Baker and Fredrick

ASTRONOMY

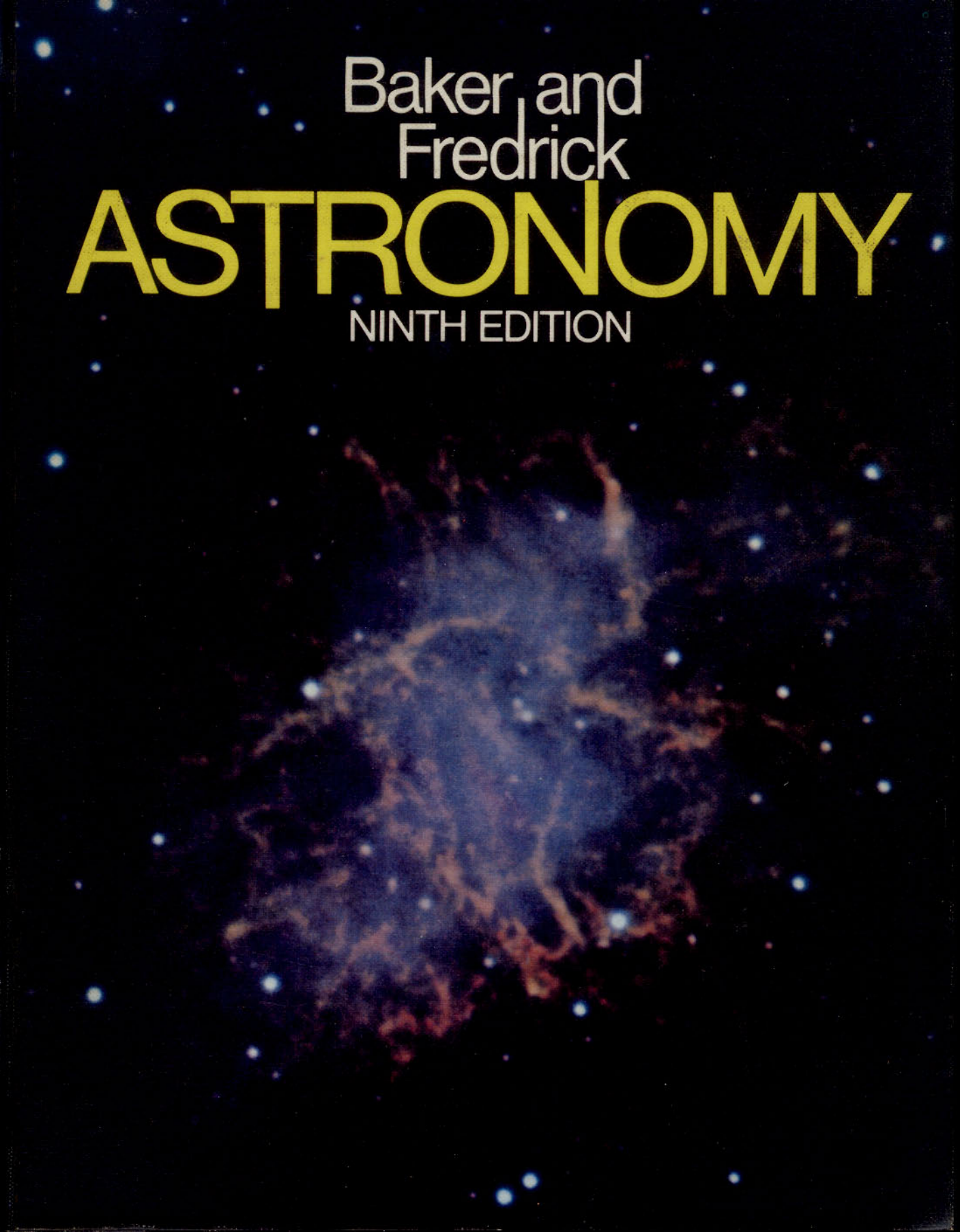
NINTH EDITION

VNR
VAN NOSTRAND REINHOLD

Baker and Fredrick

ASTRONOMY

NINTH EDITION



Robert H. Baker
Late Professor of Astronomy
at the University of Illinois

Laurence W. Fredrick
Director, Leander McCormick Observatory,
University of Virginia

ASTRONOMY

NINTH EDITION

VAN NOSTRAND REINHOLD COMPANY
New York | Cincinnati | Toronto | London | Melbourne

Van Nostrand Reinhold Company Regional Offices:
New York Cincinnati Chicago Millbrae Dallas

Van Nostrand Reinhold Company International Offices:
London Toronto Melbourne

Library of Congress Catalog Card Number: 74-127649

Manufactured in the United States of America

Published by Van Nostrand Reinhold Company
450 West 33rd Street, New York, N. Y. 10001

Published simultaneously in Canada by
Van Nostrand Reinhold Ltd.

10 9 8 7 6 5 4 3 2 1

Cover and Book Design—Saul Schnurman
Illustrations—Bert Schneider

PREFACE TO THE NINTH EDITION

This edition of Astronomy represents a considerable revision of the previous editions. An effort has been made to maintain the factual style used in earlier editions and the general scope and arrangement of the material remain as before. This is a textbook for two-semester introductory courses in astronomy at the college level and assumes very little prior background on the part of the student.

This edition contains a glossary for the first time. It is designed to aid the student with terms the instructor might use and not explain because they are so second nature to astronomers. The student is encouraged to expand the glossary by consulting additional sources. To assist the student in finding additional reading we continue to include the names of astronomers which may be used as key words in a library search. The references listed at the end of each chapter are divided into two sections, the first section contains books considered to be comparable to or simpler than the text itself, the second section contains books that range from slightly more advanced to considerably more advanced than the text. A basic reference shelf is listed after chapter 19. The entire book has been converted to the metric system, except where familiar usage overwhelmingly dictates to the contrary, e.g. the 200-inch telescope remains the 200-inch telescope. In dealing with abbreviations, units, etc., the Style Manual of the American Institute of Physics has been used.

The list of significant discoveries and improvements since the last edition is almost staggering even to astronomers. The discovery and identification of the quasi-stellar objects was so new during the last revision that it was not seen fit to mention them as such. In this edition we treat them although they remain somewhat mysterious objects. Since that edition, man has set foot upon the moon, his first feeble step into the cosmos is now behind him. A new class of objects, the pulsars have been discovered and present a whole new set of problems, X-ray stars have been discovered; one γ-ray source seems certain; the first large orbiting space telescope has been successful and has worked for more than a year revising the temperature scale for stars, etc. Long baseline interferometry is now an everyday procedure at radio wavelengths and the baselines are intercontinental; a 3°K isotropic background has been observed and it is claimed that gravity waves have been observed. Objects having masses only slightly greater than that of Jupiter have been discovered orbiting other stars.

All the achievements and the discoveries mentioned above only skim the cream off the top of the list. We should mention the discovery of a new moon of Saturn, closeup pictures of Mars and Venus and the dis-

covery of organic molecules in interstellar space. It is certain that equally exciting advances will occur during this edition.

We are indebted to an extremely long list of colleagues who have supplied pictures, diagrams, material in advance of publication, criticisms, corrections, suggestions, etc. I hope that the reader will note the credit lines on the illustrations as that is the only way we can call your attention to the people who have helped us. I would, however, like to acknowledge the assistance of Catharine Garmany and Haywood Smith in collating the revised portions, the illustrations and the problems. I would be derelict not to inform you that my wife Frances has done all the typing.

Laurence W. Fredrick

CHARLOTTESVILLE, VIRGINIA
JANUARY 1971

CONTENTS

INTRODUCTION

Astronomy, the "science of the stars," is concerned with the physical universe. This science deals with planets and their satellites, including the earth and moon, with comets and meteors, with the sun, the stars and clusters of stars, with the interstellar gas and dust, with the system of the Milky Way and the other galaxies that lie beyond the Milky Way.

The most comprehensive of the sciences, astronomy is also regarded as the oldest of all. People of ancient times were attentive watchers of the skies. They were attracted by the splendor of the heavens, as we are today, and by its mystery that entered into their religions and mythologies. Astrology, the pseudo-science which held that the destinies of nations and individuals were revealed by the stars, provided at times another incentive for attention to the heavens.

An additional incentive to the early cultivation of astronomy was its usefulness in relation to ordinary pursuits. The daily rotation of the heavens furnished means of telling time. The cycle of the moon's phases and the westward march of the constellations through the year were convenient for calendar purposes. The pole star in the north served as a faithful guide to the traveler on land and sea. These are some of the ways in which the heavens have been useful to man from the earliest civilizations to the present.

