dark matter is also found in the galaxies as their dynamics of rotation does not tally with the visible mass. We will learn about this in the following sections.

19.5 GALAXIES

A galaxy is a collection of huge matter floating in space, an enormous gravitationally-bound system consisting of planets, stars, interstellar gas, and dust and possibly unseen dark matter. Our solar system is a part of the Milkyway galaxy (also called the Galaxy). A galaxy typically contains 10^7 to 10^{12} stars apart from other objects orbiting around the center of gravity of the galaxy. Additionally, most galaxies contain a large number of multiple star systems, cluster of stars, and various types of nebula. The spread (diameter) of the galaxies varies from several thousand to several hundred thousand light years. The distances of separation between galaxies are of the order of millions of light years. The space between galaxies is known as *intergalactic space* is usually filled with tenuous plasma having an average density of less than one atom per cubic meter. It is estimated that there could be 10^{11} galaxies in the observable Universe. The dynamics of the galaxies has been a mystery, which indirectly support the theory of existence of dark matter. There are different physical structures of the galaxies and are broadly categorized into three elliptical galaxy, spiral galaxy, and irregular galaxy. Fig. 19.11 shows a spiral galaxy photographed by NASA's Hubble Space Telescope. The elliptical and spiral galaxies are further classified by the specific properties of the individual galaxy: the degree of the ellipse, and number of spirals or definition of bar. This system of galaxy classification is known as the Hubble sequence and the diagram is known as the Hubble *tuning fork* diagram as shown in Fig. 19.12. It was developed by Edwin Hubble in 1936.

Figure 19.11: A majestic spiral galaxy NGC 4414 imaged by the Hubble Space Telescope in 1995. It is situated in the constellation Coma Berenices, having a diameter of nearly 56,000 light years, and is roughly at a 60 million light years distance. (Image created by NASA.)

19.5.1 Classification of Galaxies Based on Hubble Sequence

As shown in the tuning fork diagram in Fig. 19.12, this classification of galaxies was developed by Edwin Hubble in 1936. It begins from the left with elliptical galaxies, which form the base. Note that the tuning fork sequence has nothing to do with the evolution of galaxies. But there are reasons to believe that elliptical galaxies, in general, are older than spiral galaxies. The letter E stands for *elliptical*, while the number attached to it shows the eccentricity of the ellipse multiplied by 10. Elliptical galaxies are named from E0 to E7. E0 has the shape of a ball, which could be a giant globular cluster. E7 is of the shape of a discus.

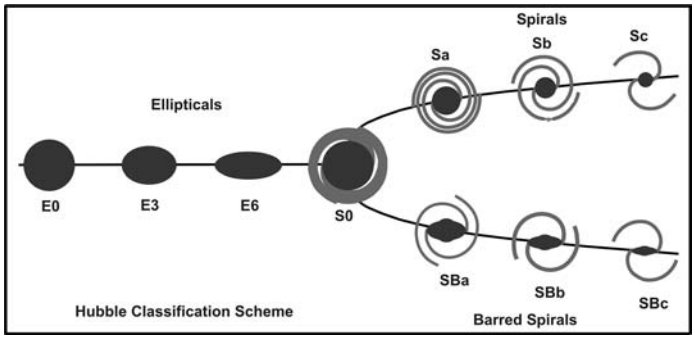


Figure 19.12: The tuning fork diagram (Hubble Classification Scheme).

After the elliptical section the diagram splits into two arms of the tuning fork. The upper arm contains the spiral galaxies, which start from S0 at the junction of the tuning fork. The arms of spiral galaxies seem to emerge directly from the core. The letter S stands for *spiral*, while the number 0 attached to it means no arms. In the other cases we do not have numbers but a lowercase letter that denotes how wound up the arms are. The range of letters is explained below:

(*i*) ***Sa.*** A bright central disk with smooth but tightly wound arms.

(*ii*) ***Sb.*** Slightly better defined spiral arms than Sa.

(*iii*) ***Sc.*** Much looser wound spiral arms than Sb.

(*iv*) ***Sd.*** Very loose arms generating most part of the luminosity.

The lower branch is for barred spiral galaxies. A barred spiral galaxy is defined as a spiral galaxy with a band of bright stars emerging from the center and running across the middle of the galaxy. The spiral arms appear as if having emerged from the ends of the *bar* in these galaxies. The letter B stands for *barred*. The subscript lowercase letters indicate how heavily defined the bar is as given below:

(*i*) ***SBa.*** A bright central region and tightly spaced spirals.

(*ii*) ***SBb.*** Relatively better defined arms than SBa, which are more loosely wound.

(*iii*) ***SBc.*** A much dimmer central portion having even looser arms than SBb.

As was mentioned, our Milkyway Galaxy was earlier believed to be a Sb galaxy like the Andromeda Galaxy, but today it is believed to be a SBb galaxy. Note that Hubble's classification was based on photographs of the galaxies, which made him initially think that elliptical galaxies were an early form of galaxies, which might have later evolved into spiral galaxies. Although today's understanding suggests that the situation is more or less opposite his early belief left its imprint in the astronomical world where people still speak of an *early type* or a *later type* of galaxy based upon whether a galaxy's type appears to the left or to the right in the diagram. We shall describe the galaxies in more detail in the next sections.

19.5.2 General Classification and Properties of Galaxies

Apart from Hubble's classification, which is limited to certain types of galaxies, galaxies as a whole may be classified as follows:

(*a*) ***Elliptical galaxy*:** The range can be E0 to E7. The number indicates only how the galaxies appear in the sky and not its true geometry. E0 galaxies are almost round in shape, while E7 galaxies are greatly flattened. The elliptical galaxies have an ellipsoidal shape with fairly even distribution of stars throughout. The subscript number n relates to the eccentricity of the ellipse having a semi-major axis a and a semi-minor axis b and is evaluated as:

$$n = 10\left(1 - \frac{b}{a}\right). \tag{19.2}$$

(*b*) ***Lenticular galaxy*:** Only the S0 and SB0 fall into this category. These galaxies appears to have disk-type structures with a spherical bulge at the center projecting out of it, and not show any spiral arms.

(*c*) ***Spiral galaxy*:** These galaxies range can be from Sa to Sd. These galaxies have a central bulge and an outlying disk consisting of the spiral arms. The arms are centered about the bulge, and vary from tightly wound (Sa) to very loosely wound (Sc and Sd). Sc and Sd have less pronounced central bulges.

(*d*) ***Barred spiral galaxy*:** These galaxies range from SBa to SBd. These galaxies have a spiral structure similar to the spiral galaxies, but the arms instead of oozing out from the bulge are projected out from the ends of a bar running through the bulge. SBa have the most tightly wound arms. Here again, SBa through SBd refer to how tightly these arms are wound.

(*e*) ***Irregular galaxy*:** These galaxies are symbolized as Irr and classified into two categories: Irr-I and Irr-II. The Irr-I type of galaxies show a spiral structure but are deformed in some way, whereas the Irr-II are those galaxies which do not fit into any other said category.

The known properties of the galaxies are listed in Table 19.3.

Table 19.3: Known properties of the galaxies.

S.N.	Galaxy Type	Mass (M_Θ)	Luminosity (L_Θ)	Diameter (kpc)	Stellar Population	Observed Galaxies (%)
1.	Spiral & Barred Spiral	$10^9 - 10^{11}$	$10^8 - 10^{10}$	5–250	Halo: Population-II Disk: Population-I	77.0
2.	Elliptical	$10^5 - 10^{13}$	$10^5 - 10^{11}$	1–205	Population-II	20.0
3.	Irregular	$10^8 - 10^{10}$	$10^7 - 10^9$	1–10	Population-I	3.0

Today's observations have added information to the types of galaxies. It seems that elliptical galaxies generally contain a fairly low amount of gas and dust. These galaxies are composed mostly of older stars. The spiral galaxies generally have plentiful supplies of gas and dust. These galaxies contain a mix of older and younger stars. On the other hand, the irregular galaxies are fairly rich in gas, dust, and young stars. Based on this, a theory of galaxy evolution has been constructed, which suggests that the elliptical galaxies are formed as a result of collisions between spiral and/or irregular galaxies, which strip out much of the gas and dust and randomize the orbits of the stars. We will discuss more about them in the next sections.

Elliptical Galaxies

The size and mass of these galaxies widely varies. They can be as small as 0.1 kpc to over 100 kpc, while their masses may vary from 10^7 to about 10^{13} M_Θ. The smallest types among these are called dwarf elliptical galaxies, which may not be larger than even a typical globular cluster but may contain considerable mass in the form of dark matter. The motion of the stars within the frame of the galaxy is dominated by randomness. These galaxies contain few young stars, very little interstellar matter, and a few open star clusters. They also contain old stars, the so-called Population-II stars. It has been found that the larger elliptical galaxies have a system of globular clusters, which indicates an old population. It is believed

that in these galaxies the star formation has come to an end. If there is any star formation, the rate seems to be very low. These galaxies appear yellow-red in color, which is in contrast to the distinct blue tinge of typical spiral galaxies (which may be largely from young stars).

As seen from the tuning fork diagram, elliptical galaxies were originally thought to vary from spherical to highly elongate. E0 is considered to be the most spherical. As the number increases from E0 to E7 the corresponding galaxies become thinner and thinner. It is now understood that the majority of elliptical galaxies are moderately thin and the classification by Hubble is a result of the angle of observation of the galaxy. The elliptical galaxies have two distinct physical structures, *boxy* or giant elliptical galaxy and *disky* or normal and low luminosity elliptical galaxies. The boxy structure is a result of the random motion of the stars within the galaxy, which could be large in a particular direction, whereas in the disky structured elliptical galaxies, the random motions are nearly isotropic but appear flattened due to rotation. The dwarf elliptical galaxies are probably not true ellipticals and many astronomers refer to them as dwarf spheroidal galaxies, a debatable issue. Elliptical galaxies and the bulges of disk-type galaxies have similar properties, and are generally regarded as the same physical phenomenon. Fig. 19.13 shows an image of an elliptical galaxy from NASA.

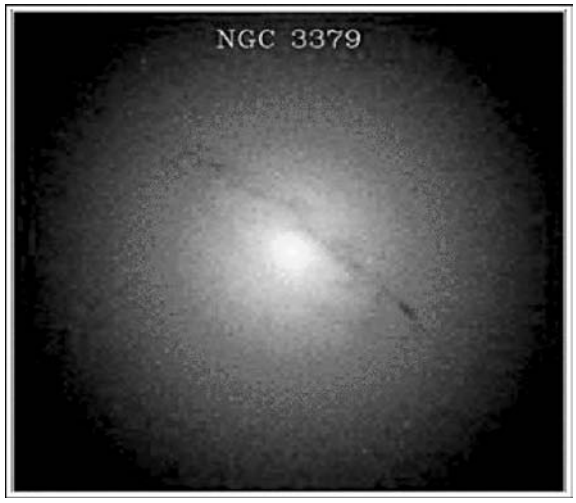


Figure 19.13: An image of an elliptical galaxy, Messier 105 (also known as NGC 3379), in the constellation of Leo, from NASA.

The elliptical galaxies are generally found in the cores of galaxy clusters and in compact groups of galaxies. Recent observations show blue star clusters within some elliptical galaxies together with other structures. These evidences have made scientists today think that an elliptical galaxy is formed when two galaxies

of any type possessing comparable masses collide or merge with each other. This phenomenon is called as *galaxy mergers*. It is also thought that many such galaxy mergers have taken place in earlier times and are still occurring today but infrequently. Minor galaxy mergers involve two galaxies of different masses and are not limited to the giant ellipticals. Even our Milkyway galaxy is known to be digesting a couple of small galaxies currently. Examples of elliptical galaxies are M32, M49, M59, M60, M87, M89, M105, and M110.

Spiral Galaxies

These galaxies consist of a dense central part (a bulge) attached to arms, which rotate around its center. The galaxy appears like a spiral from a distance. The arms form a disk-like structure with a bulge at its center. Fig. 19.11 shows a majestic spiral galaxy, NGC 4414, imaged by the Hubble Space Telescope in 1995. The bulge resembles an elliptical galaxy, containing many old stars (Population-II stars), and possibly a super-massive black hole at its center. The disk is a flat rotating assembly consisting of interstellar matter, young stars (Population-I stars), and open star clusters.

In 1964 the first acceptable theory came in from C. C. Lin and Frank Shu. Their suggestion was that the spiral arms were manifestations of spiral density waves. They presumed that stars travel in slightly elliptic orbits and the orientations of their orbits are correlated. In other words, the ellipses vary in their orientation with respect to one another in a smooth way with increasing distance from the galactic center. This is shown in Fig. 19.14. It can be understood that the elliptical orbits come close together in certain areas to give the look of the arms. Thus, the stars do not remain in a fixed position, which we now see, but they all pass through the arms as they travel in their own orbits around the galactic center. There are also alternative theories based on waves of new star formations.

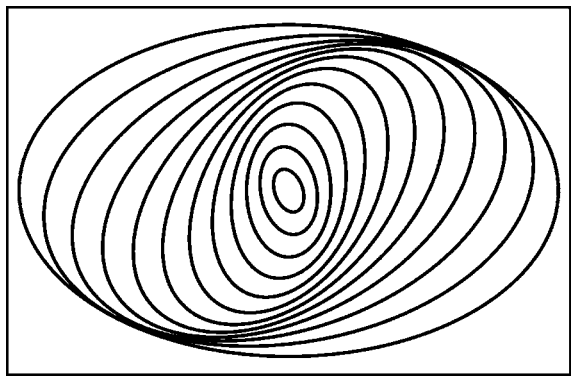


Figure 19.14: Different elliptical orbits of stars around the galactic center are shown, which makes the galaxy look like a spiral from a distance.

The new stars are formed in their bright arms within the disk, which extends in a logarithmic fashion from the bulge. The disks of these galaxies are generally surrounded by large spheroid halos of Population-II stars of which many are concentrated in the globular cluster (orbiting the galactic center). For a long time our own Milkyway galaxy was thought to be a pure spiral galaxy, but recent research from the Spitzer Space Telescope suggests that it is a barred spiral. Examples of spiral galaxies are Triangulum, Whirlpool, Milkyway, Andromeda, and Sunflower galaxies.

Irregular Galaxies

The irregular galaxies are those which do not fall into the Hubble classification for galaxies. In other words, these are neither spiral nor elliptical and often have a chaotic appearance. Neither a nuclear bulge nor any spiral arm can be traced out from the structure. Collectively, they form about one-fourth of all the galaxies. It is believed that most irregular galaxies were once spiral or elliptical in nature but were deformed by the action of gravitation. Fig. 19.15 shows an image of an irregular galaxy, Messier 82 (also known as NGC 3034 or the Cigar Galaxy), from NASA.

Examples of irregular galaxies are IC 1613, Irregular Galaxy IC 10, Leo A, Messier 82 (The Cigar Galaxy), NGC 1569, NGC 3109, Pegasus Dwarf Irregular Galaxy, Phoenix Dwarf, Sculptor Dwarf Irregular Galaxy, Sextans A, and Wolf-Lundmark-Melotte.

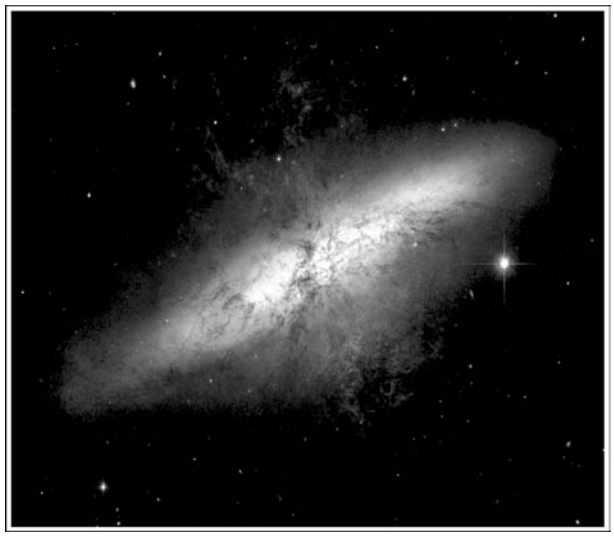


Figure 19.15: An image of an irregular galaxy Messier 82 (also known as NGC 3034 or the Cigar Galaxy) from NASA.

19.5.3 Our Milkyway Galaxy

The Milkyway is the galaxy in which we live. The name Milkyway originated from the appearance of the galaxy as a hazy band of white light on the sky running like a road. This band is a result of the lights from billions of stars existing within the galaxy in its galactic plane. It appears brightest in the direction of Sagittarius, which is toward the galactic center. Fig. 19.16 shows an artist sketch of the Milkyway galaxy. With respect to the celestial equator, the Milkyway passes as far North as the constellation of Cassiopeia and as far South as the constellation of Crux, indicating the high inclination of the Earth's equatorial plane and the plane of the

Figure 19.16: An artist sketch of the Milkyway galaxy as viewed from Earth in a dark night with a clear sky.

ecliptic relative to the galactic plane. The Milkyway divides the night sky into two roughly equal hemispheres, which indicates that our Solar System lies close to the galactic plane. The diameter of the disk of the Milkyway galaxy spans between 80,000 to 100,000 light years. The thickness of the galactic core is about 1000 light years. It is composed of 200 to 400 billion stars. The Galaxy[21] is thought to be a large barred spiral galaxy of Hubble type SBbc (loosely wound barred spiral arms) having a mass of $5.8 \times 10^{11}\, M_{\odot}$. Its integrated absolute visual magnitude has been

[21] Note that the term *Galaxy* beginning with a capital G is generally used as a substitute name for the Milkyway galaxy, whereas the term *galaxy* beginning with a lowercase *g* is used with any galaxy. This convention has been followed in many textbooks and is also adapted by us except in headings where *Galaxy* doesn't necessarily substitute the Milkyway.

estimated to be –21.3. The currently estimated age of the Galaxy based on research performed in 2004 is 13.6×10^9 years, which is nearly the age of the Universe itself. Current estimates tell the distance from the Sun to the galactic center as 26,000 ± 1400 light years.

The central object of the Galaxy is strongly suspected to be a super-massive black hole called *Sagittarius A**, which is a bright and very compact source of radio emission. Like many other galaxies, the distribution of mass in the Galaxy is such that the orbital speed of most of the stars within the Galaxy does not depend strongly on its distance from the center. Away from the central bulge, the typical stellar velocity ranges between 210 to 240 km/s. The orbital period of the stars follows a different law compared to those in our Solar System. The theoretical explanation suggests the presence of dark matter inside the Galaxy.

The bar of the Galaxy is estimated to be about 27,000 light years in length. It primarily consists of red stars, which are believed to be very old. A ring called the *5-kpc ring* surrounds the bar. It contains a large fraction of the molecular hydrogen within the Galaxy. Most of the star formation takes place there. If viewed from an extremely large distance such as from the neighboring Andromeda galaxy, this ring would be the brightest feature of the Milkyway. The galactic center is shown in Fig. 19.17.

Figure 19.17: The galactic center of the Milkyway galaxy along the direction of Sagittarius. The primary stars of Sagittarius are marked in red circles.

It is believed that there are four major spiral arms, all of which begin at the Galaxy's center. Each of the arms describes a logarithmic spiral nature having a pitch of roughly 12°. The arms have been named Perseus-Arm, Norma-Arm,

Scutum-Arm, and Sagittarius Arm. Additionally, there are smaller arms like the Orion-Arm containing the Sun. The Norma-Arm also has several other names: Cygnus-Arm, Outer-Arm, Cygnus-Norma-Arm, and kpc-Arm (only for the inner region). A sketch of the measured and extrapolated arms of the Galaxy is shown in Fig. 19.18. Note the position of our Sun in the Orion-Arm. The Solar System takes nearly 225 to 250 million years to complete one orbit rotation. This amount of time is called a galactic year. The orbital velocity of our Solar System is 217 km/s.

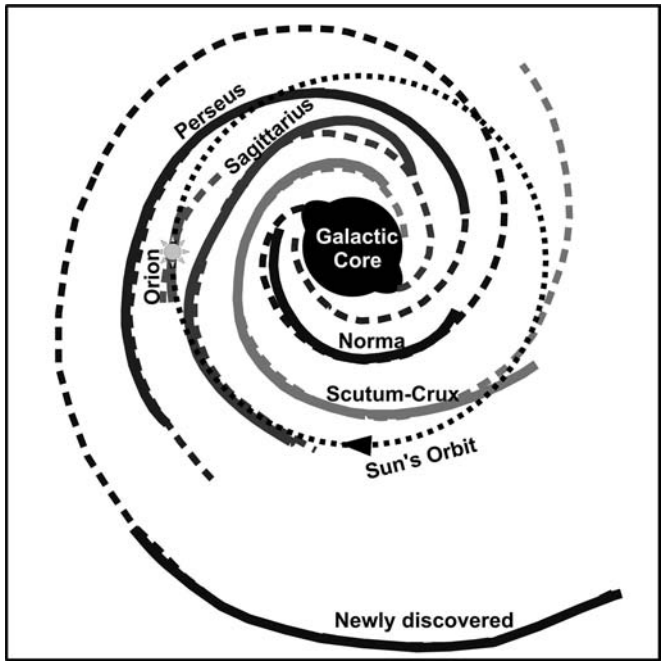


Figure 19.18: A rough map of the observed (solid lines) and extrapolated (dotted lines) structure of the spiral arms of the Milkyway galaxy. The position of the Sun is shown in the Orion-Arm.

Fig. 19.19 shows a picture of the Galaxy provided by NASA, which is an artist's conception of the Milkyway Galaxy as viewed along its axis. As can be imagined by looking at the picture the galactic disk is surrounded by a spheroid halo (this will be clear from Fig. 19.20, which shows the NGC 7331[22] galaxy). The halo consists of the oldest stars and is believed to have formed during the early age of the Galaxy. This halo is also filled with a highly diffused hot ionized gas. The extremely hot gases are thought to produce a halo in the gamma-ray wavelength, too. The halo extends out to hundreds of thousands of light years. However, the

[22] It is thought to have close resemblance with the outlook of our Milkyway galaxy at a very large distance outside the Galaxy. The NGC 7331 galaxy is often referred to as *the Milkyway's twin*.

Figure 19.19: A picture of the Galaxy provided by NASA: an artist's conception of the Milkyway Galaxy as viewed along its axis.

halo is not understood extensively. Extensive studies of the rotation of the Galaxy indicate that the halo dominates the mass of the Galaxy, but the material is not visible, which is called *dark matter*. As can be observed from the figures, the disk of the Galaxy is more or less a flat rotating system that contains the Sun and other young and intermediate aged stars. The formations of stars take place in the disk,

Figure 19.20: The NGC 7331 galaxy, which is often referred to as *the Milkyway's twin*. This is how our Milkyway galaxy might look like to an observer positioned outside our home Galaxy.

especially in the spiral arms. The Sun is nearly 2/3 of the distance from the center to the edge of the disk, which is about 25,000 light years by most modern estimates. The disk also contain atomic hydrogen (H-I regions) and molecular hydrogen (H-II regions) gas and dust.

Note that today's astronomical studies suggest that our neighboring Andromeda galaxy is approaching us at a speed of 300 km/s. This means that the Milkyway may collide with it in three to four billion years of time. However, the collision could possibly destroy some of the elements in both the galaxies, but both of these galaxies will finally merge together to form an elliptical galaxy.

19.5.4 Rotation of Galaxies

The orbital speed of an object going around another object is the speed at which it orbits around the barycenter[23] of the system. In the case of a small mass orbiting a much heavier mass, the barycenter is almost at the center of mass of the massive body. The orbital speed at any position in the orbit can be computed from the distance to the central body at that position. The orbital velocities for an object of mass m going around a much heavier object of mass M ($m << M$) can be expressed as in Eqs. (19.3) and (19.4), respectively, for elliptical orbits and parabolic trajectory (circular orbits), where G is the gravitational constant, a is the semi-major axis of the elliptical orbit, and r is the distance between the individual center of masses of the two objects. Note that in the case of very small eccentricity ($a \approx r$) or in circular orbits $v \propto r^{-1/2}$.

$$v = \sqrt{GM\left(\frac{2}{r} - \frac{1}{a}\right)} \tag{19.3}$$

$$v = \sqrt{\frac{GM}{r}} \tag{19.4}$$

The center of mass of a galaxy more or less resides in the nuclear bulge of the galaxy. All the elements rotate around this center. The stars revolve around the center at a constant speed over a large range of distances from the center of the galaxy. Fig. 19.21 shows the variation of rotation speed (orbital velocity) of the elements, stars, gases, etc., as a function of distance from the center of the galaxy for a general case and the Milkyway. The theoretical curve does not match the observed one. Instead of reduction in speed with increase in radii, the rotational velocity stabilizes and remains roughly constant around 220 km/s. It is also not understood whether the dip in the graph at about 2.9 Kpc is real or just an artifact

[23] The center of mass of a system constituted by distributed masses present int. It is a specific point at which the system's entire mass behaves as if it were concentrated.

introduced different motions in the gases. The flat behavior is also true in the case of many other galaxies. A typical rotational behavior is shown for some other galaxies.

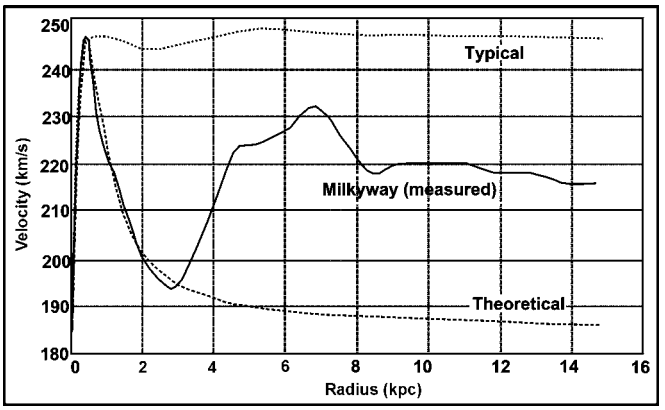


Figure 19.21: The rotational behavior of galaxies with increasing radius.

There is no direct way to measure the mass of any distant object. All we have to do is measure some parameter related to the mass, which can be observed from Earth. Thus, most of the information that reaches us is only from luminous objects like stars and hot gases. As seen in the graph the rotational velocity does not decrease in the expected manner, which makes us guess that there are substantial extra amounts of matter (could be in the form of gases), which we cannot observe directly, whose gravity is responsible for the behavior in the curve. This is some of the strongest evidence for dark matter. There are controversies on the issue of dark matter like the *MOND* theory, *Modification of Newtonian Dynamics*. This theory suggests that the physics of gravity changes on a large scale but, until recently, was not a relativistic theory.

19.6 GALAXY GROUPS, CLUSTERS, AND SUPER CLUSTERS

The early stage of the Universe is believed to have been almost smooth and featureless. With its evolution, with time, it gradually developed itself and became organized. The planets revolve around the Sun and form the Solar System. Similarly, the Sun, along with hundreds of billions of stars, gases, and dust move around the galactic center of our Milkyway galaxy. As we move up in the scale we find that galaxies organize themselves in groups both small and large. The groups include the galaxies and any other material, which is within the intergalactic space. These are considered to be the densest parts when the Universe is studied as a large

scale structure and are the largest gravitationally-bound objects. Depending on the mass, these groups of gravitationally-bound systems can be classified into three groups: galaxy groups, clusters, and super clusters. These are explained below.

(*i*) ***Groups of galaxies.*** These are the smallest aggregates of galaxies and typically contain less than 50 galaxies. The estimated mass is about $10^{13}\, M_{\Theta}$ within a diameter of 1 to 2 Mpc. The group to which our Milkyway belongs is called the *local group* and contains more than 40 galaxies. The spread of velocities for the individual galaxies is approximately 150 km/s. Note that this definition should be used as a guide only, since larger and more massive galaxy systems are sometimes classified as *galaxy groups*.

(*ii*) ***Clusters of galaxies.*** Still larger than the groups are clusters containing 50 to 1000 galaxies, hot x-ray emitting gas, and large amounts of dark matter. The distribution of these three components is roughly the same within the cluster. The estimated mass is between $10^{14}\, M_{\Theta}$ to$10^{15}\, M_{\Theta}$ within a diameter of 2 to 10 Mpc. The spread of velocities for the individual galaxies is approximately 800 to1000 km/s. Fig. 19.22 shows a picture of a massive galaxy cluster, Abell 1689.

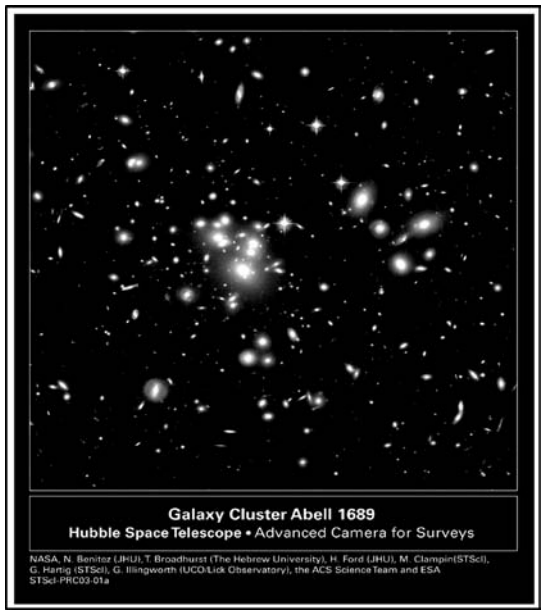


Figure 19.22: One of the most massive galaxy clusters observed is the Abell 1689 in the constellation Virgo. The yellow galaxies belong to the cluster itself, but the red and blue distorted streaks are background galaxies, which are being gravitationally lensed by the cluster. Some of the lensed galaxies are more than 13 billion light years away. The lensing zone itself is 2 million light years across.

(*iii*) ***Super clusters.*** An even larger scale of visualization is called a *super cluster* consisting of groups, clusters, and even some isolated galaxies. It was once thought of as the largest structure in nature, but today it is known be subordinate to the enormous walls or sheets sometimes called *super cluster complexes*. Super clusters range between 10 Mpc to 150 Mpc. Fig. 19.23 shows a map of the nearest super clusters.

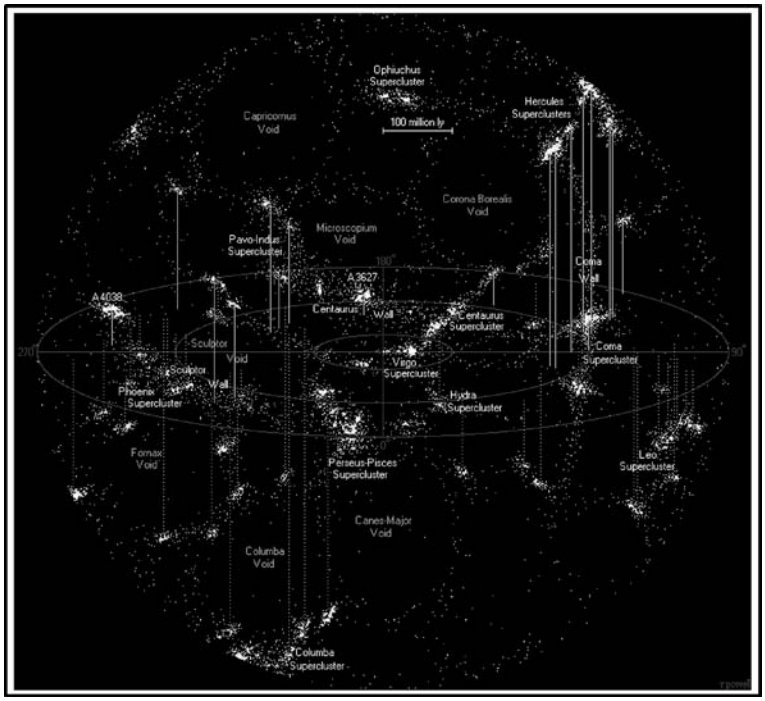


Figure 19.23: A map of the nearest super clusters (near Virgo).

19.7 GRAVITATIONAL LENSING (EINSTEIN EFFECT)

Einstein's general relativity predicts that when a large amount of matter is accumulated in space, a time-space curvature is formed around the object due to its own gravitational forces. For such actions, a substantial amount of mass is required within a small space since the gravitational force is the weak force of nature. When light or any electromagnetic radiation passes within this region, it gets bent due to the time-space curvature. Eienstein showed this phenomena during an eclipse when light from a star behind the Sun was able to reach the Earth during the eclipse. Fig. 19.23 shows the mechanism involved.

Fig. 19.24 shows light from a distant galaxy reaching the Earth after getting bent by the gravity of a massive object. The orange arrows illustrate the apparent position of the background source. The white arrows illustrate the path of the light from its true source location. The path of the light from the source is curved, which distorts the image, and hence, the apparent location of the source could be different from its true location. Additionally, multiple images of a single source might be observed. Note that if the observer, source, and massive object join in a straight line, the image of the source will appear like a ring behind the massive object. This phenomenon, usually referred to as *Einstein ring*, was first mentioned in 1924 by Orest Chwolson and later quantified by Einstein in 1936.

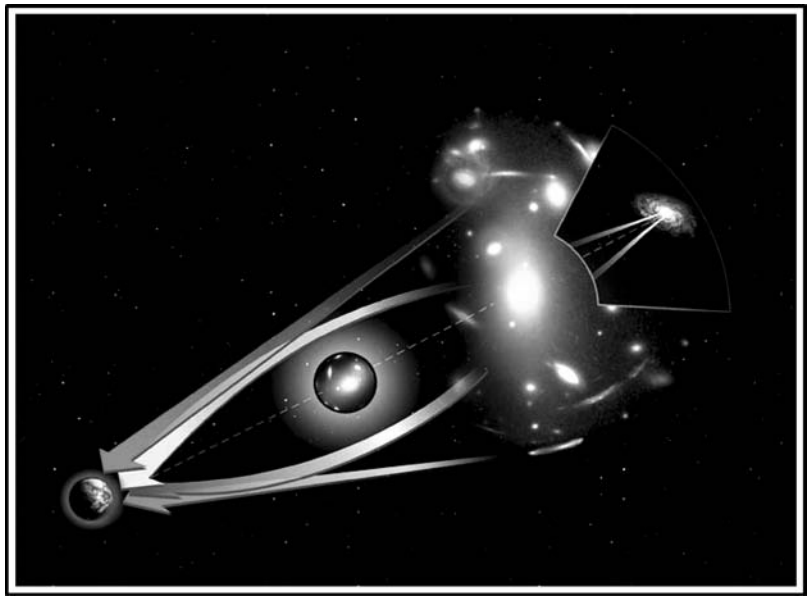


Figure 19.24: Light from a distant galaxy reaching the Earth after getting bent due to a massive object. The orange arrows illustrate the apparent position of the background source. The white arrows illustrate the path of light from its true source location.