The value of this science, however, is not measured mainly in terms of economic applications. Astronomy is concerned primarily with an aspiration of mankind, which is fully as impelling as the quest for survival and material welfare, namely, the desire to know about the universe around us and our relation to it. The importance of this service is demonstrated by the widespread public interest in astronomy and by the generous financial support that has promoted the construction and operation of large telescopes in increasing numbers. Nowhere in the college curricula can the value of learning for its own sake be more convincingly presented than in the introductory courses in astronomy.

It is the purpose of astronomy to furnish a description of the physical universe, in which the features and relationships of its various parts are clearly shown. At present the picture is incomplete. Doubtless it will always remain incomplete, subject to improvements in the light of new explorations and viewpoints. The advancing years will bring added grandeur and significance to the view of the universe, as they have in the past.

The Sphere of the Stars. As early as the 6th century B.C., Greek scholars regarded the earth as a globe standing motionless in the center of the

universe. The boundary of their universe was a hollow globe, having the stars set on its inner surface. This celestial sphere was supported on an inclined axis through the earth, on which the sphere rotated daily, causing the stars to rise and set. Within the sphere seven celestial bodies moved around the earth; they were the sun, the moon, and the five bright planets.

This view of the universe remained practically unchanged for more than 2000 years thereafter. The chief problem of astronomy was to account for the observed motions of the seven wandering bodies. The outstanding solution of the problem, on the basis of the central, motionless earth, was the Ptolemaic system.

Copernicus, in the 16th century, proposed the theory that the planets revolve around the sun rather than the earth and that the earth is simply one of the planets. The rising and setting of the stars were now ascribed to the daily rotation of the earth. The new theory placed the sun and its family of planets sharply apart from the stars. With its gradual acceptance, the stars came to be regarded as remote suns at different distances from us and in motion in various directions. The ancient sphere of the stars remained only as a convention; and the way was prepared for explorations into the star fields, which have led to the more comprehensive view of the universe that we have today.

The Solar System. The earth is one of a number of relatively small planets that revolve around the sun; the planets are accompanied by smaller bodies, the satellites, of which the moon is an example. These are dark globes, shining only as they reflect the sunlight. The nine prinicpal planets, including the earth, have average distances from the sun ranging from 0.4 to 40 times the earth's distance. Thousands of smaller planets, the asteroids, describe their orbits mainly in the middle distances. Comets and meteroids also revolve around the sun; their orbits are usually more elongated than are those of the planets, and they often extend to greater distances from the sun.

These bodies together comprise the solar system, the only known system of its kind, although others may well exist. A similar planetary system surrounding the very nearest star could not be discerned with the largest telescope. Likewise, the telescopic view of our own system from the nearest star would show only the sun, then having the appearance of a bright star.

The Stars. The sun is one of the multitude of stars, representing a fair average of the general run. It is a globe of intensely hot gas, 1,392,000 kilometers in diameter, and a third of a million times as massive as the earth. Some stars are much larger than the sun and some others are much smaller. Blue stars are hotter than the sun, which is a yellow star. Red stars are cooler; but all are hot as compared with ordinary standards, and they

are radiating great amounts of energy. The stars are the power plants and building blocks of the universe.

Vast spaces intervene between the stars. If the size of the sun is represented by one of the periods on this page, the sun's nearest neighbor among the stars, the double star Alpha Centauri, would be shown on this scale by two dots 16 kilometers away. The actual distance exceeds four light years; that is, a ray of light, having a speed of 299,793 kilometers a second, spends more than four years in traveling from that star to the sun. This is a sample of the spacing of stars in the sun's neighborhood.

The interstellar spaces are not perfectly empty. In our vicinity and in many other regions they contain much gas and dust. Clouds of this material made luminous by nearby stars constitute the bright nebulae. The dust clouds can be detected by their dimming and reddening of stars behind them; they are responsible for the dark rifts that cause most of the variety in the Milky Way.

Our Galaxy is the assemblage of 100,000 million stars to which the sun belongs. Its most striking feature, as we view it from inside, is the glowing band of the Milky Way. This system consists in the main of a spheroidal central stellar region surrounded by a flat disk of stars 30,000 parsecs in diameter. Imbedded in this disk are spiral arms, containing stars, gas, and dust. The sun is near the principal plane of the Galaxy; its distance is about 10,000 parsecs from the center, which is situated in the direction of the constellation Sagittarius. The sun is included in a spiral arm.

As would be inferred from its flattened form, the Galaxy is rotating around an axis through its center. In the rotation the sun is now moving swiftly toward the direction of Cygnus. The period of the rotation in the sun's neighborhood is of the order of 200 million years. The more nearly spherical and more slowly rotating halo of the Galaxy includes the globular clusters.

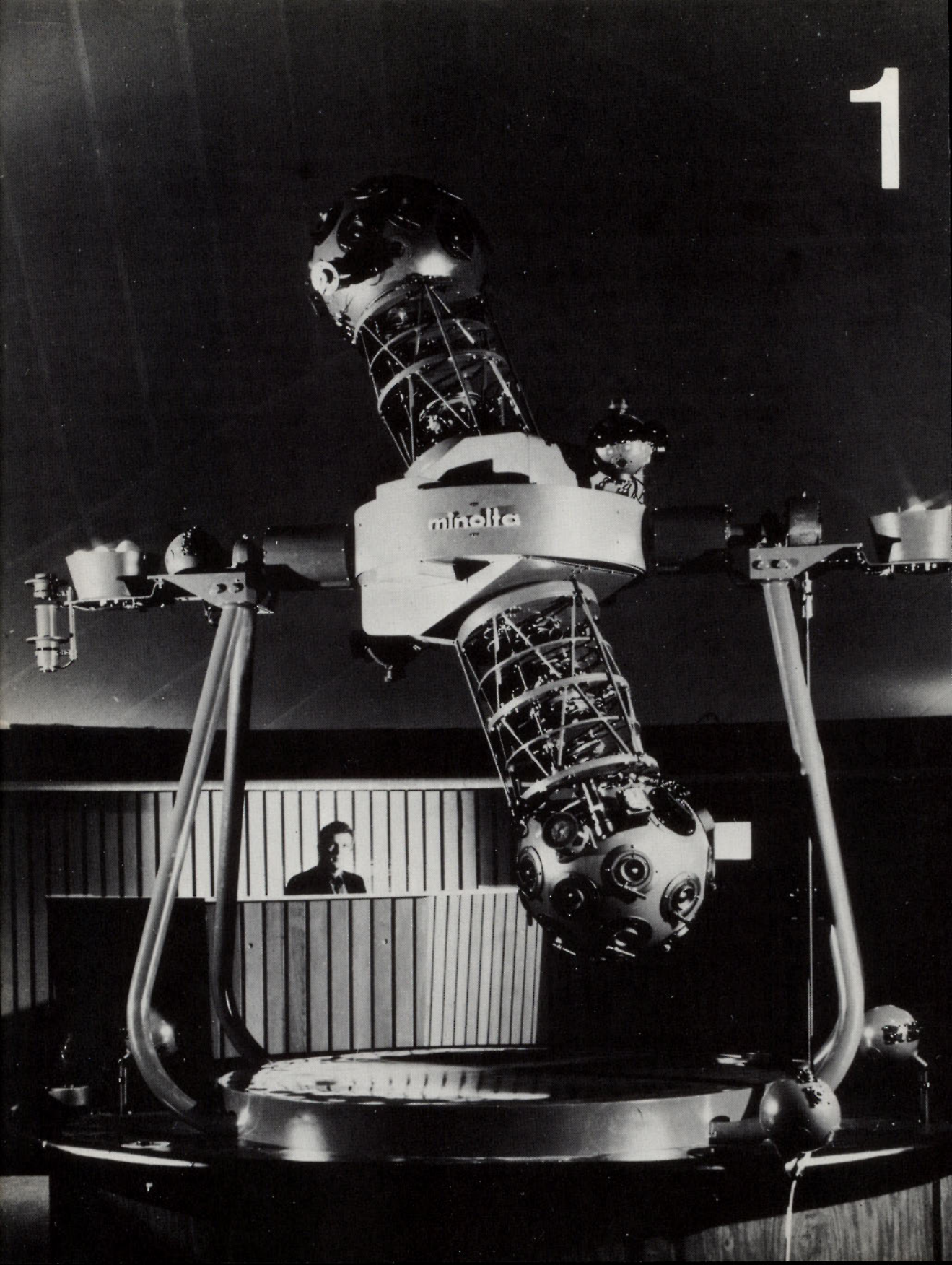
The Galaxies. Our spiral galaxy is one of the major building blocks of the universe. Millions of other galaxies extend into space as far as the largest telescopes can explore, and many of these are also spirals. They are retreating from us, and their speeds of recession increase as their distances are greater. This is the basis for the spectacular theories of the expanding universe.

Astronomy is a physical science closely related to the others. Its interests range from the structure and transmutation of the atom to the constitution and evolution of the universe. Its large optical telescopes are cameras for recording the proceedings in the laboratory of space, where quantities of material, extremes of conditions, and lengths of time involved in the ex-

periments transcend those of operations in the terrestrial physical and chemical laboratories. The telescopes also have important uses in collecting the radiations of celestial bodies and in funneling them into physical apparatus, such as the spectroscope and photoelectric cell.

Radio telescopes of increasing effectiveness are recording cosmic radiations of longer wavelength, extending the inquiries into the daytime and through cloudy skies. They inform us of features of the universe that may be entirely concealed from optical telescopes.

ASTRONOMY


1
minolta

ASPECTS OF THE SKY

THE CELESTIAL SPHERE; ITS APPARENT DAILY ROTATION — DIURNAL CIRCLES OF THE STARS — THE SUN'S APPARENT ANNUAL PATH — THE CONSTELLATIONS

At the beginning of the study of astronomy it is convenient to regard the earth as a sphere situated at the center of a vast spherical shell on which the stars are set. The stars may accordingly be represented on the surface of a globe and their positions may be referred to circles on the globe, just as the positions of towns and ships are denoted on the globe of the earth. It is in order at first to recall the circles that are employed on the earth's surface and then proceed to other coordinate systems.

medium-sized planetarium projector, with a capacity to repro-
uce faithfully the entire celestial sphere and the apparent
ıotion of all celestial bodies, and also the trajectory of man-
ıade spacecraft. *(Courtesy of the Minolta Corporation).*

THE TERRESTRIAL SPHERE

The earth, regarded as a sphere, has an imaginary coordinate system placed upon it for convenience in locating places and navigating between two points by ship or air. The coordinates are called longitude and latitude.

1.1 Circles of the Terrestrial Sphere

The earth rotates daily from west to east on an axis joining the north and south poles. Because of its rotation the earth is somewhat flattened at the poles and bulged at the equator. For the present purpose the flattening and also the surface irregularities are neglected. If the earth is regarded as a sphere, any plane passing through its center cuts the surface in a *great circle.* A plane through the earth but not through the center cuts the surface in a *small circle.*

The *equator* is the great circle of the terrestrial sphere halfway between the north and south poles and therefore 90° from each. *Meridians* are halves of great circles joining the poles and are therefore perpendicular to the equator. The *meridian of Greenwich,* which passes through the original site of the Royal Greenwich Observatory in England, is taken as the *prime meridian* for reckoning longitude. *Parallels of latitude* are small circles parallel to the equator, which diminish in size with increasing distance from the equator.

1.2 Longitude and Latitude

The position of a point on the earth's surface is denoted by the longitude and latitude of the point. The *longitude* is measured in degrees along the equator east or west from the prime meridian to the meridian through the point. Its value ranges from 0° to 180° either way, and the direction is indicated by the abbreviation E. or W., or by the minus or plus sign. If the longitude is 60° W., the point is somewhere on the meridian 60° west from the Greenwich meridian.

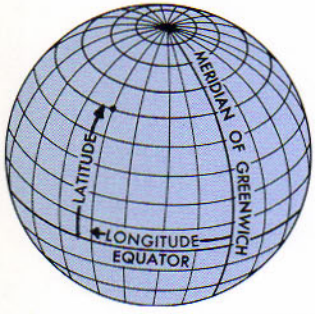


Figure 1.2
Circles of the terrestrial sphere. The position of a point on the earth is denoted by its longitude and latitude.

The *latitude* of a point is its distance in degrees north or south from the equator, measured along the meridian through the point. Its value ranges from 0° at the equator to 90° at the poles, and the direction is indicated by the abbreviation N. or S., or by the plus or minus sign. If the latitude is 50° N., the point is somewhere on the parallel 50° north from the equator. When the longitude is also given, the position is uniquely defined. As an example, the Yerkes Observatory is in longitude 88° 33′ W. and latitude 42° 34′ N.

Positions of stars on the apparent globe of the sky are similarly denoted. Although a single system of circles based on the equator suffices for the

earth, four different systems are required in the sky for the various purposes of astronomy. The three to be described in this chapter are based on the horizon, celestial equator, and ecliptic. The fourth system, based on the galactic equator will be discussed in chapter 17.

THE CELESTIAL SPHERE: ITS APPARENT ROTATION

The stars, apparently at infinity, appear to lie on a sphere. Coordinate systems can be imagined to be upon this sphere, which seems to rotate around the earth a little more than once in a day. The coordinates on the celestial sphere, analogous to those on the terrestrial sphere, are called right ascension and declination.

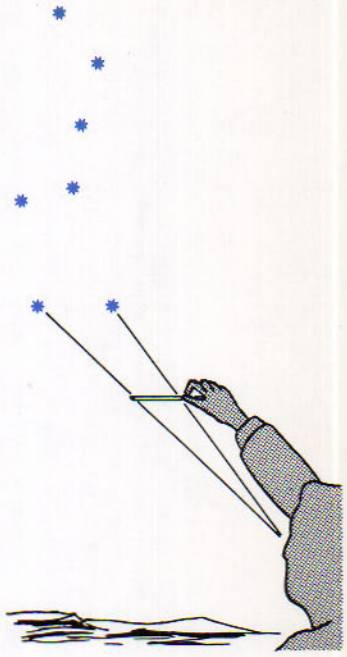


Figure 1.3
The distance between the dipper's pointers is about 5 degrees.

1.3 The Celestial Sphere

The Celestial Sphere is the conventional representation of the sky as a spherical shell on which the celestial bodies appear projected. Evidently of very great size, it has the properties of an infinite sphere. Its center may be anywhere at all and is often taken as the observer's position. Parallel lines are directed toward the same point of the sphere, just as the rails of a track seem to converge in the distance. Each star has its *apparent place* on the sphere, where it appears to be; this specifies only the star's direction. The *apparent distance* between two stars is their angular separation on the celestial sphere.

Apparent places and distances are always expressed in angular measure, such as degrees, minutes, and seconds of arc. For estimating angular distances in the sky it is useful to remember that the apparent diameters of the sun and moon are about half a degree. The pointer stars of the Great Dipper are somewhat more than 5° apart (Fig. 1.3).

1.4 Horizon and Celestial Meridian

The *zenith* is the point on the celestial sphere that is vertically overhead. The *nadir* is the opposite point, vertically underfoot. These points are located by sighting along a plumb line, or vertical line.

The *celestial horizon* is the great circle on the celestial sphere halfway between the zenith and nadir and, therefore, 90° from each. This is the

horizon of astronomy as distinguished from the visible horizon, the frequently irregular line where the earth and sky seem to meet. The horizon is an example of the circles that are imagined to be on the celestial sphere for the purpose of denoting the places of celestial bodies, just as circles such as the equator are imagined on the terrestrial sphere. The horizon is the basis of the horizon system of coordinates.

Vertical circles are great circles that pass through the zenith and nadir and are, therefore, perpendicular to the horizon. One of these is the observer's *celestial meridian,* the vertical circle that crosses his horizon at the north and south points. Another is the *prime vertical,* which crosses the horizon at the east and west points.

Celestial circles and coordinates may be somewhat confusing to the reader at first, because they often have unfamiliar names and also because they are represented in two dimensions in the diagrams. The use of a celestial globe or of a blank globe on which the circles can be drawn is likely to contribute to a clearer understanding of features of the sky described in this and other chapters. It should preferably be a globe that can be rotated and that has a movable meridian, so that the direction of the rotation axis can be varied.

1.5 Azimuth and Altitude

The position of a star is denoted in the horizon system by its azimuth and altitude. The *azimuth* of a star is the angular distance measured from the north point toward the east along the horizon to the vertical circle of the star. It is measured completely around the horizon from 0° to 360°. Azimuth is sometimes reckoned from the south instead of the north and occasionally from the north eastward through 180° and then from the south westward through 180°.