The effects of gravitational lensing may be noted in Fig. 19.23 and more clearly in Fig. 19.25. In the latter figure, a cluster of galaxies called Abell 2218 is shown, which acts as a powerful gravitational lens, magnifying all galaxies lying behind its core. The gravitationally lensed galaxies appear to be stretched along the cluster's center and some of them are imaged multiple times. The multiple images usually appear in pairs along with a generally faint third counter image, as is the case for very distant objects.

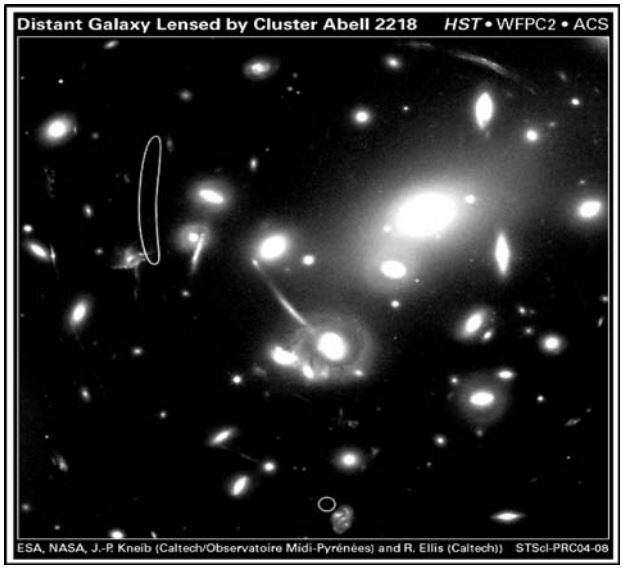


Figure 19.25: Abell 2218, a cluster of galaxies nearly 3 billion light years away in the constellation Draco. It acts as a powerful gravitational lens, magnifying all galaxies lying behind its core. The lensed galaxies appear to be stretched along the cluster's center and some of them are imaged multiple times.

The gravitational lensing may be classified into three categories: strong lensing, weak lensing, and micro lensing depending on the characteristic properties of the image. These are listed in Table 19.4.

Table 19.4: Properties of a gravitational lens.

Lensing Type	Properties
Strong	The distortion of the background object is easily visible, i.e., formation of Einstein rings, arcs, and multiple images.
Weak	The distortions are much smaller and can only be detected by analyzing large numbers of objects. The lensing action shows up statistically as a stretching of the background objects located perpendicular to the direction to the center of the lens.
Micro	The distortions are not visible in shape, but the amount of light received from a background object varies with time. The background source can be star, a quasar, or any distant emitting object within or outside our Galaxy, while the lens can also be star, a quasar, or any massive object coming in between.

19.8 RED-SHIFTS

When the wavelength of visible light increases due to some physical phenomena it is called red-shift. In other words, if the frequency of the light observed is less than that of the transmitting source, it is called a red-shift. This is shown in Fig. 19.26. The opposite of red-shift is blue-shift where the observed frequency is more than the source frequency (wavelength decreases). The terms *red-shift* and *blue-shift* are not restricted to use for visible light alone but are also used for any electromagnetic waves. These may be quantified by the relative difference between the observed and emitted wavelengths (or frequency) of an object. In astronomy this change is measured relatively such that it is dimensionless. If λ and λ_0 are the measured and emitted wavelengths then the red-shift z is given as in Eq. (19.5). The red-shift z can also be expressed in terms of frequency as shown in Eq. (19.6), where ν and ν_0 are the measured and emitted wavelengths. If z is positive it is a red-shift and if z is negative it is a blue-shift.

$$z = \frac{\lambda - \lambda_0}{\lambda_0} = \frac{\lambda}{\lambda_0} - 1 \tag{19.5}$$

$$z = \frac{\nu_0 - \nu}{\nu} = \frac{\nu_0}{\nu} - 1 \tag{19.6}$$

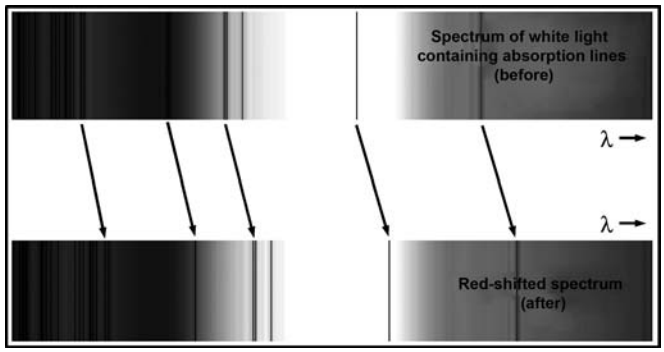


Figure 19.26: The top portion shows a spectrum of light containing absorption lines (dark lines). The bottom is the spectrum obtained after undergoing a change (increase) in wavelength. Note that the lines have moved to the right.

19.8.1 Mechanisms for Red-Shifts and Blue-Shifts

There are various mechanisms by which the source frequency may differ from the observed frequency, which are listed as follows:

(*i*) ***Relativistic Doppler Red-Shift.*** If a source of electromagnetic radiation moving away at a relativistic speed[24] with a relative velocity v with respect to the observer then the frequency observed is less than emitted and is a red-shift as expressed in Eq. (19.7), where γ is the Lorentz factor as expressed in Eq. (19.8) and c is the velocity of light. At non-relativistic speeds γ is equal to unity. Consequently, for a source moving at an angle θ to the observer ($\theta = 0$ is directly away from the observer), the complete form for the relativistic Doppler effect can be expressed as in Eq. (19.9).

$$z = \left(1 + \frac{v}{c}\right)\gamma - 1 \tag{19.7}$$

$$\gamma = \sqrt{1 - \frac{v^2}{c^2}} \tag{19.8}$$

$$z = \left(1 + \frac{v \cos\theta}{c}\right)\gamma - 1 \tag{19.9}$$

(*ii*) ***Cosmological Red-Shift.*** If we consider the Universe to be expanding then the distance between any two points for a particular instant will be different at a different instant. This suggests that the space between these two points is expanding with time. If we take multiple points, the same principle applies and the distances between individual points will increase with time. If the velocity of light were fixed, then due to the continuous expansion with time a light source at one point as seen from any other point will constantly change its frequency with time. The frequency will be reducing with time. In other words, a red-shift will be observed. Since the distance is a function of time in this case, if we know the red-shift then we may also obtain the distance of separation. This type of red-shift is also known as *cosmological red-shift* or *Hubble red-shift*. On the other hand, if the Universe was contracting then the blue-shift would have been observed. The expression for cosmological red-shift is given in Eq. (19.10), where a_{T_1} and a_{T_2} are the values of the cosmic scale factor at times T_1 and T_2, respectively. A red-shift occurs if $T_2 > T_1$ and a blue-shift occurs if $T_2 < T_1$.

$$z = \frac{a_{T_2}}{a_{T_1}} - 1 \tag{19.10}$$

[24] Speed comparable to the velocity of light.

(*iii*) ***Gravitational Red-Shift.*** It has been observed that in the presence of high gravitational fields produced by massive objects, the time-space dilation takes place. It may be thought that the time-space density near the gravitating object is more. Thus, frequency radiated from within a high gravitational field when observed at a reasonable distance where the effect of gravity is relatively weaker, the frequency is found to increase. The reverse is also true. If the source is outside the strong gravity and the observer at a strong gravitational field measures the frequency he would find it less than the actual frequency of radiation. The gravitational red-shift can expressed in Eq. (19.11), where G is the gravitational constant, M is the mass of the gravitating object, and r is the radial coordinate of the observer (which is analogous to the classical distance from the center of the object).

$$z = \frac{1}{\sqrt{1 - \frac{2GM}{rc^2}}} - 1 \tag{19.11}$$

19.9 HUBBLE'S LAW IN RELATION TO THE EXPANDING UNIVERSE

Hubble's law says that the recessional velocity of an astronomical object determined from the red-shift is directly proportional to its distance. As said previously, the Universe is expanding according to the Big Bang theory. But a mere confirmation came in 1929 when Edwin Hubble discovered a rough proportionality of the objects' distances with their red-shifts. This proportionality factor is called Hubble's constant and is usually represented as H_0. However, its value has changed over time and the current value is roughly 70 km/s/Mpc. To put it into a mathematical form, if the recessional velocity is v and D is the distance then Hubble's law expressed as in Eq. (19.12) is

$$v = H_0 D. \tag{19.12}$$

EXERCISES

1. Give names of some mysterious objects of the universe.
2. Briefly explain the objects:
   (*a*) Galaxies.
   (*b*) Dark matter.

3. What is a supernova? What are its two categories and what are their characteristic differences?
4. Discuss the properties of a Type I supernova.
5. Give the names of different novas detected from 1970 onwards.
6. Discuss the properties of a Type II supernova and point out the characteristic differences with Type I.
7. What is the order of energy released by a Type II supernova as compared to that of the Sun?
8. What are compact objects in the sky?
9. Write a few lines about white-dwarfs, neutron stars, and black holes.
10. Write down the important properties of bright white-dwarfs.
11. How is a neutron star formed? What is the value of its typical density?
12. Give a model of the interior of a neutron star.
13. What are pulsars? Write down their important characteristics.
14. Give the basic structural model of a pulsar.
15. From the available theory, write down the basic properties and structures of black holes.
16. How are the black holes formed? Give a clear interpretation of the same.
17. What is the 'Schwarzchild radius'? What would be the value of this for our Sun?
18. Depending on the Schwarzchild radius, categorize the black holes into three types and discuss them.
19. What are quasars? Write down the characteristics of quasars.
20. Explain the mechanism for emissions from quasars.
21. What is dark matter? Give evidence for the existence of dark matter.
22. What is the typical number of stars in a galaxy? What is intergalactic space?
23. Discuss the different physical structures of galaxies. Explain the characteristics of elliptical and spiral galaxies.
24. Classify galaxies based on the Hubble sequence.
25. Make a general classification of galaxies and explain the different types.
26. Write down the known properties of the galaxies.
27. Mention the important characteristics of elliptical galaxies. Compare them with spiral galaxies.
28. What are irregular galaxies? Describe briefly our Milkyway, and then galaxies of different types.
29. What is meant by a rotation of galaxies?
30. Depending on the mass, classify the galaxy groups and explain them.
31. What is meant by a 'red-shift'? Write down its properties.
32. Explain the various mechanisms for red-shifts and blue-shifts.
33. Write down the Hubble's law in relation to the expanding universe.

Chapter 20

Problems of Radio Interference and New Developments in Radio Astronomy

20.1 INTRODUCTION

In this chapter we discuss the problems of radio interference, spectrum monitoring instruments, methods of terrestrial spectrum analysis, and some new developments taking place in the field of radio astronomy. Spectrum management is generally not covered in other texts of radio astronomy. We felt its importance and have treated the subject seriously. Finally, we describe an ongoing international project for futuristic radio astronomy involving several countries called the *Square Kilometer Array*.

20.2 RADIO INTERFERENCES IN ASTRONOMY

Radio astronomy signals are extremely weak in comparison to man-made terrestrial radio signals and radio noise. We take, for instance, the radio flux generated by

our nearest star, the Sun. Its flux is measured in units of SFU (Solar Flux Units), where 1 SFU is equal to 10^{-22} W m^{-2} Hz^{-1}. A distant radio station may produce a signal flux density of the order of several millions of SFUs. The radio power flux densities from distant radio galaxies received on Earth are generally of the order of several micro-jansky to few milli-jansky, where 1 jansky[25] is equal to 10^{-26} W m^{-2} Hz^{-1}. The astronomical sources produce signals of the nature of a random noise (Gaussian noise) generally in both polarizations and spread over a wide band of frequencies. When seen at the output of a radio telescope receiver, the spectrum appears to be more or less flat over a substantial amount of bandwidth (unless absorption lines are present). It appears like a locally terminated spectrum analyzer with a matched load (without connecting any signal source). A locally generated wideband radio noise like that produced from an electrical arc welding may be quite high in magnitude to obscure the radio astronomy signal. Any kind of signal or noise (other than that produced from astronomical sources) interfering or obscuring the astronomical signals is called RFI (Radio Frequency Interference) by radio astronomers.

RFI can enter into the radio telescope data through the side-lobes of the telescope-antenna or as a leakage through the cables/shielding of the electronic receiver system. This is one of the reasons that radio telescopes are located in very remote places where human population is rarely present. Fig. 20.1 shows the various possible ways by which the RFI can directly enter into the telescope

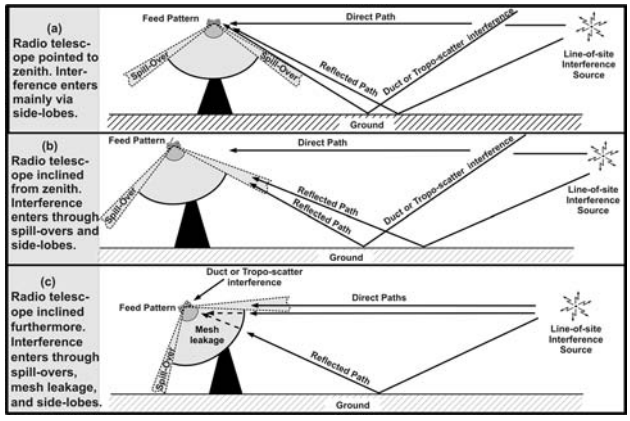


Figure 20.1: Methods by which the RFI gets coupled to the radio astronomy data through the antennas. (a) RFI signals enter through side-lobes when the antenna points to zenith. (b) RFI signals enter through spill-over and side-lobes when the antenna points at a small angle to zenith. (c) RFI signals enter through spill-over, mesh leakage, and side-lobes when the antenna points at a large angle to zenith.

[25] This unit is named after the discoverer of radio emission from extra terrestrial objects by Sir Karl Jansky in 1933.

data through the antenna in low frequency radio telescopes (VHF/UHF). In part (*a*), RFI signals enter through side-lobes when the antenna points to zenith. In part (*b*), RFI signals enter through spill-over and side-lobes when the antenna points at a small angle to zenith. In part (*c*), RFI signals enter through spill-over, mesh leakage, and side-lobes when the antenna points at a large angle to zenith.

RFI can be broadly classified into two categories: narrowband RFI and broadband RFI. Fig 20.2 shows the spectrum of narrow and broadband RFI. Narrowband RFI is generally produced from communication transmitters. In general, the communication signal may be represented in terms of weighted sine and cosine functions of frequency (by applying a Fourier transform on the time domain signal). The spectral power distribution of these signals lie concentrated around some discrete frequencies represented by sine and cosine terms. Thus, the overall spectrum appears as tall narrowband lines (communication signals) standing over a wide uniformly distributed power over the entire frequency band (radio astronomy signals). In order to process the acquired astronomical data, a simple technique generally adopted is to remove the data containing these narrowband lines and process the rest of the data.But this process is not very

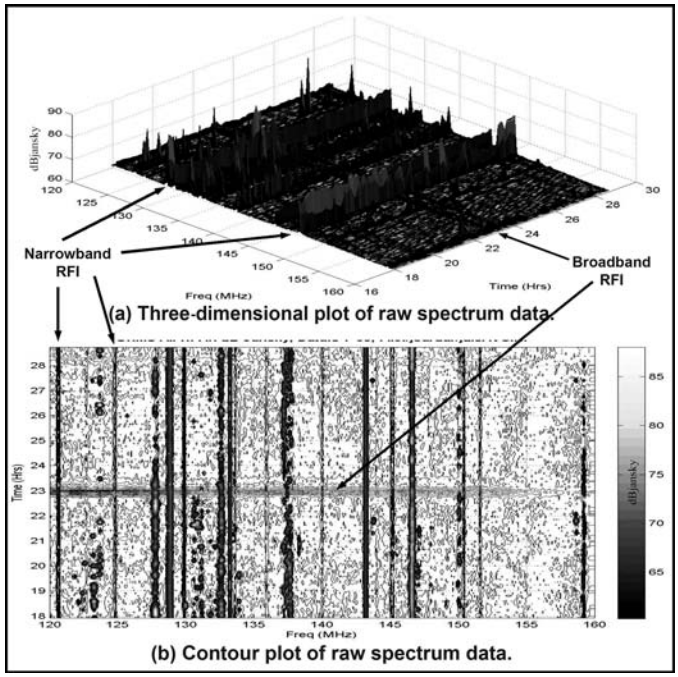


Figure 20.2: Examples of narrow and broadband RFI. The power spectrum is plotted against time and frequency. (a) Three-dimensional plot of raw RFI spectrum. (b) Contour plot of the same data.

efficient if a large portion of the data is wasted. There are other techniques to eliminate these lines, which are called RFI mitigation. The process mathematically separates the lines from the original data. A separate antenna is used to collect the signature of the RFI, which is properly scaled and vectorially (amplitude and phase) subtracted from the original data. Thus, the remaining data is free from RFI. However, not all types of RFI can be eliminated by this process. Moreover, proper signature of the RFI in quality and quantity are needed for this process.

The other type of RFI, as stated earlier, is broadband RFI. These are generated from man-made electrical processes where a large amount of current in electrical circuits is switched. In other words, di/dt is large, where i represents the current and t the time. Therefore, if any amount of inductance is present in the circuit the induced voltage will be very large and the system can radiate. Examples of such processes are electrical arc wielding, electric ignition in motor vehicles, SMPS (Switched Mode Power Supply), etc. The spectral distribution of these signals is extremely large, which may be understood by looking at the Fourier transform of a pulse. Thus, the presence of such signals can radio-jam the telescope and the entire extra-galactic data may get spoiled. This kind of RFI could be more dangerous than the narrowband type since these do not leave any scope for data usage.