The *altitude* of a star is its angular distance from the horizon. Altitude is measured along the vertical circle through the star, having values from

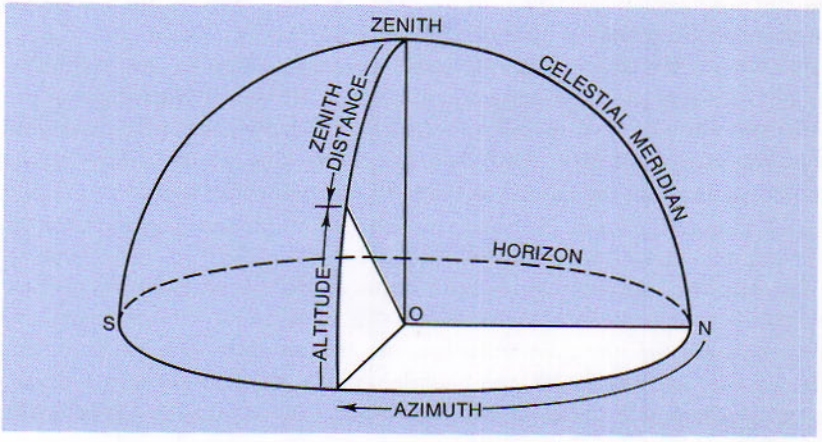


Figure 1.5
Location of a star in the horizon system by azimuth and altitude.

0° at the horizon to 90° at the zenith. Its complement, *zenith distance,* is measured downward from the zenith along the vertical circle. When its azimuth and altitude are given, the star's place in the sky is known, as the following examples illustrate.

1. Point to a star in azimuth 90° and altitude 45°.
   Answer: The star is directly in the east, halfway from the horizon to the zenith.
2. Point to a star in azimuth 180° and altitude 30°.
   Answer: The star is directly in the south, one third of the way from the horizon to the zenith.
3. State the azimuth and altitude of a star that is exactly in the southwest and two thirds of the way from the horizon to the zenith.
   Answer: Azimuth 225°, altitude 60°.

The simplicity of the horizon system recommends it for various purposes in astronomy, navigation, and surveying. It is easy to visualize these circles and coordinates in the sky. The navigator's sextant or octant and the engineer's transit operate in this system. Azimuths and altitudes of the celestial bodies are always changing, because of the apparent daily rotation of the celestial sphere; and at the same instant they change with the observer's position on the earth. This necessitates the use of other coordinate systems as well.

1.6 Apparent Daily Rotation of the Celestial Sphere

The westward movement of the sun across the sky, which causes the sun to appear to rise and set, is an example of the motion in which all the celestial bodies share. It is as though the celestial sphere were rotating daily around the earth from east to west. This apparent daily rotation, or *diurnal motion,* of the heavens is an effect of the earth's rotation on its axis from west to east.

Every star describes its *diurnal circle* around the sky daily. All diurnal circles of the stars are parallel and are described in the same period of time; but those of the sun, moon, and planets, which change their places among the stars, are not quite parallel and have somewhat different periods. The rapidity with which a star proceeds along its diurnal circle depends on the size of the circle that is described. The motion is fastest for stars that rise exactly in the east; it becomes progressively slower as the rising is farther from the east point, and vanishes at two opposite points in the sky around which the diurnal circles are described.

1.7 The Celestial Poles

The two points on the celestial sphere having no diurnal motion are the *north and south celestial poles.* They are the points toward which the earth's axis is directed. For observers in the northern hemisphere, the north

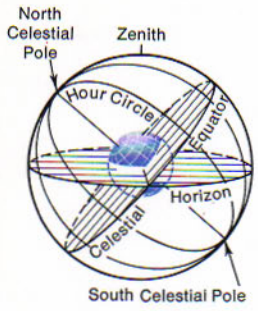


Figure 1.8
The celestial equator is in the plane of the earth's equator. It crosses the horizon at the east and west cardinal points at an inclination equal to the complement of the observer's latitude.

celestial pole is situated vertically above the north point of the horizon; its place is marked approximately by Polaris, the *pole star,* or *north star,* at the end of the handle of the Little Dipper. Polaris is now about 1° from the pole, or twice the apparent diameter of the moon.

The south celestial pole is depressed below the south horizon as much as the north pole is elevated in the northern sky. Its place is not marked by any bright star. This is the elevated pole for observers in the southern hemisphere.

It is interesting to photograph the diurnal trails of stars around the pole with an ordinary camera. Set the camera so that it points toward the pole star and expose a film for several hours on a clear evening, using the full aperture of the lens and having the focus adjusted for infinity. The trails in the picture are arcs of diurnal circles, which have the celestial pole as their common center. Increasing the exposure makes the trails longer, but shows no more stars.

1.8 Celestial Equator; Hour Circles

The *celestial equator* is the great circle of the celestial sphere halfway between the north and south celestial poles. It is in the plane of the earth's equator and is the largest of the diurnal circles. For a particular place on the earth, the celestial equator has nearly the same position in the sky throughout the day and year; it is traced approximately by the sun's diurnal motion on March 21 or September 23.

Hour circles pass through the celestial poles and are, therefore, at right angles to the celestial equator. They are half circles from pole to pole, like meridians of the terrestrial sphere; and they may be considered as fixed either on the turning celestial sphere or with respect to the observer, like his celestial meridian.

1.9 Directions in the Equator System

In the system of circles based on the celestial equator, north is the direction along an hour circle toward the north celestial pole. South is the opposite direction. West is the direction of the diurnal motion, which is parallel to the celestial equator. With these definitions in mind there will be no confusion about directions in the sky, even in the vicinity of the pole. As one faces north, the stars circle daily in the counterclockwise direction. From a star directly above the pole, north is downward and west is toward the left; from a star below the pole, north is upward and west is toward the right. It should be understood here that the equatorial system is valid only for terrestrial observers.

1.10 Right Ascension and Declination

The *right ascension* of a star is the angular distance measured in hours, or degrees, from the vernal equinox eastward along the celestial equator to the hour circle of the star. The vernal equinox is the point where the sun's center crosses the celestial equator at the beginning of spring; the

hour circle through this point serves the same purpose as the prime meridian of the earth. Unlike terrestrial longitude, which is measured both east and west from the prime meridian, right ascension is always measured toward the east, from 0^h or 0° at the vernal equinox to 24^h or 360° at the equinox again.

The *declination* of a star is its angular distance north or south from the celestial equator. Declination is measured along the hour circle of the star, its value ranging from 0° at the equator to 90° at the poles. If the star is north of the equator, the sign of the declination is plus; if the star is south of the equator, the sign is minus. The places of the celestial bodies are generally denoted by their right ascensions and declinations in maps and catalogs.

As an example, the right ascension of the bright star Sirius is $6^h\ 44^m$, and the declination is −16° 43′; the star is therefore $6^h\ 44^m$, or 101°, east of the vernal equinox and 16° 43′ south of the celestial equator. The following relations can be employed in changing from hours to degrees of right ascension, or conversely:

$$1^h = 15^\circ \qquad 1^\circ = 4^m$$
$$1^m = 15' \qquad 1' = 4^s$$
$$1^s = 15'' \qquad 1'' = .067^s$$

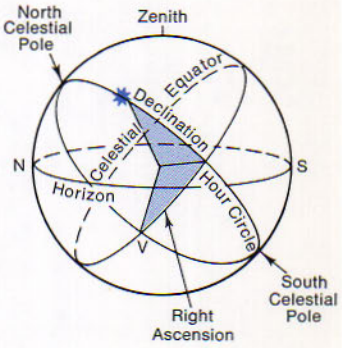


Figure 1.10
Right ascension and declination. Right ascension is measured eastward from the vernal equinox, *V*, along the celestial equator. Declination is measured north or south from the equator.

1.11 Hour Angle

The place of a star is also denoted in the equator system by its hour angle and declination. In this case the observer's celestial meridian is the circle of reference and is considered to have two branches. The *upper branch* of the celestial meridian is the half between the celestial poles that includes the observer's zenith; the *lower branch* is the opposite half that includes the nadir. A star is at *upper transit* when it crosses the upper branch of the celestial meridian, and at *lower transit* when it crosses the lower branch.

The *local hour angle* of a star is the angular distance measured along the celestial equator westward from the upper branch of the observer's celestial meridian to the hour circle of the star; or it is the corresponding angle at the celestial pole. The value of the hour angle ranges from 0° to 360°. Unlike the right ascension of a star, which remains nearly unchanged during the day, the hour angle increases at the rate of about 15° an hour, and at the same instant has different values in different longitudes.

The *Greenwich hour angle* of a star is its local hour angle as observed at the meridian of Greenwich. Greenwich hour angles of celestial bodies are tabulated in nautical and air almanacs at convenient intervals of the day throughout the year.

DIURNAL CIRCLES OF THE STARS

The stars seem to move across the sky daily, reflecting the earth's rotation. The apparent motion and its angle with the horizon depend on the latitude of the observer's position.

1.12 The Observer's Latitude Equals the Altitude of the North Celestial Pole

The *astronomical latitude* of a place on the earth is defined as the angle between the vertical line at the place and the plane of the earth's equator. Evidently this is also the angle between the directions of the zenith and celestial equator. As is seen in Fig. 1.12A, this angle equals the altitude of the north celestial pole as well as the declination of the zenith at the place. Here we have the basic rule for all determinations of latitude by sights on celestial objects.

Where irregularities of the earth affect the direction of the vertical line, the astronomical latitude requires a slight correction to obtain the *geographical latitude.* The correction rarely exceeds 30″ and is usually much smaller.

When the latitude of a place is given, the altitude of the celestial pole

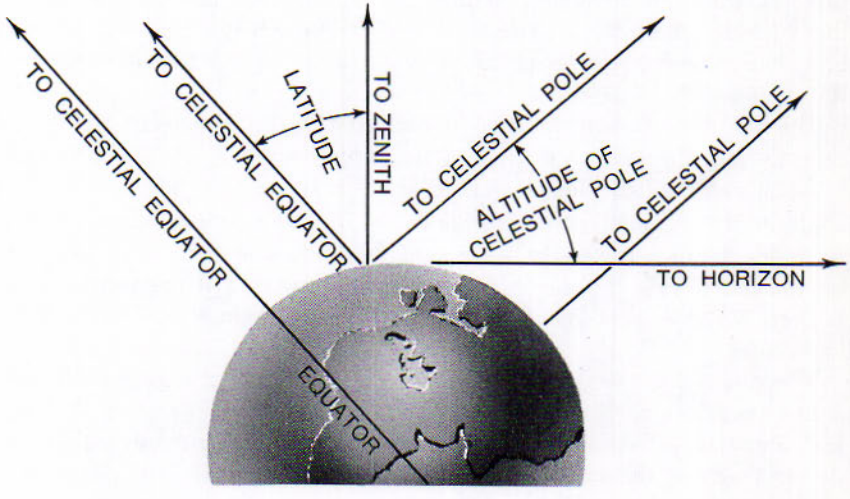


Figure 1.12
The latitude of a place on the earth equals the altitude of the north celestial pole at that place. Astronomical latitude is defined as the angle between the vertical line and the equator plane. This angle equals the altitude of the north celestial pole, because both are complements of the angle between the zenith and the celestial pole. It also equals the declination of the zenith.

is an equal number of degrees, according to the rule. Thus the positions of the celestial poles and of the celestial equator midway between them become known relative to the horizon of the place. Because the diurnal circles of the stars are parallel to the equator, we may now consider how these circles are related to the horizon for observers at different places on the earth.

Viewed from the north pole, latitude 90° N., the north celestial pole is in the zenith and the celestial equator coincides with the horizon. Here the diurnal circles are parallel to the horizon. Stars north of the celestial equator never set, and those in the south celestial hemisphere are never seen. The sun, moon, and planets, which change their places among the stars, come into view when they move north across the equator and set when they cross to the south. At the south pole, of course, the south celestial pole is in the zenith and everything is reversed.

It will be noted later that some statements in this chapter require slight modification because of refraction of starlight in the earth's atmosphere.

1.13
At the Pole, Diurnal Circles Are Parallel to the Horizon

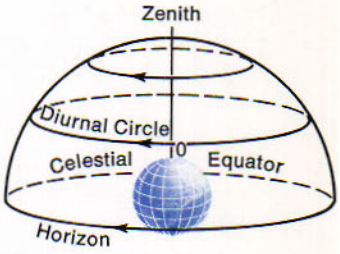


Figure 1.13
Diurnal circles observed at the north (or south) pole are parallel to the horizon.

Viewed from the equator, latitude 0°, the celestial poles are on the horizon at its north and south points. The celestial equator crosses at right angles to the horizon at its east and west points and passes directly overhead. All diurnal circles, since they are parallel to the equator, are also perpendicular to the horizon and are bisected by it. Thus every star is above the horizon for about 12 hours and is below for the same interval; the daily duration of sunlight is about 12 hours throughout the year.

It is to be noted that places on the equator are the only ones from which the celestial sphere can be seen from pole to pole, so that all parts of the heavens are brought into view by the apparent daily rotation.

1.14
At the Equator, Diurnal Circles Are Perpendicular to the Horizon

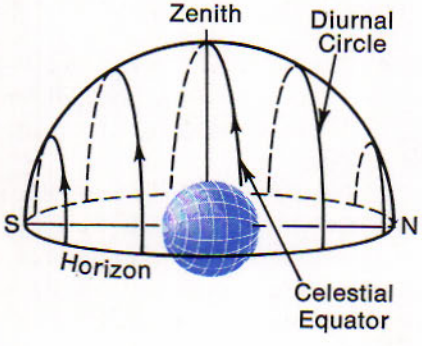


Figure 1.14
Diurnal circles observed at the equator are perpendicular to the horizon.

1.15 Elsewhere, Diurnal Circles Are Oblique

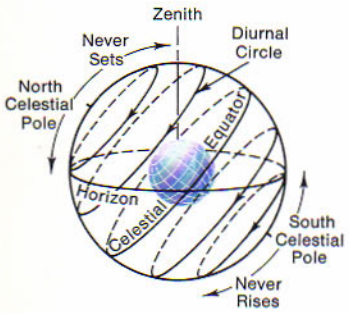


Figure 1.15
Diurnal circles observed in latitude 40° N. are oblique.

From points of observation between the pole and the equator, the north celestial pole is elevated a number of degrees equal to the latitude of the place, and the south celestial pole is depressed the same amount. Although the celestial equator still crosses the horizon at the east and west points, it no longer passes through the zenith, but inclines toward the south, in the northern hemisphere, by an angle equal to the latitude. Thus the diurnal circles of the stars cross the horizon obliquely.

The celestial equator is the only one of these circles that is bisected by the horizon. Northward, the visible portions of the diurnal circles become progressively greater, until the entire circles are in view; southward from the celestial equator they diminish in size, until they are wholly out of sight. The changing daily duration of sunlight from summer to winter serves as an example.

In this oblique aspect of the diurnal motion with respect to the horizon, the celestial sphere is conveniently divided into three parts: (1) A cap around the elevated celestial pole, having its radius equal to the latitude of the place, contains the stars that are always above the horizon; (2) a cap of the same size around the depressed pole contains the stars that never come into view; (3) a band of the heavens symmetrical with the celestial equator contains the stars that rise and set. In latitude 40° N., for example, the two caps are 40° in radius, and all stars within 50° of the celestial equator rise and set.

As one travels south from 40° N, the circumpolar caps grow smaller and finally disappear when the equator is reached. As one travels north, the circumpolar caps increase in radius, until they join when the pole is reached. Here none of the stars rises or sets.

1.16 Circumpolar Stars

If a star is nearer the celestial pole than the pole itself is to the horizon, the star does not cross the horizon; it is a *circumpolar star*. Consequently, for an observer in the northern hemisphere, a star never sets if its north polar distance (90° minus its declination) is less than the observer's latitude; and it never rises if its south polar distance is less than the latitude. The following examples illustrate the rule:

1. The Southern Cross, Decl. −60°, never rises in latitude 40° N., because its south polar distance of 30° is less than the latitude. It becomes visible south of latitude 30° N., in Florida and southern Texas.
2. The bowl of the Great Dipper, Decl. +58°, never sets in latitude 40° N., because its north polar distance of 32° is less than the latitude. Under the celestial pole its center is still 8° above the horizon. It rises and sets south of latitude 32° N.
3. The sun on June 22, Decl. +23½°, rises and sets in latitude 40° N.,

because its north polar distance of 66½° is not less than the latitude. North of about latitude 66½° N. the sun is circumpolar on that date.

The *midnight sun* is an example of a circumpolar object. The sun may be seen at midnight on June 22 about as far south as the arctic circle. Farther north it remains above the horizon for a longer period, and at the north pole it does not set for six months.