20.3 RFI SPECTRUM MONITORING SYSTEMS

Terrestrial radio spectrum monitoring assists radio astronomy by providing information such as free cosmic frequency bands, out-of-band and spurious emissions form transmitters, power line interference, and the growth of new transmitters in the region. Magnitude spectrum of radio signals available from a region is useful for generating radio availability statistics and studying radio properties over time, etc.

The GMRT[26] has a four antenna magnitude spectrum monitoring system. It also provides radio direction information using software. Fig. 20.3 and Fig. 20.4, respectively, show the antenna mountings on a tower and the block details of the omni-directional RFI spectrum monitoring system of GMRT. Four identical log-periodic dipole arrays with their maximum directivity pointing toward East, West, North, and South directions are mounted 20 meters above the ground plane. The antennas are designed to cover the VHF and lower UHF bands. These are positioned in the E-plane (parallel to the ground). The computer selects the antennas in a cyclic order, East, West, North, and South using a 4PST (Four Poles Single Throw) RF-switch. The corresponding RF powers from the antennas are

[26] Giant Meterwave Radio Telescope. It is a low-frequency radio telescope (50 MHz to 1620 MHz) located in Narayangaon, Pune, India.

Figure 20.3: A picture of the antennas of the RFI monitoring system of GMRT. Four log-periodic antennas are mounted on a tower of 20 m height in the E-plane directed along East, West, North, and South directions.

amplified suitably by an LNA (Low Noise Amplifier) and fed to the spectrum analyzer kept inside the electronics laboratory building through an RF-cable. The traces of the spectrum analyzer are re-displayed and saved by the computer with time and directional stampings using its parallel port to GPIB interface. Settings of the spectrum analyzer and observation period are user defined and set using the computer. The computer cycles the antennas over the entire period of observation and keeps recording the data in five different data files.

20.3.1 System Operating Software and Data Visualizing Software

The system is operated using software running on the Linux platform. The settings of the observation are first entered into the GUI[27] fields, i.e., duration of observation, center frequency, frequency span, resolution bandwidth, video bandwidth, number of averages, name of the file to save data, etc. Four files are created for saving individual antenna spectrum data along with a common file

[27] Graphical User Interface of the software through which the operational parameters are set and the results are displayed.

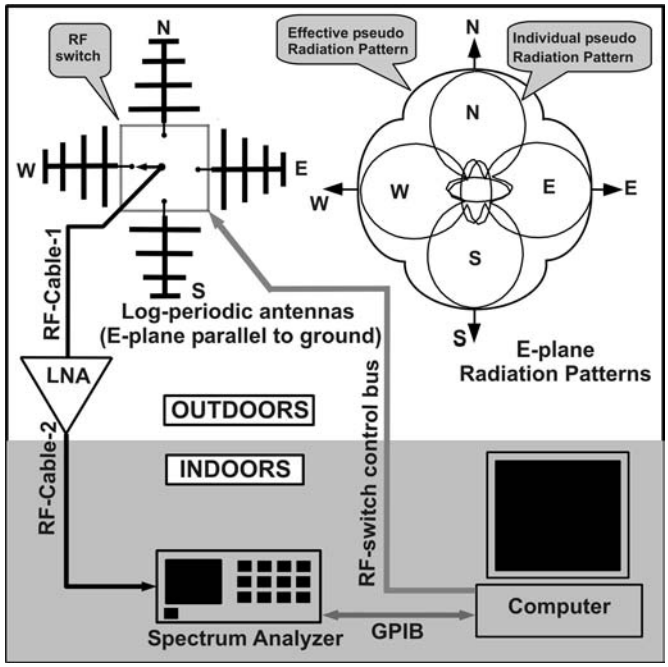


Figure 20.4: Block diagram of the omni-directional RFI Spectrum Monitoring System of GMRT.

holding all the four spectrum data. The software is accessible from the local area network. Fig. 20.5 (*a*) shows the GUI objects of the system running the software. At the back is the system settings window. The details of the spectrum analyzer's traces are visible on the top GUI. It also displays the antenna selected, time progress bar, remaining time, etc. Fig. 20.5 (*b*) shows the GUI of a data visualizing software, which can be operated from any terminal connected to the local area

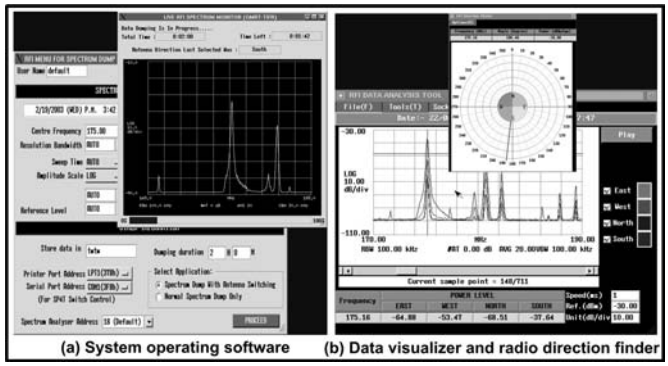


Figure 20.5: GUI of the software used. (a) GUI of the system operating the software. (b) Data visualizing and radio direction finding software.

network. This software tool is used for viewing the completed data files and can also be directly connected online to the system operating software for visualizing current data. The East, West, North, and South spectrum (one cycle) is displayed in four colors. A crosswire can be positioned by the mouse on the spectrum to visualize the power levels from all four directions for any particular frequency. The corresponding radio direction appears in a vector plot GUI sitting on top. This innovative method compares the RF-power levels with the antenna radiation patterns and finds the radio direction. It can generate time-averaged plots from completed data and also visualize the data trace-by-trace over the observed time.

20.3.2 System Analysis

As shown in the system block diagram in Fig. 20.4, the signals collected from the antennas are amplified using an LNA and fed to a spectrum analyzer. Depending on the amplified signal strength, the resolution bandwidth $\Delta\nu_{res}$ and the video bandwidth $\Delta\nu_{vid}$ of the spectrum analyzer are set. Table 20.1 lists the specifications and characteristics of the devices and instruments used for the system. The number of spectral channels in the spectrum analyzer can be expressed as in Eq. (20.1). The noise temperature of the spectrum analyzer $T_{specAna}$ is expressed in Eq. (20.2), where $T_{specAnaPhy}$ is its physical temperature. The integration time τ of the spectrum analyzer for each sweep is expressed in Eq. (20.3). The temperature contribution of the spectrum analyzer $T_{specAnaSys}$ to the system is expressed in Eq. (20.4), where G_{LNA} is the gain of the LNA, and the symbol ε with a subscript represents the RF-efficiency (output to input power ratio) of that passive device. The contribution to the system temperatures by the RF-switch, RF-cables, and the LNA are expressed in Eqs. (20.5) through (20.8), where NF_{LNA} is the noise figure of the LNA. The system temperature is expressed in Eq. (20.9). The effective integration time can be expressed as in Eq. (20.10), where N_{sweep} is the total number of video averages including post-video-averages.[28] The minimum detectable temperature ΔT_{min} of the system is expressed in Eq. (20.11). If S is the flux density of the signal at the antenna aperture in W m^{-2} Hz^{-1}, the SNR (Signal-to-Noise Ratio) of the system can be expressed as in Eq. (20.12), where k is the Boltzmann constant, G_{Ant} is the gain of the antenna, λ_{sig} is the wavelength of the signal, and $\Delta\nu_{sig}$ is its bandwidth.

$$\text{No. of Spectral Channels} = \left(\frac{\text{Frequency Span}}{\Delta\nu_{res}} \right) \tag{20.1}$$

$$T_{specAna} = \left(\frac{\text{Noise Floor}}{k\,\Delta\nu_{res}} - T_{SpecAnaPhy} \right) \tag{20.2}$$

[28] This is done in the data processing after the observation is complete.

Table 20.1: Used specifications and characteristics of the devices and instruments used.

Device	Type / Settings / Specifications
Log-Periodic Dipole Array	G_{Ant} = 4.5 dBi, Frequency Range: 50 MHz to 1800 MHz (see Table 20.2 for antenna off-source temperatures)
Low Noise Amplifier	[MAR-6, MAR-3, MAV-11A combination, from MiniCircuits] Max. noise figure NF_{LNA} = 3.5 dB, G_{LNA} = 40 dB
Spectrum Analyzer	HP-8590L, $\Delta\nu_{res}$=1 kHz, $\Delta\nu_{vid}$=1 kHz, Noise Floor = −110 dBm
RF-Switch	GSWA-4-30, Insertion loss = 1.5 dB (typical)
RF-Cable-1	RG-223, Length = 2 m (see Table 20.2 for insertion loss)
RF-Cable-2	LMR-400, Length = 120 m (see Table 20.2 for insertion loss)

$$\tau = \left(\frac{1}{2\Delta\nu_{vid}}\right) \tag{20.3}$$

$$T_{specAnaSys} = \left(\frac{T_{specAna}}{G_{LNA}\,\varepsilon_{cable1}\,\varepsilon_{cable2}\,\varepsilon_{RFswitch}}\right) \tag{20.4}$$

$$T_{RFswitchSys} = T_{RFswitchPhy}\left(\frac{1}{\varepsilon_{RFswitch}} - 1\right) \tag{20.5}$$

$$T_{cable1Sys} = \left(\frac{T_{cable1Phy}}{\varepsilon_{RFswitch}}\right)\left(\frac{1}{\varepsilon_{cable1}} - 1\right) \tag{20.6}$$

$$T_{cable2Sys} = \left(\frac{T_{cable2Phy}}{G_{LNA}\,\varepsilon_{RFswitch}\,\varepsilon_{cable1}}\right)\left(\frac{1}{\varepsilon_{cable2}} - 1\right) \tag{20.7}$$

$$T_{LNAsys} = \left(\frac{T_{LNAphy}\,(NF_{LNA} - 1)}{\varepsilon_{cable1}\,\varepsilon_{RFswitch}}\right) \tag{20.8}$$

$$T_{sys} = \left.\begin{array}{l} T_{AntOffSrc} + T_{LNAsys} + T_{specAnaSys} \\ + T_{cable1Sys} + T_{cable2Sys} + T_{RFswitchSys} \end{array}\right\} \tag{20.9}$$

$$\tau_{eff} = (N_{sweep}\,\tau), \quad N_{sweep} = 1, 2, 3... \tag{20.10}$$

$$\Delta T_{\min} = \left(\frac{T_{sys}}{\sqrt{\tau_{eff}\ \Delta\nu_{res}}} \right) \tag{20.11}$$

$$SNR = \left(\frac{S\, G_{Ant}\, \lambda_{sig}^2\, \sqrt{N_{sweep}}\, \Delta\nu_{sig}}{4\,\pi\, k\, T_{sys}\, \sqrt{2\, \Delta\nu_{res}\, \Delta\nu_{vid}}} \right), \quad \Delta\nu_{res} \le \Delta\nu_{sig} \tag{20.12}$$

Table 20.2 lists the antenna off-source temperature $T_{AntOffSrc}$, RF-cable losses, minimum detectable signal ($k\Delta T_{min}$), and the *SNR* for a signal flux density of −180 dB [w m^{-2} Hz^{-1}] at various frequencies for the system settings $\Delta\nu_{sig} \ge \Delta\nu_{res}$, $\Delta\nu_{res} = 1$ kHz, $\Delta\nu_{vid} = 1$ kHz, $N_{sweep} = 100$.

Table 20.2: Sys. Char. at −180 dB [watt/m^2/Hz], $\Delta\nu_{res} = \Delta\nu_{vid}$ =1 kHz, N_{sweep} = 100.

Freq. (MHz)	100	190	300	500	900	1300	1700
$T_{AntOffSrc}$ (°K)	419	197	170	155	152	151	151
RF-Cable-1 Insertion loss (dB)	0.27	0.36	0.46	0.65	0.86	1.06	1.24
RF-Cable-2 Insertion loss (dB)	4.8	7.1	8.4	10.8	15.4	18.2	21.1
$k\,\Delta T_{\min}$ (dB[w m^{-2} Hz^{-1}])	−207	−207	−207	−207	−207	−206	−205
SNR (dB)	70	65	61	56	50	47	43

20.3.3 Theory of Radio Direction Using Static Antenna Spectrum

Two or more antennas with their major lobes peaking along endfire directions possessing a gradual smooth falling toward the first nulls can be used for resolving radio source angles. Identical antennas are preferable. The antennas are laid on a circular arc looking toward different directions opposite to the center, in a plane parallel to the ground, but several wavelengths above it. The directions of the radio sources for any particular frequency are found by comparing powers received from the antennas with the radiation patterns. The simplest model is a two-antenna system shown in Fig. 20.6. The antennas are positioned on a circular arc of radius *R* looking toward the horizon with angle of separation α_v. The radius *R* is chosen such that the antennas do not lie in each other's near-field region as expressed in Eq. (20.13). Here, λ_{max} is the wavelength at the smallest frequency of operation, *L* is the length of the largest element, and *D* is the smallest distance of separation

between the two antennas. The antennas are kept just outside the arc of the circle. Thus, the arc-length of the semicircle between the two antennas is $R\ \alpha_v$, which may be roughly approximated to $(L + D)$ if $\alpha_v \leq 90°$. Thus, R may be expressed as in Eq. (20.14). Here, α_v has been named here as the viewing angle, since the radio sources to be resolved should fall within this angle. The maximum value of α_v is α_{max} and is determined by the maximum side-lobe level in the radiation pattern, at the operating frequency. As shown in Fig. 20.7, a circle is drawn on the major lobe with a radius equal to the maximum side-lobe level. The angle subtended by the cuts on the major-lobe is $2\ \alpha_{max}$.

$$D \geq 0.62\left(\frac{L^3}{\lambda_{max}}\right)^{\frac{1}{2}} \tag{20.13}$$

$$R \geq \left(\frac{L+D}{\alpha_v}\right) \tag{20.14}$$

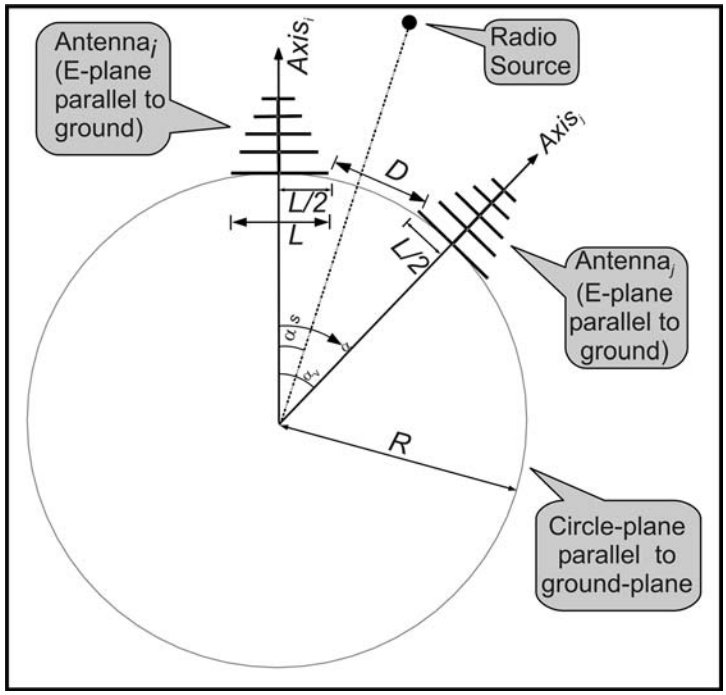


Figure 20.6: The simplest two-antenna system for radio-direction finding. Two identical log-periodic dipole arrays are positioned on the arc of a circle of radius R. α_v is the viewing angle within which the radio source must be present, α_s is the source angle, L is the length of the largest element of any antenna, and D is the minimum distance of separation.

From Fig. 20.7, considering $Axis_i$ the reference axis, let the gain patterns of the adjacent antennas (i.e., $Antenna_i$ and $Antenna_j$) within the angle α_v subtended by them be $G_i(\alpha)$ and $G_j(\alpha)$, respectively. Let their ratio be denoted by $G_R(\alpha)$ as expressed as in Eq. (20.15). It must be noted that each point of $G_R(\alpha)$ within α_v has a unique value. The antennas deliver output powers P_i and P_j corresponding to a radio source falling within α_v at the source frequency. The angle of the source can be obtained from Eq. (20.16).