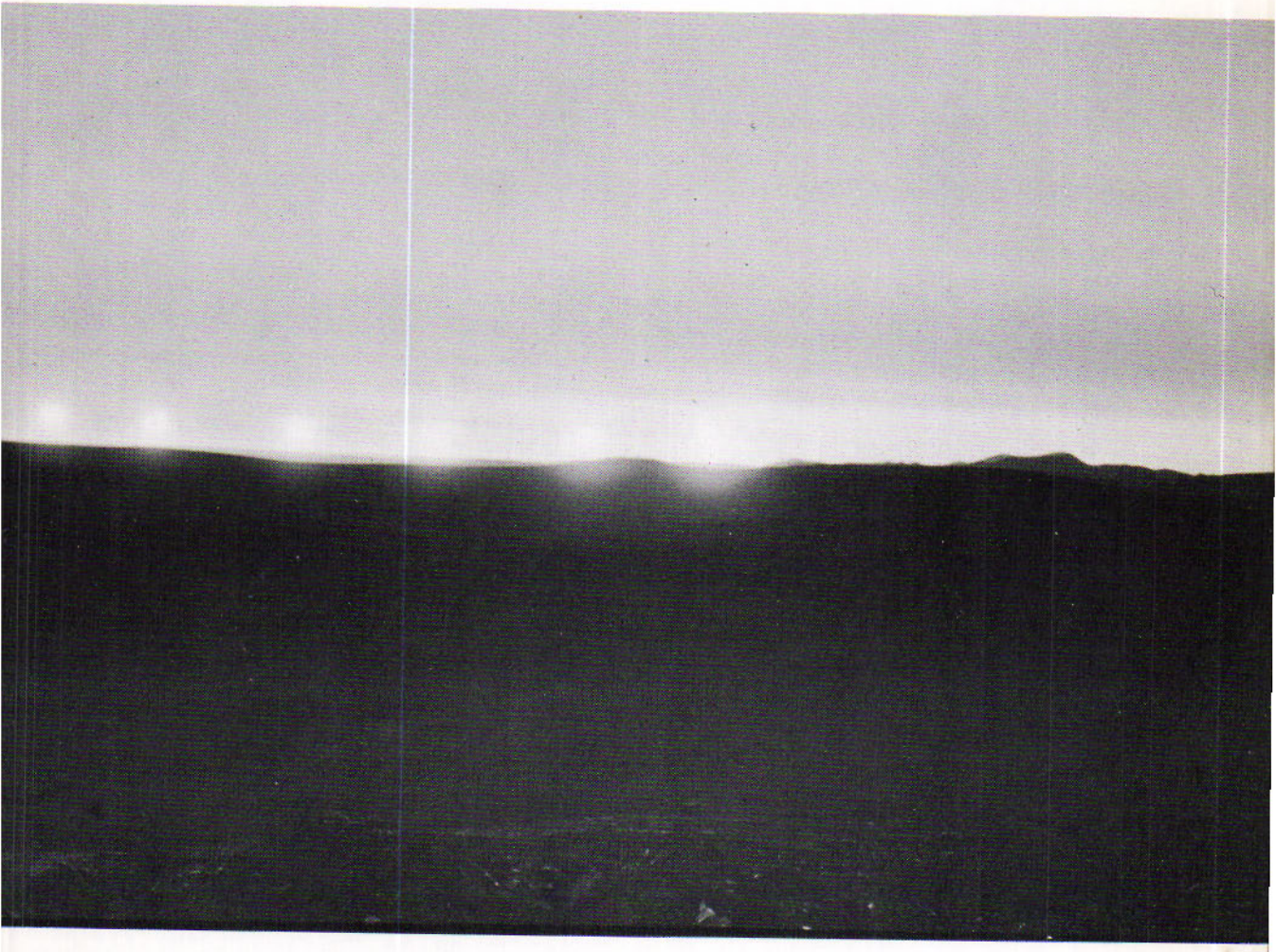


Figure 1.16
The midnight sun. Multiple-exposure photograph taken at 15-minute intervals, starting at 11:15 PM on June 19, 1967 at Eagle Summit, Alaska. *(Photograph by P. Sheridan.)*

THE SUN'S APPARENT ANNUAL PATH

The sun appears to move due to the earth's orbital motion around it. The tilt of the earth's axis on rotation causes the sun to appear to move up and down as well.

1.17 Westward Advance of the Constellations

In addition to their daily westward circling around us, the stars are a little farther west each evening than they were at the same time the evening before. A constellation steps forward gradually, night by night, until it has finally disappeared below the western horizon at that time. For example, the familiar oblong figure of Orion is seen rising in the east in the early evening in December. Late in the winter, Orion has moved into the south at the same hour of the night. As spring advances, it comes out in the western sky and sets soon after nightfall.

This steady westward march of the constellations with the changing seasons results from the sun's apparent eastward movement among the stars. If the stars were visible in the daytime, as they are in the sky of the planetarium, we could watch the sun's progress among them. We would observe that the sun is displaced toward the east about twice its diameter from day to day, and completely around the heavens in the course of a year. The sun's apparent annual movement around the celestial sphere is a consequence of the earth's annual revolution around the sun.

Not only does the sun move eastward with respect to the stars, but it oscillates to the north and south as well during the year. The sun's apparent path is inclined to the celestial equator.

1.18 The Ecliptic; Equinoxes and Solstices

The *ecliptic* is the apparent annual path of the sun's center on the celestial sphere. It is a great circle inclined 23½° to the celestial equator.

Four equidistant points on the ecliptic are the two *equinoxes*, where the circle intersects the celestial equator, and the two *solstices*, where it is farthest from the equator. The *vernal equinox* is the sun's position about March 21, when it crosses the celestial equator going north; the *autumnal equinox* is the sun's position about September 23, when it crosses on the way south. The *summer solstice* is the most northern point of the ecliptic, the sun's position about June 22; the *winter solstice* is the southernmost point, the sun's position about December 22. These dates vary slightly because of the plan of leap years.

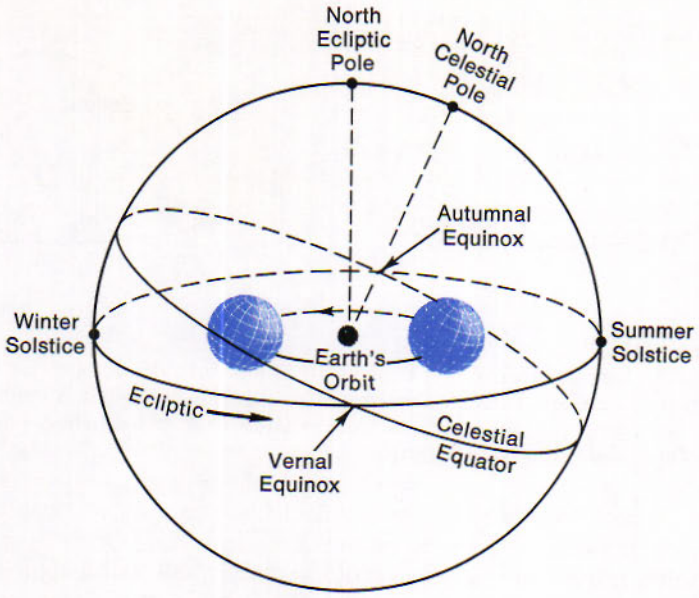


Figure 1.18
Relation of ecliptic and celestial equator. The inclination of the ecliptic to the celestial equator is the same (23½°) as the inclination of the earth's orbit to its equator.

The north and south *ecliptic poles* are the two points 90° from the ecliptic. They are 23½° from the celestial poles.

The relation between the ecliptic and celestial equator is explained in Fig. 1.18, in which the earth's orbit is viewed nearly edgewise. Because parallel lines meet in the distant sky, the celestial poles, toward which the earth's axis is directed, are not displaced by the earth's revolution around the sun; similarly the celestial equator is unaffected. Evidently the angle between the ecliptic and celestial equator is the same as the angle between the earth's orbit and equator. This inclination, or *obliquity*, of the ecliptic is 23° 27′; it is at present decreasing at the rate of 1′ in 128 years.

1.19 Relation Between Ecliptic and Horizon

The inclination of the celestial equator to the horizon of a particular place remains almost unaltered; this angle is the complement of the latitude of the place. Thus in latitude 40° N. the celestial equator is inclined 50° to the horizon. The ecliptic, however, takes different positions in the sky during the day.

Because the ecliptic is inclined 23½° to the celestial equator, its inclination to the horizon can differ as much as 23½°, either way, from that of the equator. It can be seen with the aid of a globe that the greatest and least angles between the ecliptic and the horizon in middle northern or southern latitudes occur when the equinoxes are rising or setting.

In latitude 40° N., when the vernal equinox is rising and the autumnal equinox is setting, as at sunset on September 23, the angle between the

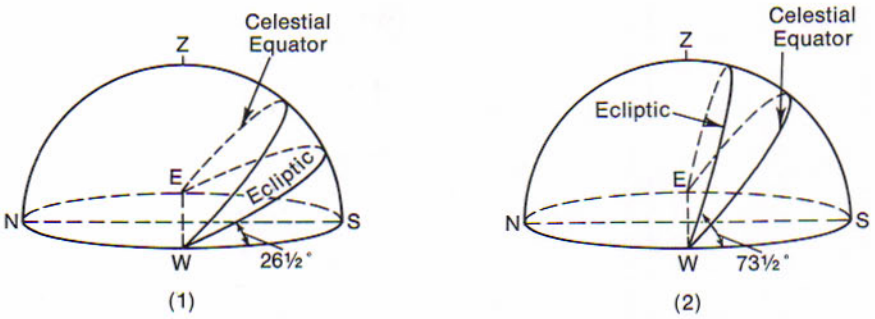


Figure 1.19
Relation between ecliptic and horizon. (1) The ecliptic is least inclined to the horizon in middle northern latitudes when the vernal equinox is rising and the autumnal equinox is setting. (2) The ecliptic is most inclined when the autumnal equinox is rising and the vernal equinox is setting.

ecliptic and horizon is $50° - 23½° = 26½°$. The visible half of the ecliptic lies below the celestial equator. When the autumnal equinox is rising and the vernal equinox is setting, as at sunset on March 21, the angle is $50° + 23½° = 73½°$. The visible half of the ecliptic is then above the celestial equator.

It will be noted later that the variation in the angle between the ecliptic and horizon is involved in the explanations of the harvest moon, the appearance of the planet Mercury as evening and morning star, and the favorable seasons for observing the zodiacal light.

1.20 Celestial Longitude and Latitude

The observations of early astronomers were confined for the most part to the sun, moon, and bright planets, which are never far from the ecliptic. It was accordingly the custom to denote the places of these objects with reference to the ecliptic by giving their celestial longitudes and latitudes. *Celestial longitude* is angular distance from the vernal equinox, measured eastward along the ecliptic to the circle through the object that is at right angles to the ecliptic. *Celestial latitude* is the angular distance of the object from the ecliptic, measured to the north or south along the perpendicular circle.

The earlier coordinates still find use in problems of planetary motions. They have been supplanted for most purposes by right ascension and declination, which are the counterparts of terrestrial longitude and latitude. The newer coordinates might well have been named celestial longitude and latitude instead, except that these names had already been appropriated.

THE CONSTELLATIONS

The stars form interesting patterns, which are well known to many people. There are dippers, crosses, and a variety of other figures easy to identify and to remember. In the original sense, the constellations *are these configurations of stars.*

1.21 The Primitive Constellations

Two thousand years ago, the Greeks recognized 48 constellations with which they associated the names and forms of heroes and animals of their mythology. The earliest nearly complete account of them that can be found in libraries today is contained in the *Phenomena*, written about 270 B.C. by the poet Aratus. In the writings of Hesiod, more than 500 years earlier, and in the Homeric epics the more conspicuous figures, such as Orion, the Pleiades, and the Great Bear, are mentioned familiarly. There are reasons for supposing that practically the whole scheme of primitive constellations was transmitted to the Greeks, having originated several thousand years before among the peoples of the Euphrates valley. The 48 original constellations, with certain changes made by the Greeks, are described in Ptolemy's *Almagest* (about A.D. 150), which specifies the positions of stars in the imagined creatures.

The ancient star-creatures have nothing to do with the science of astronomy. Their names, however, are still associated with striking groupings of stars, which attract attention now just as they did long ago.

1.22 Constellations as Regions of the Celestial Sphere

The original constellations did not cover the entire sky. Of the 1028 stars listed by Ptolemy, 10 per cent were "unformed," that is, not included within the 48 figures. Moreover, a large area of the celestial sphere in the south, which never rose above the horizons of the Greeks, was uncharted by them. In various star maps that appeared after the beginning of the 17th century, new configurations were added to fill the vacant spaces and received names not associated with mythology. At present, 88 constellations are recognized (Table 3, Appendix), of which 70 are visible at least in part from the latitude of New York in the course of a year.

For the purposes of astronomy, the constellations are regions of the celestial sphere set off by arbitrary boundary lines. These divisions are useful for describing the approximate positions of the stars and other celestial bodies. The statement that Vega is in the constellation Lyra serves the same purpose as the information that Cleveland is in Ohio. We know about where it can be found.

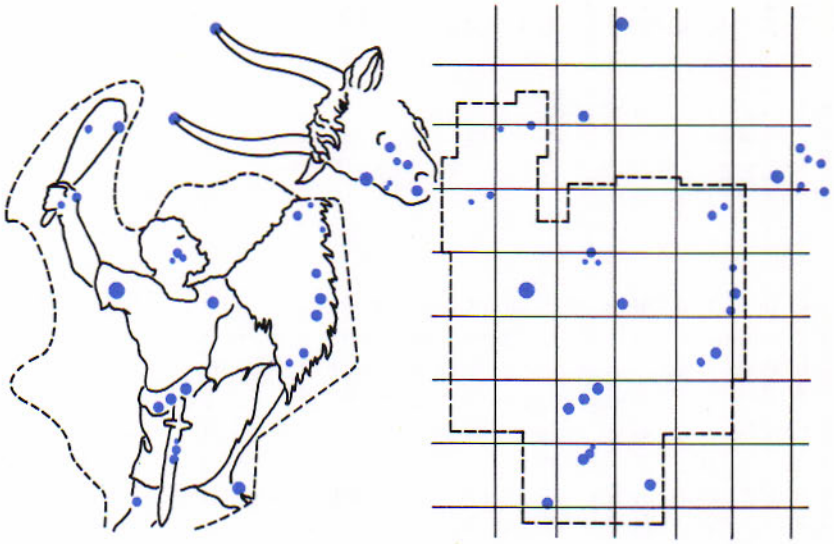


Figure 1.22
Old and new boundaries of Orion.

The boundaries of the majority of the constellations were formerly irregular. Revised by action of the International Astronomical Union, in 1928, the boundaries are now parts of circles parallel and perpendicular to the celestial equator. The boundary lines are not shown in the star maps of this book.

1.23 Names of Individual Stars

Fifty or more of the brighter stars are known to us by the names given them long ago. Some star names, such as Sirius and Capella, are of Greek and Latin origin; others, such as Vega, Rigel, and Aldebaran, are of Arabic derivation. The influence of the Arabs in the development of astronomy is indicated by the frequent appearance of their definite article *al* in the names of stars (Algol, Altair, etc.).

Some star names now regarded as personal were originally expressions giving the positions of stars in the imagined constellation characters. These descriptive terms, translated from Ptolemy's catalog into Arabic, degenerated later into single words. Examples are Betelgeuse (perhaps meaning armpit of the Central One), Fomalhaut (mouth of the Fish), Deneb (tail, of the Bird).

1.24 Designations of Stars by Letter and Number

The star maps of Bayer's *Uranometria* (1603) introduced the present plan of designating the brighter stars of each constellation by small letters of the Greek alphabet. In a general way, the stars are lettered in order of brightness, and the Roman alphabet is drawn upon for further letters. The full name of a star in the Bayer system is the letter followed by the possessive of the Latin name of the constellation. Thus α Tauri is

the brightest star in Taurus. When several stars in the constellation have nearly the same brightness, they are lettered in order of their positions in the mythological figure, beginning at the head. Thus the seven stars of the Great Dipper, which are not much different in brightness, are lettered in order of position.

Another plan, adopted in Flamsteed's *Historia Coelestis* (1729), in which the stars are numbered consecutively from west to east across a constellation, permits the designation of a greater number of stars. The star 61 Cygni is an example. In modern maps of the lucid stars it is usual to employ the Bayer letters as far as they go, giving the specific names of the brightest and most notable stars, and to designate some fainter stars by the Flamsteed numbers.

These are means of identifying the few thousand stars visible to the unaided eye. Stars are also often referred to by their running numbers in various catalogues. Probably the most widely used catalogue is the Bonner Durchmusterung, which is also a set of maps for the epoch 1855. Stars in this catalogue have the prefix BD. Many special purpose catalogues have developed for identifying objects such as double stars, stars with large motions, etc. Stars become more and more numerous as we go to fainter limits and the practice now is to give a finding chart if the object is fainter than the limits of the well known catalogues. Thus, the many interesting faint objects found by W. Luyten, H. Giclas and others are given numbers in their respective catalogues and finding charts as well. The finding charts may be actual photographs or very careful drawings.

Usually catalogues list their objects in order of increasing right ascension either for the whole sky or for specified strips of declination. Thus, the early radio source catalogues such as the third Cambridge catalogue (numbered sources preceded by the prefix 3C) use the former method. As the catalogues extend to fainter and fainter limits, more refined listings will have to be used and eventually the latter method will probably prevail.

1.25 Magnitudes of the Stars

It is easier to identify a star correctly when its brightness is known as well as its place in the sky. From Ptolemy's early catalogue to the modern catalogues and maps of the stars, it has been the custom to express the relative brightness of a star by stating its *magnitude.* For radio sources brightness is expressed in terms of flux units which are the amount of energy received from the source in terms of a specified unit energy. Magnitudes could have and eventually will have a similar positive identification. In fact, magnitudes do have such an identification but it is the result of tacit agreement and as such rarely expressed, much to the consternation and confusion of the engineer and physicist.