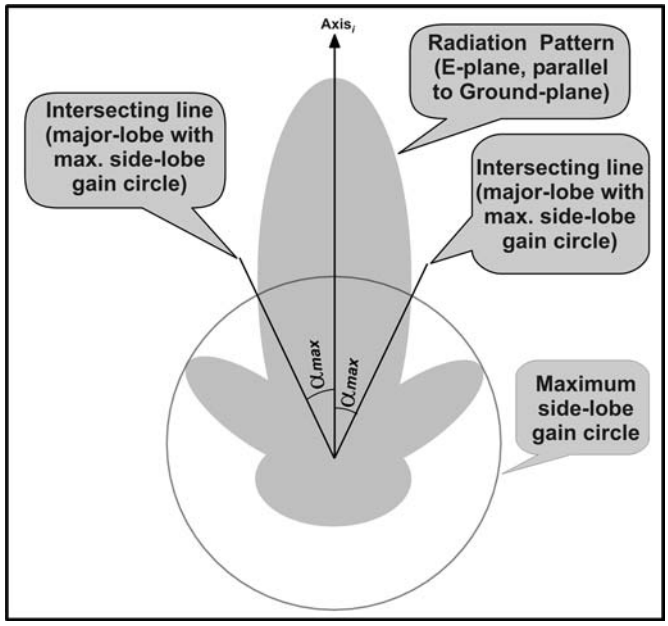


Figure 20.7: Determination of the maximum angle of separation α_{max} between two adjacent identical antennas. A circle having a radius equal to the maximum side-lobe level cuts the major-lobe at two points. These points are joined to the center with straight lines. The angle subtended is 2 α_{max}.

$$G_{R_{ij}}(\alpha) = \left(\frac{G_i(\alpha)}{G_j(\alpha)}\right), \; 0 \le \alpha \le \alpha_v \tag{20.15}$$

$$\alpha_S = G_R^{-1}\left(\frac{P_i}{P_j}\right) \tag{20.16}$$

In order to increase the effective viewing angle, more antennas could be positioned on the arc of radius R at angles 2 α_v, 3 α_v, 4 α_v, etc., with respect to $Axis_i$ until the entire 360° (2π of arc) is covered. The number of antennas required is expressed in Eq. (20.17). However, if N is not an integer, α_v can be reduced so

that N takes the value of the next possible integer. An RF-switch of the type *SPNT* can be used to multiplex the antenna outputs into the spectrum analyzer. The RFI monitoring system of GMRT (see Fig. 20.4) is a practically working system with $N = 4$. For resolving a radio source, the RF-switch is swept from $Antenna_i$ to $Antenna_N$ and correspondingly RF-powers P_i to P_N are obtained, respectively, for any particular frequency of observation. The source lies within the angle subtended by the two adjacent antennas, which deliver the highest and second highest powers. Thus, for a multi-antenna system α_s can be expressed as in Eq. (20.18).

$$N = \left(\frac{2\pi}{\alpha_v}\right) \tag{20.17}$$

$$\alpha_s = (i-1)\,\alpha_v + G_R^{-1}\left(\frac{P_i}{P_j}\right), \quad i = 1, 2, \ldots N \tag{20.18}$$

The power flux density (power per unit area per unit of bandwidth) available on the circle of radius R at the angle α_s with respect to $Axis_i$ in the direction of the radio source shown in Fig. 20.6 can be expressed as in Eq. (20.19). With an antenna gain at α_s in dBi, power at antenna terminals in dB[watt], frequency of operation in MHz and resolution bandwidth in kHz, and the flux density in dB[W m^{-2} Hz^{-1}] can be expressed in Eq. (20.20). For direction finding, the signal from both the adjacent antennas should be above the noise floor for detection. In the worst case the signal may be along one particular antenna axis so that the gain ratio of the antennas is at its maxima. Under the said condition, the signals received from both the antennas will be visible if the minimum signal-to-noise ratio of the system at that frequency is expressed as in Eq. (20.21). This is sometimes achieved by spectrum frame averaging of a large quantity of data.

$$\left.\begin{aligned} S &= \left(\frac{4\pi P_m}{G_m(\chi)\,\lambda^2\,\Delta\nu_{res}}\right) \\ &\chi = \alpha_x \text{ if } m = i,\ \chi = (\alpha_v - \alpha_s) \text{ if } m = j \end{aligned}\right\} \tag{20.19}$$

$$\left.\begin{aligned} S_{\mathrm{dB[watt/m^2/Hz]}} = {} & -68.55 + 20\log_{10}(f_{\mathrm{MHz}}) + P_m(\chi)_{\mathrm{dB[watt]}} \\ & - G_m(\chi) - 10\log_{10}\left([\Delta\nu_{res}]_{\mathrm{kHz}}\right) \end{aligned}\right\} \tag{20.20}$$

$$SNR \geq \left(\frac{G_i(0)}{G_i(\alpha_v)}\right) \tag{20.21}$$

20.3.4 Modeling and Error Analysis in Radio Direction Finding

The gain variation as a function of angle for the major-lobes of the radiation pattern of most of the directional antennas can be modeled. We have slightly modified

the formula for modeling wide-beam patterns as expressed in Eq. (20.22), where B_0 represents the maximum value of antenna gain and κ is the angle scaling factor.[29] Thus, with reference to Fig. 20.4, the power ratio for a source under observation can be expressed as in Eq. (20.23). From this the source angle is derived as in Eq. (20.24). The angular accuracy of the computed source angle due to a deviation in the radiation pattern of one of the antennas from its ideal value is analyzed here. The relative change in source angle with respect to viewing angle due to a small change in antenna output power ratio is expressed in Eq. (20.25). This is obtained by differentiating Eq. (20.24) with respect to ρ_{ij}. The usable portion of the averaged E-plane radiation pattern of the four antennas for the system described earlier is shown in Fig. 20.8. The side-lobes and back-lobe reside within the shaded area and are of less concern to us since their gains are well below the operating region. Eq. (20.22) fairly models this pattern with $n = 6$, $\kappa = 0.69$ and $B_0 = 2.82$ (4.5 dBi). Table 20.3 lists the percentage error in source angle relative to viewing angle α_v due to relative deviation of ρ_{ij} for a gain radiation pattern modeled for Fig. 20.7 using Eq. (20.22). It is observed that the error at half viewing angle is nearly one and a half times of that at its minimum and maximum values.

$$\left.\begin{aligned} &G(\theta, \phi) = B_0 \cos^n(\kappa\theta); \\ &0 \le \theta \le \frac{\pi}{2}, 0 \le \phi \le 2\pi, n \ge 1, \kappa \ge 1 \end{aligned}\right\} \tag{20.22}$$

$$\left.\begin{aligned} &\rho_{ij} = \left(\frac{\cos(\kappa \alpha_s)}{\cos\left[\kappa(\alpha_v - \alpha_s)\right]} \right)^n \\ &\text{also} \quad \rho_{ij} = \left(\frac{P_i}{P_j} \right) \end{aligned}\right\} \tag{20.23}$$

$$\alpha_s = \frac{1}{\kappa} \tan^{-1}\left(\left[\rho_{ij}\right]^{\frac{1}{n}} \frac{1}{\sin(\kappa \alpha_v)} - \cot(\kappa \alpha_v) \right) \tag{20.24}$$

$$\left.\begin{aligned} &\epsilon_\alpha = \left(\frac{\dfrac{1}{n \alpha_v \kappa \sin(\kappa \alpha_v)} \left[\rho_{ji}\right]^{\frac{1}{n}} \left(\dfrac{d\rho_{ji}}{\rho_{ji}} \right)}{1 + \left\{ \left[\rho_{ij}\right]^{\frac{1}{n}} \dfrac{1}{\sin(\kappa \alpha_v)} - \cot(\kappa \alpha_v) \right\}^2} \right) \\ &\text{also } \epsilon_\alpha = \frac{d\alpha_s}{\alpha_v} \quad \text{and} \quad d\rho_{ji} = \left(\rho_{ji_{Actual}} - \rho_{ji} \right) \end{aligned}\right\} \tag{20.25}$$

[29] For low-gain antennas having wide major-lobes, (first nulls appear after 90° on either side).

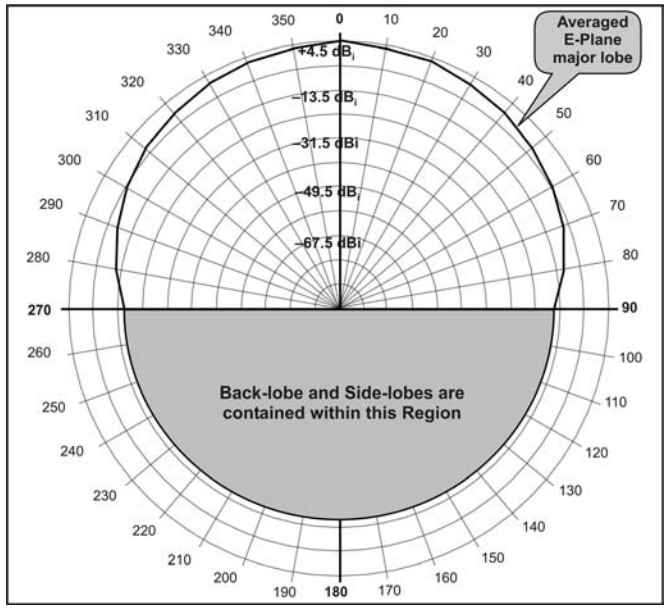


Figure 20.8: Usable region of the averaged E-plane power radiation pattern of the antennas used in the system shown in Fig. 20.4.

Table 20.3: Percentage error in α_s relative to α_v due to relative deviations of ρ_{ji} for a gain radiation pattern model with $n = 6$, $B_0 = 2.82$, $\kappa = 0.69$, $\alpha_v = 90°$ using Eq. (20.25).

$\frac{d\,\rho_{ji}}{\rho_{ji}}$	$\frac{\rho_{ji_{Actual}}}{\rho_{ji}}$	$\%\in_\alpha$ $\alpha_s = 0°$	$\%\in_\alpha$ 15°	$\%\in_\alpha$ 30°	$\%\in_\alpha$ 45°	$\%\in_\alpha$ 60°	$\%\in_\alpha$ 75°	$\%\in_\alpha$ 90°
0.25	0.9691	2.04	2.65	3.05	3.19	3.05	2.65	2.04
0.50	1.7609	4.07	5.30	6.10	6.39	6.10	5.30	4.07
0.75	2.4304	6.11	7.95	9.17	9.58	9.17	7.95	6.11
1.00	3.0103	8.14	10.60	12.21	12.77	12.21	10.60	8.14

20.3.5 Some Results of Radio Direction Finding

The averaged raw spectrum data obtained at the GMRT site from the East, West, North, and South directions between 70 MHz and 300 MHz using the RFI spectrum monitoring system of GMRT are expressed in the top portion of

Fig. 20.9. The narrowband spectral lines are separated as shown in the bottom half of the figure using some computer algorithms. Note that, except where the spectral lines are present, all other portions are removed. For each spectral line, the largest and the second largest powers are determined and the corresponding antennas are identified. If they come from adjacent antennas, the directions and absolute fluxes are determined using Eq. (20.18) in combination with Eqs. (20.24) and (20.20), respectively, otherwise the data is rejected. This is done using software specially

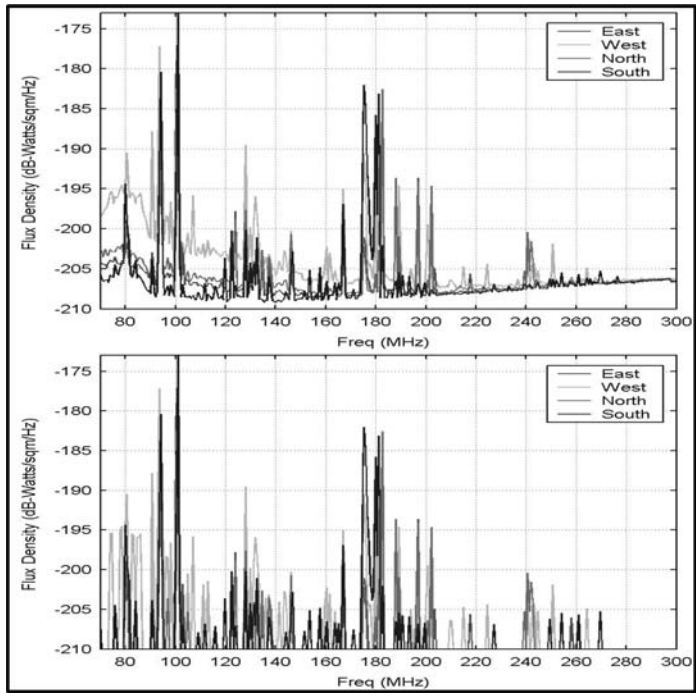


Figure 20.9: Averaged spectrums from the four-antenna system of Fig. 20.4. Top: Raw spectrum data from different directions. Bottom: Narrowband signal spectrums separated from raw data, $\Delta\nu_{res}$ = 1 kHz, $\Delta\nu_{vid}$ = 1 kHz, N_{sweep} = 100.

prepared based on the equations. The resolved radio source angles and flux densities over the entire frequency band are shown in Fig. 20.10. The impedance, radiation pattern, etc., of the log-periodic dipole antennas repeats with a logarithm of frequency. For the system shown in Fig. 20.4, the maximum antenna gain deviation from the averaged pattern (shown in Fig. 20.8) is within 1.5 dB. If the deviations in both the adjacent antenna patterns are opposite, the maximum error that can occur is 12.77%, which corresponds to the row with serial number 4 of Table 20.3. The maximum error corresponding to that is 11.5° from the actual radio direction if the model pattern is used. Table 20.4 lists the actual errors measured

for some of the FM radio and TV spectrums. Note that the percentage errors are calculated with $\alpha_v = 90°$. Since the system is omni-directional, the effective viewing angle is 360°, in which case the tabulated errors should be divided by 4.

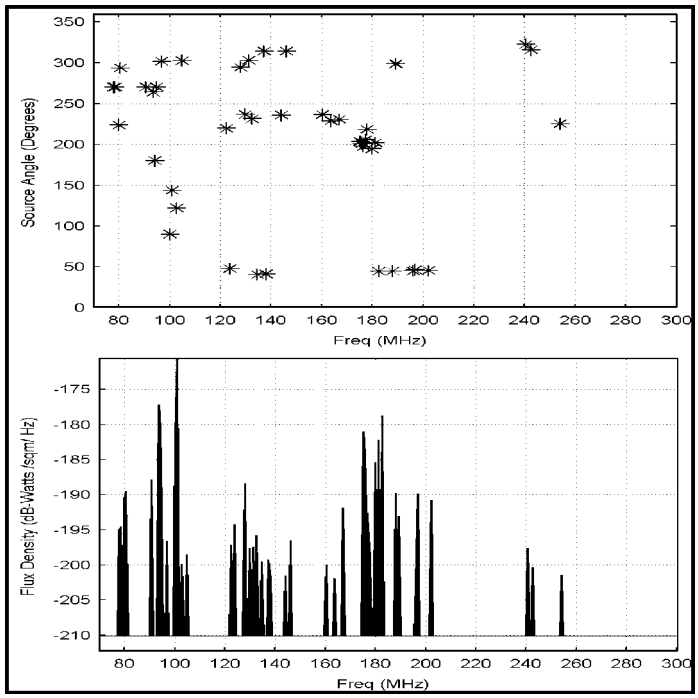


Figure 20.10: Resolved terrestrial radio sources. Top: Angular positions of the sources in a Horizon coordinate system. Bottom: Flux density at the vicinity of the antennas.

Table 20.4: Measured percentage error in α_s relative to α_v (90°) for a few transmissions.

Frequency (MHz)	93.5	100.1	175.25	180.75	196.25	201.75
Transmission Type	Radio	Radio	TV	TV	TV	TV
$\alpha_{s_{Actual}}$ **(Degree)**	266	130	200	200	55	55
$\alpha_{s_{Estimated}}$ **(Degree)**	263.6	121.7	200	201.5	44.7	45.1
Percentage error $100 \times (\alpha_{s_{Actual}} - \alpha_{s_{Estimated}})\alpha_v$	+2.67	+9.22	0.0	-1.67	+11.44	+11.0

20.3.6 Categorical Spectrum Analysis

Offline spectrum analyzing software has been developed for extracting narrowband and broadband spectrum features and performing statistical analysis. The trace data output from a spectrum analyzer usually consists of several discrete power values of corresponding frequency channels. The power versus frequency plot is produced by joining these discrete points. The RFI spectrum monitoring system saves each of these traces with a time stamp indicating the time of acquisition. The raw data after having been calibrated and integrated[30] is represented in matrix form as shown in Eq. (20.26). The first trace is saved in the first row, and the final trace in the m^{th} row. Thus, the total observation time is divided into m discrete values. The first and the n^{th} column, respectively, contain the spectral powers at the lowest and highest frequencies. As mentioned in the footnote, if the data integration is applied, the number of traces averaged N_{sweep} should be less than or equal to 10% of the number of traces in which the particular interference phenomena under study is spread out $N_{EMIspread}$ in the data, and is expressed in Eq. (20.27). Another important condition but not the least to be satisfied is determined from the ratio χ of the spectrum analyzer's frequency span $\Delta\nu_{span}$ to the maximum possible bandwidth of a narrowband spectral line $\Delta\nu_{MaxNarrowBand}$, which should be greater than 10 as expressed in Eq. (20.28). In other words, several narrowband spectrums can be accommodated within the observed spectrum range. It is equally important that each of the narrowband spectrums must be distinctly resolved within the raw data and should not overlap each other so that they appear broad enough and eventually reduce χ. This may sometimes be achieved by reducing the resolution bandwidth $\Delta\nu_{res}$ of the spectrum analyzer below the minimum possible bandwidth of a narrowband signal $\Delta\nu_{MinNarrowBand}$ as expressed in Eq. (20.29).

$$\left.\begin{aligned} X = &\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \\ &\text{where,} \\ &m = \text{ no. of time stamped traces,} \\ &n = \text{ no. of data points/trace} \end{aligned}\right\} \tag{20.26}$$

$$N_{sweep} \leq 0.1 N_{EMIspread} \tag{20.27}$$

[30] It is an optional choice that may help in improving the signal-to-noise ratio. The maximum number of traces that can be integrated/averaged is decided based on the number of traces occupied by the spectrum feature, which is under study.

$$\chi = \left(\frac{\Delta\nu_{span}}{\Delta\nu_{MaxNarrowBand}} \right), \ \chi \geq 10 \tag{20.28}$$

$$\Delta\nu_{res} \leq \Delta\nu_{MinNarrowBand} \tag{20.29}$$

For further processing, the data in Eq. (20.26) is now reorganized in groups of sub-matrices of X_{ij} as in Eq. (20.30). There are k numbers of sub-matrices in each row. Each of the X_{ij} sub-matrix is a row matrix containing (n/k) columns. The ratio (n/k) and k must be integers. Strictly speaking, the value of k should be chosen based on the definition set for the minimum bandwidth of the broadband spectrum under study in terms of the number of data points it occupies in any single time stamped trace or in any row of the matrix X. But for all practical purposes, the value of k can be chosen greater than or equal to 10 times of $n_{MaxNarrowBand}$, which is the maximum number of data points occupied by a narrowband spectrum.

$$\left.\begin{aligned} X &= \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & \vdots & \vdots \\ X_{m1} & \cdots & X_{mk} \end{bmatrix}, \\ &\text{where} \\ X_{ij} &= \left[x_i\left\{\frac{n}{k}(j-1)+1\right\} \quad x_i\left\{\frac{n}{k}(j-1)+2\right\} \quad x_i\left\{j\frac{n}{k}\right\}\right] \\ &\left(\frac{n}{k}\right) \text{is an integer, } k \geq 10\, n_{MaxNarrowBand} \end{aligned}\right\} \tag{20.30}$$

Coarse Extraction of Broadband Spectrum Features from Raw Data

As a first step toward extraction of the broadband spectrum features, the narrowband time-walls in the raw data are chopped above the untidily spread carpet of Fig. 20.2. When each of the narrowband spectral lines are distinctly visible, it is experienced that the majority of the data points in a terrestrial spectrum lie close to the noise floor of the instrument or near the broadband spectrum features, whichever is larger. Thus, the median of the data has a tendency to reside closer to the untidily spread carpet. A matrix called median matrix M derived from matrix X of Eq. (20.30) is expressed in Eq. (20.31). The sizes of M and X are identical. The elements of M are also row matrices M_{ij} each of which consist of (n/k) columns. All the elements in any particular row matrix M_{ij} are identical and are equal to the median $\widetilde{X}_{ij}$ of the corresponding row matrix of X_{ij} in Eq. (20.30). The first level of broadband information is extracted by the chopping operation shown in

Eq. (20.32) and is saved as matrix $X1$. This process of chopping may be repeated several times by substituting the original data matrix X of Eq. (20.27) with $X1$ of Eq. (20.32) and process the Eqs. (20.27) through (20.32) with smaller values of p until the narrowband spectrum features diminish toward acceptable limits.

$$\left.\begin{aligned} M &= \begin{bmatrix} M_{11} & \cdots & M_{1k} \\ \vdots & \vdots & \vdots \\ M_{m1} & \cdots & M_{mk} \end{bmatrix} = \begin{bmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \vdots & \vdots \\ m_{m1} & \cdots & m_{mn} \end{bmatrix} \\ \text{where, } M_{ij} &= \begin{bmatrix} \tilde{X}_{ij} & \tilde{X}_{ij} & .... & \tilde{X}_{ij} \end{bmatrix}, \text{ size: } 1 \times \left(\frac{n}{k}\right) \end{aligned}\right\} \tag{20.31}$$

$$\left.\begin{aligned} X1 &= \begin{bmatrix} x1_{11} & \cdots & x1_{1n} \\ \vdots & \vdots & \vdots \\ x1_{m1} & \cdots & x1_{mn} \end{bmatrix} \\ \text{where, } x1_{ij} &= m_{ij} \text{ if } x_{ij} \geq \rho\ m_{ij} \\ x1_{ij} &= x_{ij} \text{ otherwise} \\ \text{where, } \rho &\geq 1 \end{aligned}\right\} \tag{20.32}$$

Smoothening the Broadband Spectrum Features

If data in X1 is plotted, it may be observed that the slope of each of the power versus frequency curve in each row changes abruptly on either side of the locations where the major power of the narrowband spectrum features existed earlier. In other words, traces of the narrowband spectrum features still exist however less in quantity. In order to further smoothen, a discrete/fast Fourier transform operation is performed first. Next, the high frequency components in the Fourier transform output are attenuated substantially and then a discrete inverse Fourier transform is taken. The magnitude of the inverse Fourier transform is smooth and contains the broadband spectrum features.

However, if the transformations are directly applied to the data, undesirable effects near the edges (channel 1 and channel n) due to the edge discontinuities takes place. In order to prevent them, two matrices, namely P and Q, containing the extreme end data values are padded on either side of the matrix $X1$ and saved as matrix $Y1$ as shown in Eq. (20.33). The Fourier transform is applied to every individual time stamped row of matrix $Y1$, and the results are stored in matrix F as shown in Eq. (20.34). Thereafter, attenuation is applied to the high-frequency components by suitably selecting the values of α_{ij} and storing products in matrix $F1$ as shown in Eq. (20.35). Following this, an inverse Fourier transform is applied

on each of the time stamped data rows of $F1$, and the results are saved in matrix $Y2$ as expressed in Eq. (20.36). The values in the extra padded sections in $Y2$ are removed and the magnitudes of the remaining complex data are saved in $X2$ as expressed in Eq. (20.37). The entire smoothening operation involving Eqs. (20.33) through (20.37) are called second-level broadband processing. It is observed that the operations involving Eqs. (20.35) through (20.37) may produce small dips at certain regions. These may be safely corrected by comparing with the raw data thereby saving the final values in matrix $X3$ as expressed in Eq. (20.38). This operation is called third-level broadband processing.

$$\left.\begin{aligned} Y1 = [P \quad X1 \quad Q], \quad \text{where } P = \begin{bmatrix} x1_{11} & \cdots & x1_{11} \\ \vdots & \vdots & \vdots \\ x1_{m1} & \cdots & x1_{m1} \end{bmatrix}, \\ Q = \begin{bmatrix} x1_{1n} & \cdots & x1_{1n} \\ \vdots & \vdots & \vdots \\ x1_{mn} & \cdots & x1_{mn} \end{bmatrix}, \; F = \begin{bmatrix} f_{11} & \cdots & f_{1(n+2k)} \\ \vdots & \vdots & \vdots \\ f_{m1} & \cdots & f_{m(n+2k)} \end{bmatrix} \end{aligned}\right\} \tag{20.33}$$

$$\left.\begin{aligned} f_{il} = \left(\frac{1}{n+2k}\right) \sum_{j=1}^{n+2k} y1_{ij} \exp\left(\frac{-2\pi(-1)^{0.5}(l-1)(j-1)}{n+2k}\right) \\ 1 \le i \le m, \;\; l = 1, 2, \ldots (n+2k) \end{aligned}\right\} \tag{20.34}$$

$$\left.\begin{aligned} F1 = \begin{bmatrix} \alpha_{11} f_{11} & \cdots & \alpha_{1n} f_{1n} \\ \vdots & \vdots & \vdots \\ \alpha_{m1} f_{m1} & \cdots & \alpha_{mn} f_{mn} \end{bmatrix} \\ \alpha_{ij} \le 0.01 \text{ if } \zeta \le j \le (2k + n - \zeta), \\ \text{where } \zeta < \left(\frac{n}{2}\right), \;\; \alpha_{ij} = 1 \text{ otherwise} \end{aligned}\right\} \tag{20.35}$$

$$\left.\begin{aligned} Y2 = \begin{bmatrix} y2_{11} & \cdots & y2_{1(n+2k)} \\ \vdots & \vdots & \vdots \\ y2_{m1} & \cdots & y2_{m(n+2k)} \end{bmatrix} \\ y2_{il} = \sum_{j=1}^{n+2k} d_{ij} \exp\left(\frac{2\pi(-1)^{0.5}(l-1)(j-1)}{n+2k}\right), \\ 1 \le i \le m, \;\; l = 1, 2, \ldots (n+2k) \end{aligned}\right\} \tag{20.36}$$

$$\left.\begin{aligned} X2 &= \begin{bmatrix} x2_{11} & \cdots & x2_{1n} \\ \vdots & \vdots & \vdots \\ x2_{m1} & \cdots & x2_{mn} \end{bmatrix} \\ x2_{ij} &= \left| y2_{i(j+k)} \right| \quad \text{where} \quad 1 \le i \le m,\ 1 \le j \le n \end{aligned}\right\} \tag{20.37}$$

$$\left.\begin{aligned} X3 &= \begin{bmatrix} x3_{11} & \cdots & x3_{1n} \\ \vdots & \vdots & \vdots \\ x3_{m1} & \cdots & x3_{mn} \end{bmatrix}, \\ x3_{ij} &= x_{ij} \text{if } x2_{ij} > \left[\kappa(\text{smallest } x_{ij}) + x_{ij} \right], \\ & \quad 0 \le \kappa \le 1, \\ x3_{ij} &= x2_{ij} \text{otherwise} \end{aligned}\right\} \tag{20.38}$$

The processed data in $X3$ contains the broadband spectrum features. Fig. 20.11 shows the three-dimensional and contour plots of the extracted broadband spectrum features from the raw data displayed in Fig. 20.2.

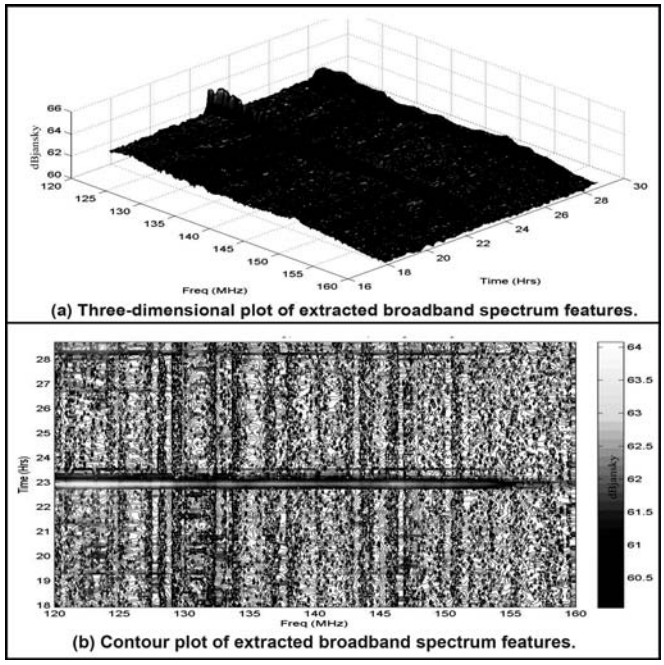


Figure 20.11: A three-dimensional plot and contour plot of the extracted broadband spectrum features in time-frequency-power space for the raw data displayed in Fig. 20.2.

Extraction of Narrowband Spectrum Features

The narrowband spectrum features are extracted by comparing the broadband spectrum features with the raw data. The narrowband spectrum of the signals appears as time-walls erected over the instrument's noise-floor in the time-frequency planes. For valid data, it may be assumed that the signal spectrum is well above the instrument's noise floor. Hence, we may define the noise-floor of the instrument to be less than or equal to the smallest value in raw data as expressed in Eq. (20.39). The pseudo scale factor β in the worst case should be 1 and for an ideal case 0.[31] The three-dimensional image envelope produced out of $X3$ is now slightly shifted up and compared with the raw data in X, thus enabling us to distinguish the narrowband spectrum features, which are then saved in $X4$ as shown in Eq. (20.40). If X exceeds the envelope, the information is of narrowband spectrum, otherwise it is noise. Fig. 20.12 shows a three-dimensional and contour plot of the extracted narrowband spectrum features from the raw data shown in Fig. 20.2.

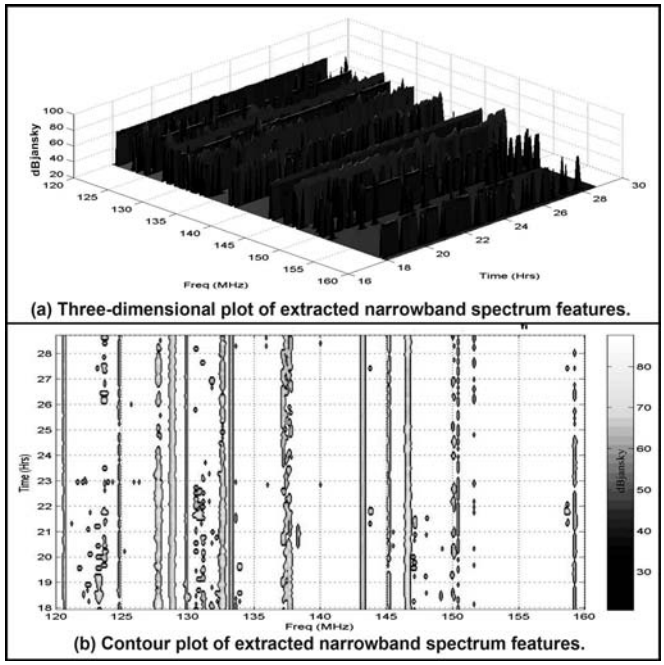


Figure 20.12: A three-dimensional plot and contour plot of the extracted narrowband spectrum features in time-frequency-power space for the data shown in Fig. 20.2.

[31] It is assumed that the noise-floor of the instrument is a constant across the entire observed band. An ideal instrument is noise-free.

$$\left.\begin{aligned}&\text{Noise-floor} = \beta\ (\text{smallest}\ \ x_{ij})\\&0 \le \beta \le 1.0,\ \ 1 \le i \le m,\ \ \ 1 \le j \le n\end{aligned}\right\} \tag{20.39}$$

$$\left.\begin{aligned}&X4 = \begin{bmatrix} x4_{11} & \cdots & x4_{1n} \\ \vdots & \vdots & \vdots \\ x4_{m1} & \cdots & x4_{mn} \end{bmatrix}\\&x4_{ij} = x_{ij}\ \text{ if }\ x_{ij} > \left[\gamma(\text{smallest}\ \ x3_{ij}) + x3_{ij}\right],\\&\qquad 0 \le \gamma \le 1\\&\text{otherwise}\ \ x4_{ij} = \text{Noise-floor},\\&\qquad 1 \le i \le m,\ \ \ 1 \le j \le n\end{aligned}\right\} \tag{20.40}$$

20.3.7 Determination of Spectral Occupancy

The spectral occupancy is defined here as the ratios of the total amount of time within which the particular spectrum features were present to the total observed time. Therefore, the data has been categorized into two types of events, existence and non-existence of spectral signature for every frequency and time. For any particular frequency, the ratio of existence events to total events is its spectral occupancy at that frequency. A value called the *MaskLevel* is set for comparing with the data and deciding for the existence or non-existence of a spectral feature as expressed in Eq. (20.41). This is compared with the respective data matrix, $X3$ or $X4$, as shown in Eq. (20.42), and the interference existence matrix E is prepared. The rows of E are averaged and multiplied by 100 to form a single row spectral occupancy matrix O as shown in Eq. (20.43). Fig. 20.13 shows the spectral statistics of the broadband and narrowband spectral features for the raw data displayed in Fig. 20.2.

$$\left.\begin{aligned}&S = X3 \text{ or } X4\\&\text{MaskLevel} = \delta(\text{smallest}\ \ s_{ij}),\ \ \ 1 \le \delta \le 2\end{aligned}\right\} \tag{20.41}$$

$$\left.\begin{aligned}&E = \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \vdots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{bmatrix}\\&\text{if}\ \ s_{ij} > \text{MaskLevel},\ \ e_{ij} = 1\\&\text{else}\ \ e_{ij} = 0,\quad 1 \le i \le m,\ 1 \le j \le n\end{aligned}\right\} \tag{20.42}$$

$$\left.\begin{aligned} O &= \left[o_{11} \;\; o_{12} \;\; .... \;\; o_{1n} \right] \\ o_{1j} &= \left(\frac{100}{n} \right) \sum_{i=1}^{m} e_{ij} \,, \;\; 1 \le j \le n \end{aligned}\right\} \tag{20.43}$$

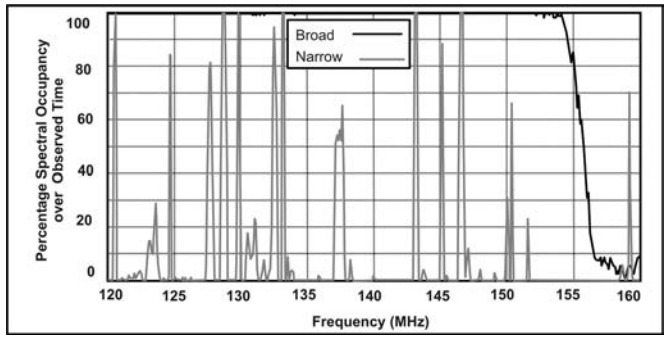


Figure 20.13: Percentage spectral occupancy over entire observed time for the data shown in Fig. 20.2.

20.3.8 Results of Some Categorically Analyzed RFI Spectrum

Figs. 20.11, 20.12, and 20.13 show the detailed categorically analyzed interferences existing at the lower frequency end of the GMRT. In Fig. 20.11, it is found that the broadband interference is present throughout observed time covering most of the frequency band. The signature resembles that of the power line interference, which generally spreads over a wideband and is also weather dependent. The possible sources are a few high tension AC and DC transmission lines running across the GMRT array, and an AC power sub-station inside the central square area of GMRT. Around the 23rd and 29th hour, a sudden increase in broadband interference is observed, which could be due to coronal arching in the nearby poles of the sub-station. Fig. 20.14 shows some analyzed results at different observing frequency bands of the GMRT. In part (*a*) of this figure, the presence of the broadband interference in the lower part of the spectrum is seen. In part (*b*), the spectral signatures of the various AM/FM and TV transmission are observed. Part (*c*) also shows another TV transmission. Part (*d*) shows the cellular communication links. The broadband spectral signatures are not detected above 250 MHz, and hence, the plots are not included here. Figs. 20.13 and 20.14 (*e*) show the spectral occupancies of the extracted features. The various factors used for the analysis have been listed in Table 20.5. Like in any other data processing, certain parameters have to be set from the inferences gathered from the dirty plot of the raw data. We have tried our best to simplify for use by quantifying the range of these parameters. To the best of our knowledge, the mathematical details provided here are adequate for anyone

who wishes to write his/her own computer code. It is felt that these algorithms can also be extended for other EMI/EMC analysis apart from radio astronomy.

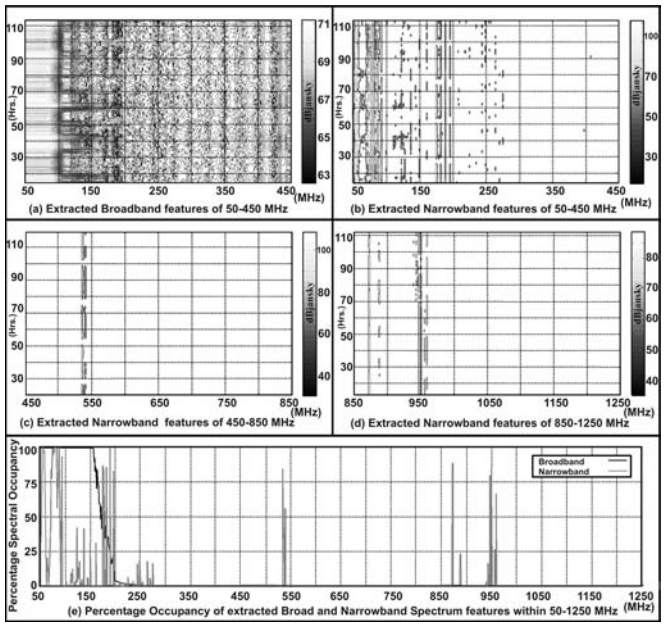


Figure 20.14: Categorically analyzed of terrestrial spectrum data collected at the GMRT site.

Table 20.5: Parameter settings for the various results of data analysis presented.

Parameter	Fig. 20.11 Fig. 20.12 120–160 MHz	Fig.20.14 (*a*) & (*b*) 50–450 MHz	Fig. 20.14 (*c*) 450–850 MHz	Fig. 20.14 (*d*) 850–1250 MHz
m	5662	12000	12000	12000
n	400	400	400	400
N_{sweep}	56	120	120	120
Δv_{res} (kHz)	100	30	30	30
χ	10	10	10	10
k	50	50	50	50
ρ	1.45	2.5	1.45	1.45

ς	20	20	20	20
α_{ij} (H-freq.)	0.0001	0.0	0.0	0.0
κ	0.0	0.6	0.6	0.6
β	0.00001	0.0001	0.0001	0.0001
γ	0.15	0.2	0.2	0.2
δ	1.45	1.45	1.45	1.45

20.4 BANDWIDTHS FOR RADIO ASTRONOMY

In general, the allocated bandwidth for radio astronomy is only a few MHz centered at certain frequencies important from radio astronomy's point of view. There is a huge ongoing battle between the communication industries and radio astronomers for capturing unallocated spectrum bands. There is very little bandwidth left for allocation at higher frequencies. As the atmosphere is not transparent to the whole electromagnetic spectrum, the spectrum use for radio astronomy is very limited. Out of this, the majority of the bandwidth is used for commercial purposes due to which radio astronomers are taking refugees in deserted places where the communication signals do not reach the radio telescopes. There is also some thinking among astronomers to launch satellites and also possibly use the Moon for setting up radio telescope antennas. But these are an extremely adventurous and costly affair. Radio astronomers have formed international committees to present their claims for spectrum usage globally.

Certain frequency bands are not always occupied by communication users. Radio astronomers feel that these frequency bands could be used efficiently for exploring the space when they are free. The major requirement of such technologies is an ultra wideband feed fitted on a parabolic dish, which can cover decades of bandwidths. The free spectrum can therefore be extracted by setting a selectable bandwidth band-pass filter. This futuristic view of acquiring increased bandwidth of the front-end system of radio telescopes has led many engineers to develop highly sensitive antenna feeds covering decade bandwidths. Some are reported to be partially successful and have shown the possibility of attaining perfection. One such feed has also been reported from the GMRT. This feed has already attained the perfection for use in communication systems.

20.5 NEWLY DEVELOPED ULTRA WIDEBAND ANTENNA-FEED FOR GMRT

The detail of the antenna feed is shown in Fig. 20.15. It consists of two log-periodic planar toothed trapezoidal structures (for orthogonal polarizations) with a Step-Lane reflector (thick lines) or a Pyramidal reflector (dotted lines).[32] An optional secondary reflector can also be seen, which boosts the gain at the lowermost frequencies. The antenna has been fabricated first on printed circuit board and later again on a metallic sheet cut, shaped, and fixed on a low permittivity thin dielectric surface (for mechanical support). In order to further reduce the effect of the dielectric, slots were made by removing the dielectric between the dipoles. This has also reduced the resistance to air flow. For any particular design, only one of the frequency independent reflectors, Step-Lane or Pyramidal, is positioned below the antennas. The step widths of the Step-Lane reflector are kept identical

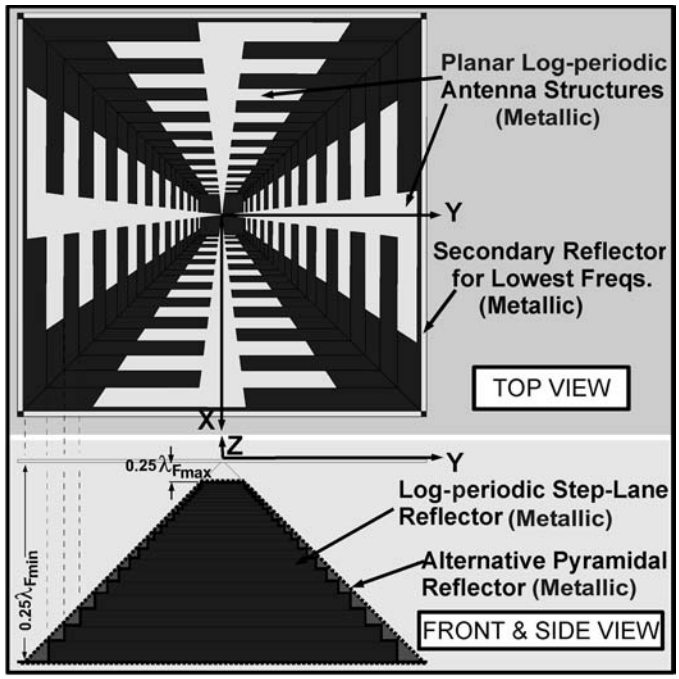


Figure 20.15: Elevation and plan of the dual polarized planar log-periodic antenna structures associated with a Step-Lane reflector (dark lines) or a Pyramidal reflector (dotted lines). An optional secondary reflector is provided for improving the gain at the lowermost frequencies.

[32] The names *Step-Lane* and *Pyramidal* has been given by the inventors of the frequency independent reflectors for use with frequency independent antennas.

to the dipole widths of the antennas, such that the former hide themselves behind the antennas when viewed from the top. For any particular frequency, the Z-axis distance between a dipole and its corresponding reflecting step is about a quarter of an wavelength. Similarly, for the case of the Pyramidal reflector, the Z-axis distance between any dipole and the reflector surface is maintained as a quarter of a wavelength. Thus, for any radiation falling directly perpendicular to the antenna plane (along Z-axis) within the designed frequency range, there exists a dipole and a reflecting surface to enhance the gain at that frequency. Essentially, within the designed spectrum band, the flux escaping the respective frequency dipoles get reflected back on them with an identical phase for every frequency, and hence named *frequency independent reflector*. Effectively, the reflector converts the bidirectional radiation pattern into a unidirectional pattern with improved average antenna gain approximately by a factor of 2.0.

The gain simulations were done using an antenna and electromagnetic modeling software script on MATLAB. Two computational models were prepared; one with the Step-Lane reflector and other with the Pyramidal reflector. The secondary reflector was included in both the models. Each of these models was broken into more than 4000 finite element triangles. The modeling and meshing were done using the GNU GMSH software. Fig. 20.16 shows the simulated and measured[33] gains over the 1:10 frequency bandwidth. The quantity *F* stands for the minimum cut-off frequency of the ultra-wide band. The average gain over the ultra-wide band is 8.9 and 9.1, respectively, for a Step-Lane reflector and a Pyramidal

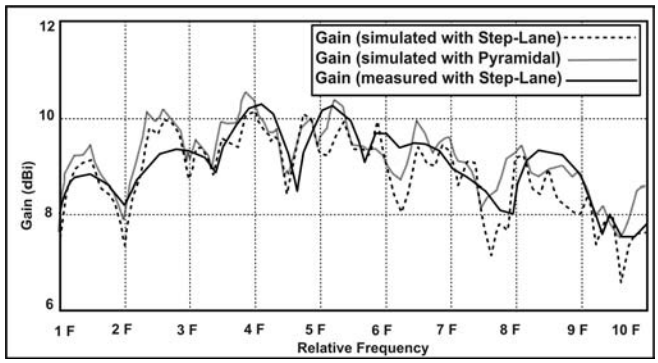


Figure 20.16: Gain variation of the antenna-feed over the ultra-wide band. Please note that the measured gain is for an antenna-feed designed for a frequency range of 200 MHz to 2000 MHz for use on the GMRT dishes.

[33] The measured gains stand for a design specially done for GMRT using a Step-Lane reflector. The frequency range of this antenna-feed is from 200 MHz to 2000 MHz. The Pyramidal reflector version of the antenna-feed is under the process of fabrication and yet to be tested.

reflector backed antenna-feed. The characteristics of both the polarizations very much resemble each other and polarization leakage isolation was measured to be around 20 dB and varies over the ultra-wide band with a peak value of 30 dB. For all the experimental measurements ultra-wide band baluns were used to convert the balance terminals of the antenna-feed into 50 Ohms unbalanced coaxial terminals. The VSWR (Voltage Standing Wave Ratio) was measured as below 2 during the course of the measurements. Hence, results are presented for one polarization.

20.6 THE SQUARE KILOMETER ARRAY (SKA)

The sensitivity of a radio telescope is dependent on the power it can collect from the astronomical sources. Thus, the more light (radio wave) collecting area a telescope can have, the more will be its sensitivity. The angular resolution of any telescope is measured roughly as λ/D, where λ is the wavelength and D is the diameter of the reflector. The interferometers can further increase the angular resolution, which is a function of the baselines. Unlike optical telescopes that operate in wavelengths of angstrom units, radio telescopes operate in wavelengths from a few mm to a few meters (the lower part of the spectrum of radio telescopes are of great importance). The major problem of building large radio telescopes is the huge size of the paraboloid reflectors. The sensitivity of a telescope depends on the total light collecting area. In radio wavelengths, the signals are extremely weak and require a huge collecting area. Thus, radio telescopes are fabricated in pieces—many antennas are placed in an array. The array powers may be combined in interferometer mode or the phased array mode depending on applications. For example, for observing distant sources like a radio galaxy which is an extended radio source, the interferometer mode would be useful since one requires high angular resolution for producing an intensity map of the extended radio source. On the other hand, for observing a pulsar, one may use the phased array mode since its electromagnetic properties are of interest to the observer and not the shape of the source. But in either case, the total collecting area plays a major role in improving sensitivity.

The Square Kilometer Array, or SKA, is a proposed integrated radio telescope, which intends to produce a light collecting area in squares of kilometers over the frequency range roughly covering 100 MHz to 25 GHz. The frequency range may further be extended below and above the said limits. The super collecting area will make it 50 times more sensitive than any other radio telescope array that exists today. Various engineering techniques such as electronic beam switching and antenna band sharing are proposed to be implemented so that several radio astronomers may use the instrument simultaneously for looking into different parts of the sky and at different frequencies of observations. Interferometer

techniques will be used to make images of much distant sources, which are not possible today with present instruments. An international consortium consisting of representatives from various countries is working on the design aspects of the SKA, which is to be finalized in 2008.

The array is expected to comprise many antennas spread over an area of several thousands of kilometers. Half of its collecting area will be from a compact core of antennas placed within 150 km. The remaining antennas will be spread up to a few thousand kilometers to increase its angular resolution by increasing the baseline. A choice has to be made between the two sites short listed for building the array, Mileura station in Western Australia and Karoo Basin in the Northern Cape area of South Africa. After the site is selected, construction of the SKA will begin in 2010, with observations to start from 2015 onwards. The complete functionality is expected to be available from 2020 onwards. The instrument will be sensitive to detect every AGN out to a red-shift of 6, when the Universe was less than one billion years old.

EXERCISES

1. What is meant by radio interference in radio astronomy?
2. What are the possible ways by which the RFI enters into data using the radio telescope antenna?
3. Describe the properties of narrow and broadband RFI and give examples.
4. Describe a spectrum monitoring system with a block diagram.
5. Why does a spectrum monitoring system use multiple antennas?
6. Convert the power 1×10^{-14} W/m^2/Hz to jansky and dBjansky.
7. Two antennas of identical patterns are positioned making an angle $\alpha_v = 90°$ with respect to each other, and their E-planes are parallel to the ground. A radio frequency interference signal polarized in E-plane enters within this angle. The radiation patterns of the antennas 1 and 2 in the E-plane can be respectively expressed as $G_1(\alpha) = G_0 \cos(\alpha)$ and $G_2(\alpha) = G_0 \cos(\alpha_v - \alpha)$. Let P_1 and P_2 be the power outputs from antennas 1 and 2 respectively due to the RFI. Find the value of α if the ratio of output powers $\rho_{21} = P_2/P_1$ (from antennas 2 and 1) is found as:
   (*a*) 2. (*b*) 1. (*c*) 0.5.
8. What are the advantages for analyzing the RFI spectra categorically?
9. An RFI line spectrum is seen for 8 hours within 24 hours of observation. What is its spectral occupancy?
10. A radio astronomer has an ultra wideband antenna-feed mounted on his radio telescope dish. He also has a spectrum monitoring station. How can the radio astronomer utilize both of these instruments together for getting more useful data?
11. Describe the square kilometer array in brief.

APPENDIX A

BACKGROUND INFORMATION, FORMULAE, RELATIONSHIPS

A.1 CHARACTERISTICS OF THE SUN'S RADIATION

In the Sun the gases that are at high temperatures produce electrically charged particles. Due to their collisions, electromagnetic waves of different wavelengths are emitted by the Sun, which extend from gamma rays ($\lambda = 10^{-13}$ m) to radio waves ($\lambda = 10^{3}$ m). However, about 99% of solar radiation lies between 0.15 to 4.0 μm. Out of this: (*i*) about 46% lies in IR (0.7 μm to 2.3 μm), which causes radiant heat with some chemical effect; (*ii*) about 44% lies in the visible light (0.38 μm to 0.7 μm), violet to red; (*iii*) about 9% lies in the UV radiation (0.15 μm to 0.38 μm), which causes photochemical effects, sunburn, bleaching, etc. All these radiations travel with 3×10^{8} m/s in an average surface temperature of 6,000° K, while the Earth emits at its average surface temperature of 15° C or 288° K.

From Wein's law we can write

$$\lambda_{max} = \frac{\text{constant}}{T},$$

where the constant = 2897 μmK ≈ 3000 μmK.

For the Sun,

$$\lambda_{max} = \frac{3000\ \mu mK}{6000\ K} = 0.5\ \mu m.$$

This shows that the Sun emits a maximum amount of radiation energy corresponding to the wavelength of 0.5 μm.

On the other hand, we can write for the Earth:

$$\lambda_{max} = \frac{3000\ \mu mK}{300\ K} = 10\ \mu m \qquad \text{[considering } T = 300\ K\text{]}.$$

This shows that the Earth emits a maximum amount of radiant energy corresponding to the wavelength of 10 μm.

From the above calculation, we understand why solar radiation is called shortwave radiation and terrestrial radiation is called long-wave radiation.

The mean Earth-Sun distance is $a_{\oplus} = 1\ AU \approx 1.5 \times 10^{11}$ m. Also, the angular diameter of the Sun as observed from the Earth = 9.3×10^{-3} radian, which is slightly more than half a degree. Hence,

$$\frac{arc}{radius} = \text{radian;} \quad \text{i.e., arc} = \text{radius} \times \text{radian.}$$

Now, the Sun's diameter = Earth-Sun distance × angular diameter of the Sun

$$= 1.5 \times 10^{11} \times 9.3 \times 10^{-3}\ m = 13.95 \times 10^{8}\ m.$$

The radius of the Sun $\approx 6.97 \times 10^{8}$ m.

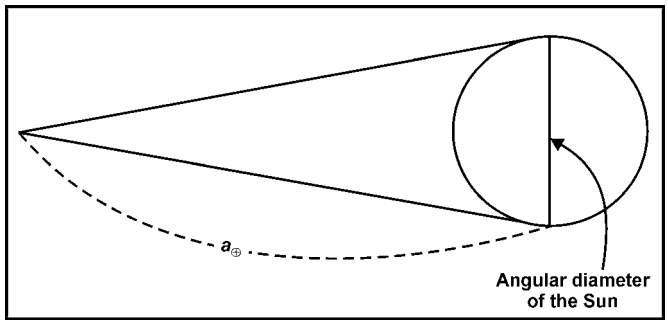


Figure A.1.1: Earth-Sun distance vs angular diameter of the Sun.

A.2 ACCELERATION DUE TO GRAVITY OF THE SUN

We have

the mass of the Sun, $M_\Theta = 1.99 \times 10^{30}$ kg, and

the radius of the Sun, $R_\Theta = 6.97 \times 10^8$ m.

Now using the formula,

$$g_\Theta = \frac{G M_\Theta}{R_\Theta^2} \qquad [\text{where, } G = 6.67 \times 10^{-1}\ \text{Nm}^2/\text{kg}],$$

we get $g_\Theta = 30\, g_\oplus$, where $g_\oplus$ = Earth's acceleration due to gravity

A.3 THE DENSITY OF THE SUN

If ρ_Θ represents the density of the Sun then,

$$\rho_\Theta = \frac{M_\Theta}{V_\Theta} = \frac{3}{4}\frac{M_\Theta}{\pi {R_\Theta}^3} = \frac{3}{4} \times \frac{1.99 \times 10^{30}}{3.14[6.96 \times 10^8]^3} \approx 1.41 \times 10^3\, \text{kg/m}^3$$

$$= 1.41 \times \text{density of water.}$$

It may be mentioned here that the average density of the Earth = 5.5×10^3 kg/m^3.

A.4 SOLAR CONSTANT

The amount of solar radiation (i.e., energy incident on a unit area in unit time on a surface held at right angles to the solar beam at the outer boundary of the atmosphere) is called the solar constant. If it is denoted by S_Θ then,

$$S_\Theta = 1.36 \times 10^3\ \text{w/m}^2 = 1360\ \text{joules/m}^2$$

[where unit area = 1 m^2 and unit time = 1 second].

The solar constant varies ±2% due to variation in the solar output and varies about 3.5% owing to changes of the Earth-Sun distance.

A.5 SOLAR LUMINOSITY

If L_Θ represents the solar luminosity and $a_\oplus$ the Earth-Sun distance then

$$L_\Theta = 4\pi {a_\oplus}^2 \times S_\Theta = 4 \times 3.14 \times (1.5 \times 10^{11})^2 \times 1.36 \times 10^3 \approx 3.83 \times 10^{26}\ \text{w.}$$

A.6 THE SUN'S BRIGHTNESS

If S denotes the brightness of the Sun, then

$$S = \frac{L_\Theta}{4\pi R_\Theta{}^2} = \frac{3.83\times10^{26}}{4\times3.14[6.96\times10^{8}]^2} \approx 6.29\times10^{7}\ \text{w/m}^2$$

It has been noted that mass of the Sun consists of about 80% hydrogen, 19% helium, and 1% other elements. The Sun is a burning gaseous globe with a radius of about 7×10^{15} km and a mass of about 2.0×10^{30} kg.

A.7 SOLAR ENERGY

It has been estimated the solar energy received per year is 1.5×10^{15} Mwh. On a clear sunny day, 1 acre of land in the daytime receives solar energy equivalent to 16 bbl (bbl = billion barrels) of crude oil. The total solar energy that can be intercepted by the Earth = 2.55×10^{18} cal min^{-1} = 3.67×10^{21} cal day^{-1} .

The deposition of solar radiation in the Earth atmosphere is given by

$$Q_s = C_r + A_r + C_a + A_a + (Q + q)\,(1 - \alpha) + (Q + q)\alpha.$$

Here, Q_s = amount of solar radiation incident at the top of the atmosphere = 263 kly/yr.

Q = Direct beam of solar radiation incident on horizontal surface at the ground = 82 kly/yr.

q = Diffused solar radiation incident on horizontal surface at the ground = 58 kly/yr.

C_r = Solar radiation reflected and scattered back to space by clouds.

A_r = Solar radiation reflected and scattered back to space by aerosol (i.e., air molecules, dust, and water vapor of the atmosphere).

C_a = Solar radiation absorbed by the clouds.

A_a = Solar radiation absorbed by atmospheric aerosols.

α = Albedo of the surface ≈ 0.1143.

Solar energy per unit area is usually expressed in langely (ly) or kilolangely (kly), where

$$1\ \text{ly} = 1\ \text{cal/cm}^2$$

$$1\text{kly} = 1000\ \text{ly}$$

It is assumed that in each minute the Sun radiates about 56×10^{26} cal of energy in all. Now the solar constant

$$S_\Theta = \frac{56 \times 10^{26}}{4\pi(a_\oplus{}^2)} \text{ cal/min} = \frac{56 \times 10^{26}}{4 \times 3.14 \times [1.5 \times 10^{13}]^2}$$
$$= 2.0 \text{ ly/min} = 1359 \text{ w/m}^2.$$

The total solar energy intercepted by the Earth in unit time = $\pi R_\oplus^2 S_\Theta$

$$= \pi[6.40 \times 10^8 \text{ cm}]^2 S_\Theta \approx 2.55 \times 10^{18} \text{ cal/min} \approx 3.67 \times 10^{21} \text{ cal/day}.$$

The total solar energy intercepted by the Earth is illustrated in the figure below.

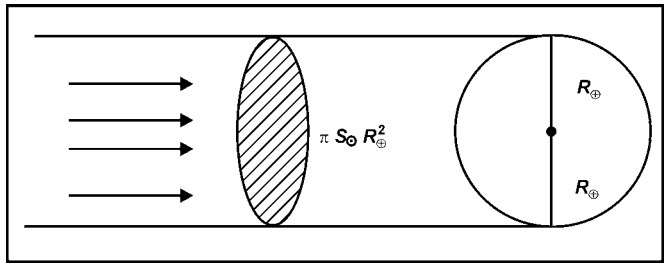


Figure A.7.1: Total solar energy intercepted by the Earth.

The amount of solar energy received at the top of the atmosphere in unit area in unit time is given by,

$$Q_s = \frac{\pi R_\oplus{}^2 S_\Theta}{4\pi R_\oplus{}^2} = \frac{S_\Theta}{4} = 0.5 \text{ ly/min}$$

$$= \frac{0.5 \times 60 \times 24 \times 365.25}{1000} = 263 \text{ kly/year} = 339.75 \text{ w m}^{-2}.$$

Again, $$Q_s = \frac{S_\Theta}{4} = \frac{1359}{4} = 339.75 \text{ w m}^{-2}.$$

The average albedo of the atmosphere = 32%. Thus, the isolation incident on the surface of the Earth = 68% of 339.75 w $m^{-2} \approx 231$ w m^{-2}.

It may be noted that the efficiency of the atmosphere is conversion of insulation is only about 1% to 2%. The annual average solar energy received in northern hemisphere for the layer 1000 to 75 hpa is given by,

$$\text{PE} = 567.5 \times 10^6 \text{ J m}^{-2}.$$

$$\text{Internal energy} = 1674.8 \times 10^6 \text{ J m}^{-2}.$$

$$\text{Total PE} = 2242.3 \times 10^6 \text{ J m}^{-2}.$$

$$\text{KE} = 1153.4 \times 10^3 \text{ J m}^{-2}.$$

$$\text{Therefore, an average amount of KE} = \frac{1153.4 \times 10^3}{2242.3 \times 10^6} \approx \frac{1}{2000} \text{ of total PE.}$$

A.8 CIVIL DAY AND SIDEREAL DAY

It is seen that the stars take about 3 minutes and 56 seconds less than the Sun to go once around the Earth. So we can write,

$$1 \text{ civil day} = 24 \text{ hours.}$$

$$1 \text{ sidereal day} = 23^{\text{h}}\ 56^{\text{m}}\ 4^{\text{s}}.$$

A.9 SOLAR YEAR AND CIVIL CALENDAR

The time taken by the Sun to complete a revolution around the Earth, as observed from the Earth, is defined as a solar year. Depending on the points chosen we have different types of solar years, out of which we consider here the sidereal solar year and the tropical solar year.

The duration of the sidereal year = 365.256364 days.

The duration of a tropical year = 365.242190 days.

It is to be noted that

$$T \text{ mean solar day} = (T + 1) \text{ sidereal days.}$$

Thus, approximately

$$1 \text{ sidereal day} = \frac{T}{T+1} = \frac{365.25}{366.25} \text{ civil day} = \frac{365.25 \times 24}{366.25} \text{ hours}$$

i.e., $1 \text{ sidereal day} = 23^{\text{h}}\ 56^{\text{m}}\ 4.1^{\text{s}}.$

A.10 SOLAR MONTH AND LUNAR MONTH

The angular extent of the ecliptic is 360° and the Sun takes about 365.25 civil days to complete one revolution. A solar month is equal to about 365.25/12 days. To be precise, we have

1 mean sidereal solar month = 365.256364/12 days = 30.43803 days.

The Moon like the Sun moves from the West to the East with reference to the fixed stars. The motion of the Moon is much faster than that of the Sun. The Moon moves along the orbit inclining toward the ecliptic at an angle of about 5°8′. Since the angle is small it is assumed that the Moon is moving along the ecliptic itself. It is found that the Moon takes an average period of 27.3216615 days to complete a revolution with reference to the fixed stars. This time interval is known as the sidereal period of the Moon.

A.11 TIMES OF SUNRISE AND SUNSET

If H is the hour angle, then $\cos H = (\sin a - \sin \phi \sin \delta)/\cos \phi \cos \delta$, where a is the altitude, δ is the declination, and ϕ is the latitude at the time of sunrise and sunset, and the Sun's altitude $a = 0$.

Therefore, $\cos H = -\tan \phi \tan \delta$.

This is the required relations to be used. We give an example below.

Find the times of sunrise and sunset and also the duration of day and night at Bangalore ($\phi = 13°$ N) on May 15, given that the Sun's declination is 18°47′N.

Solution:

We have $\cos H = -\tan(+13°) \tan(+18°47') = -0.078519$.

Therefore, $H = 95.503444° = 6^h\ 18^m\ 01^s$.

Hence, we have,

(*i*) Time of sunrise = $12^h - 6^h\ 18^m\ 01^s = 5^h\ 41^m\ 59^s$ a.m.

(*ii*) Time of sunset = $12^h + 6^h\ 18^m\ 01^s = 6^h\ 18^m\ 01^s$ p.m.

(*iii*) Duration of the day = $2H = 2\ (6^h\ 18^m\ 01^s) = 12^h\ 36^m\ 02^s$

(*iv*) Duration of night = $24^h - 12^h\ 36^m\ 02^s = 11^h\ 23^m\ 58^s$.

APPENDIX B

STANDARD SYMBOLS

n	Symbol of a neutron
p^+	Symbol of a proton
e^-	Symbol for an electron
e^+	Symbol for a positron
λ	Symbol of wavelength
μ	Symbol for refractive index
γ	Symbol of gamma ray
ν	Symbol of frequency
${}^{12}_{6}C$	Symbol of nuclei for carbon isotope. The superscript represents the atomic mass (in amu) and the subscript is the atomic number.
G	Gravitational constant = 6.670×10^{11} N m^2 kg^{-2}
h	Planck's constant = 6.6262×10^{-34} J s
k	Boltzmann's constant = 1.3806×10^{-23} J K^{-1}
m_e	Mass of an electron = 9.1096×10^{-31} kg
amu	Atomic mass unit = 1.6605×10^{-27} kg
c	Velocity of light in vacuum = 2.9979×10^8 m s^{-1}
e	Charge of an electron 1.6022×10^{-19} C
eV	Electron Volt = 1.6022×10^{-19} J
U,B,V	Magnitudes through the standard U B V photometric system
ε_0	Permittivity of free space = 8.854×10^{-12} F m^{-1}
μ_0	Permeability of vacuum = $4\pi \times 10^{-7}$ H m^{-1}

APPENDIX C

ABOUT THE CD-ROM

- Included on the CD-ROM are figures from the text(most in color) and other files related to topics for astronomy and astrophysics
- See the “README" files for any specific information/system requirements related to each file folder, but most files will run on Windows XP or higher

REFERENCES

Adams D. J. (1980). Cosmic X-ray Astronomy (Bristol: Adam Hilger).

Alderman S. J. et al. (ed) (1992). Automated Telescopes for Photometry and Imaging (Astronomical Society of the Pacific Conference Series) vol. 28.

Alloin D. M., Mariotti J. M. (1989). Diffraction limited imaging with very large telescopes (Darderchet: Kluwer).

Barden S. C. (ed) (1988). Fiber Optics in Astronomy (Astronomical Society of the Pacific Conference Series vol. 3).

Barford N. C. (1985). Experimental Measurements: Precision, Error, and Truth (New York, Wiley).

Birney D. S. (1991). Observational Astrophysics (Cambridge: Cambridge University Press).

Bhattacharya A. B., Kar S. K. and Bhattacharya R. (1997). Theoretical Applied Climatology, vol. 58, p. 95.

Bode M. F. (1995). Robotic Observations (New York, Wiley).

Bray D. E. (1992). The Observations and Analysis of Stellar Photosperes (Cambridge: Cambridge University Press).

Buddin E. (1993). Astronomical Photometry (Cambridge: Cambridge University Press).

Comte G. and Marcelin M. (ed) (1955). Tridemensional Optical Spectroscopic Methods in Astrophysics (Astronomical Society of the Pacific Conference Series vol. 71).

Cooper W. A. and Walker E. N. (1989). Getting the Measure of Stars (Bristol: Adam Hilger).

Davis Phillip A. G. et al. (ed) (1990). CCDs in Astronomy II (L Davis Press).

di Gesu V., Scarsi L., Crane P., Friedman J. H. and Levialdi S. (ed). (1985). Data Analysis in Astronomy IV (New York: Plenum).

di Gesu V., et al. (1992). Data Analysis in Astronomy IV (New York: Plenum).

Edward W. K. and Michael S. T. (1990). The Early Universe (Westview Press).

Emerson D. T. and Payne J. M. (ed) (1995). Multi-Feed Systems for Radio Telescopes (Astronomical Society of the Pacific Conference Series vol. 75).

Evans J. V. and Hagfors T. (1968). Radar Astronomy (Dordrecht: Reidee).

Goddard D. E. and Milne D. K. (ed) (1994). Parkes: Thirty Years of Radio Astronomy (CSIRO).

Golay M. (1974). Introduction to Astronomical Photometry (Dordrecht: Reidel).

Guyenne T. (ed) (1995). Future Possibilities for Astronomy in Space (European Space Agency).

Harwit M. (1973). Astrophysical Concepts (New York: Wiley)

Hearnshaw J. B. (1996). Measuring of Starlight (Cambridge: Cambridge University Press).

Hiltner W. A. (ed) (1962). Astronomical Techniques (Stars and Stellar Systems vol. II) (Chicago: University of Chicago Press).

Hopkins J. (1976). Glossary of Astronomy and Astrophysics (Chicago: University of Chicago Press).

James B. and Scott T. (1987). Galactic Dynamics (Princeton University Press).

Jeschek C. and Murtagh F. (ed) (1990). Errors, Bias and Uncertainties in Astronomy (Cambridge: Cambridge University Press).

Joardar S. and Bhattacharya A. B. (2006). *Journal of Electromagnetic Waves and Applications*, vol. 20, p. 1195 & p. 1465.

Joardar S. and Bhattacharya A. B. (2006). *Journal of Electromagnetic Waves and Applications*, vol. 20, No. 11, 1465–1479, 2006.

Joardar S. and Bhattacharya A. B. (2007). *Journal of Electromagnetic Waves and Applications*, vol. 21, p. 441.

Joardar S., Datta A., Bhattacharya A. B. and Bose S. K. (2007). *IETE Journal of Research*, vol. 53, no. 1, p. 83.

Joardar S. and Bhattacharya A. B. (2007). Progress In Electromagnetics Research, PIER 72, p.127.

Kitchin C. R. (1995). Telescopes and Techniques (Berlin: Springer).

Kitchin C. R. and Ferrest R. F. (1997). Seeing Stars (Berlin: Springer).

Kitchin C. R. (2005). Astrophysical Techniques (New Delhi: Overseas Press India Pvt. Ltd.)

Kraus J. D. (1966). Radio Astronomy (New York: McGraw Hill).

Lang K. R. (1978). Astrophysical Formula (Berlin: Springer).

Mackintosh A. (1986). Advanced Telescope Making Techniques, vol. I & II (William Bell).

Manly P. L. (1991). Unusual Telescopes (Cambridge: Cambridge University Press).

Manly P. L. (1994). The 20 cm Schmidt Cassegrain Telescope (Cambridge: Cambridge University Press).

McClean I. S. (1994). Electronic and Computer Aided Astronomy (Chichester: Ellis Howard).

Michael Z. and Stephen A. G. (1998). Introductory Astronomy Astrophysics (Harcourt Brace College Publishers, U.S.).

Perley R. A. et al (ed) (1989). Synthesis Imaging in Radio Astronomy (Astronomical Society of the Pacific Conference Series, vol. 6).

Richard T., James M. M. and George W. S. (2001). Interferometry and Synthesis in Radio Astronomy (John Wiley & Sons).

Rohlfs K. and Wilson T. L. (1986). Tools of Radio Astronomy (Berlin: Springer).

Roy A. and Clarke D. (1982). Astronomy: Principles and Practice (2^{nd} ed) (Bristol: Adam Hilger).

Rudolf K. (1994). Discovering the Secrets of the Sun (John Wiley & Sons).

Wall J. V. (1993). Optics in Astronomy (Cambridge: Cambridge Univerity Press).

Wall J. V. and Boksenberg A. (1990). Modern Technology and Its Influence on Astronomy (Cambridge: Cambridge University Press).

Warner B. (1988). High Speed Astronomical Photometry (Cambrige: Cambridge University Press).

Zensus J. A. et al (ed) (1995). Very Long Baseline Interferometry and the VLBA (Astronomical Society of the Pacific Conference Series, vol. 82).

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z