1 flux unit $= 10^{-26}$ watt/meter^2/Hz

At first, the stars were divided arbitrarily into six classes, or magnitudes, in order of diminishing brightness. About 20 of the brightest stars

were assigned to the 1st magnitude; Polaris and stars of the Great Dipper were representatives of the 2nd magnitude; and so on, until stars barely visible to the naked eye remained for the 6th magnitude.

With the invention of the telescope, permitting the observation of still fainter stars, the number of magnitudes was increased, while greater precision in measurement of brightness began to call for the use of decimals in denoting the magnitudes. Eventually, a factor slightly greater than 2.5 was adopted as the ratio of brightness corresponding to a difference of 1 magnitude. Thus a star of magnitude 3.0 is about 2½ times as bright as a star of magnitude 4.0.

The magnitudes assigned to the naked-eye stars by the early astronomers are not altered greatly by modern practice, except those of the very brightest stars. The original 1st-magnitude stars range so widely in brightness that the more brilliant ones have been promoted by the modern rule to brighter classes, and so to smaller numbers. The visual magnitude of the brightest star, Sirius, is −1.4; Canopus is −0.7; Arcturus, Vega, and Capella, the brightest stars of the north celestial hemisphere, are not far from magnitude 0; Spica is of magnitude +1.0. We will return to the question of stellar brightness in more detail in chapter 11.

In maps that follow, the brightness of stars is denoted to whole magnitudes by the symbols, the meanings of which are given in the key adjoining the circular maps. In the interest of simplicity, two stars of around the minus 1st magnitude are designated as of magnitude 0, and a few 5th-magnitude stars as of magnitude 4. Stars fainter than the 4th magnitude are not generally shown in these maps.

1.26 The North Circumpolar Map

Map 1 represents the appearance of the heavens to one facing north in middle northern latitudes. At the center is the north celestial pole, the altitude of which equals the observer's latitude. Hour circles radiating from the center are numbered with hours of their right ascensions. Parallels of declination appear as circles at intervals of 10°, from declination +90° at the center of the map to +50° at its circumference.

The names of the months around the circumference of the map facilitate its orientation to correspond with the sky at any time. If the map is turned so that the date of observation is uppermost, the vertical line through the center of the map represents the observer's celestial meridian at about 9 P.M., standard time. The constellations then have the same position in the map as they have in the northern sky at that hour.

To orient the map for a later hour, turn it counterclockwise through as many hours of right ascension as the standard time is later than 9 P.M. For an earlier hour, turn the map clockwise. Thus the map may be made to represent the positions of the constellations in the northern sky at any time during the year.

1.27 Star Maps for the Different Seasons

Maps 2, 3, 4, and 5 show the constellations that appear in the vicinity of the observer's celestial meridian in the evening during each of the four seasons. The maps extend from the north celestial pole, at the top, down to the south horizon of latitude 40° N. Hour circles radiating from the pole are marked in hours of right ascension near the bottom of the map. Circles of equal declination go around the pole; their declinations are indicated on the central hour circle.

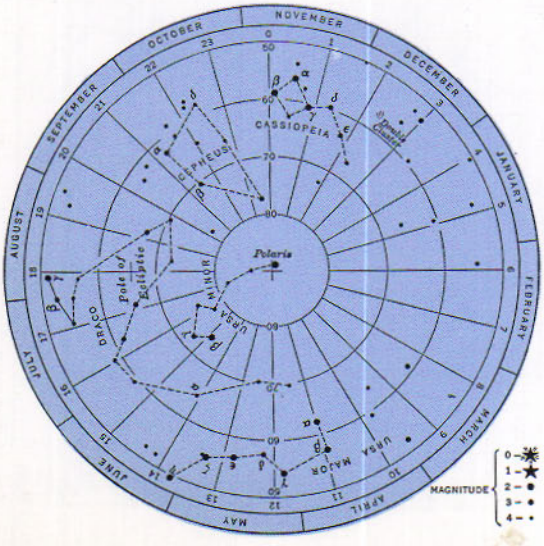
Select the map for the desired season; face south and hold it in front of you. The hour circle above the date of observation coincides with the celestial meridian at about 9 P.M., standard time. The stars near this hour circle are at upper transit at about this time. If the observer is in middle northern latitudes and is facing south, the stars represented in the upper part of the map are behind him. The northern constellations are arranged more conveniently, however, in Map 1; they are repeated in the seasonal maps to show how they are related to the constellations farther south.

Map 6 shows the region around the south celestial pole that is not visible from latitude 40° N.

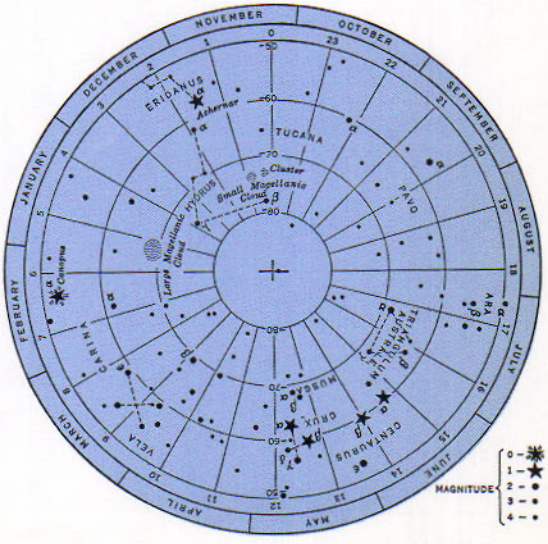
α	alpha	ι	iota	ρ	rho
β	beta	κ	kappa	σ	sigma
γ	gamma	λ	lambda	τ	tau
δ	delta	μ	mu	υ	upsilon
ε	epsilon	ν	nu	ϕ	phi
ζ	zeta	ξ	xi	χ	chi
η	eta	o	omicron	ψ	psi
θ	theta	π	pi	ω	omega

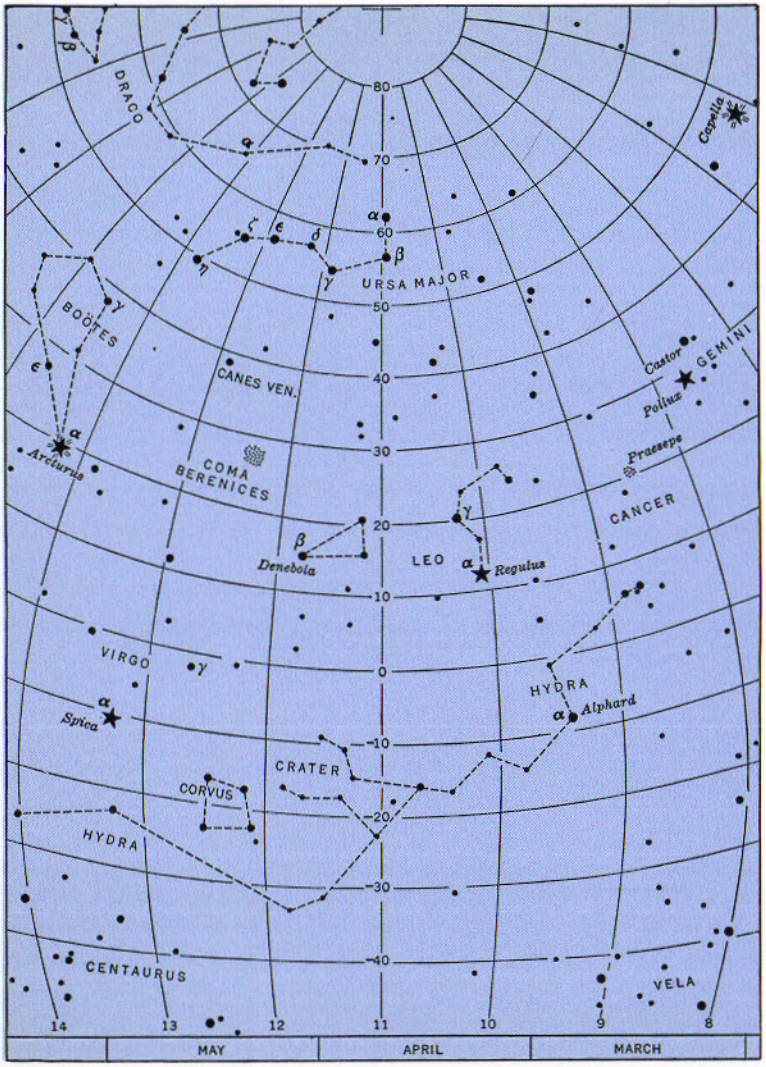
TABLE 1.I
Greek Alphabet (Small Letters)

Map 1
The northern constellations.

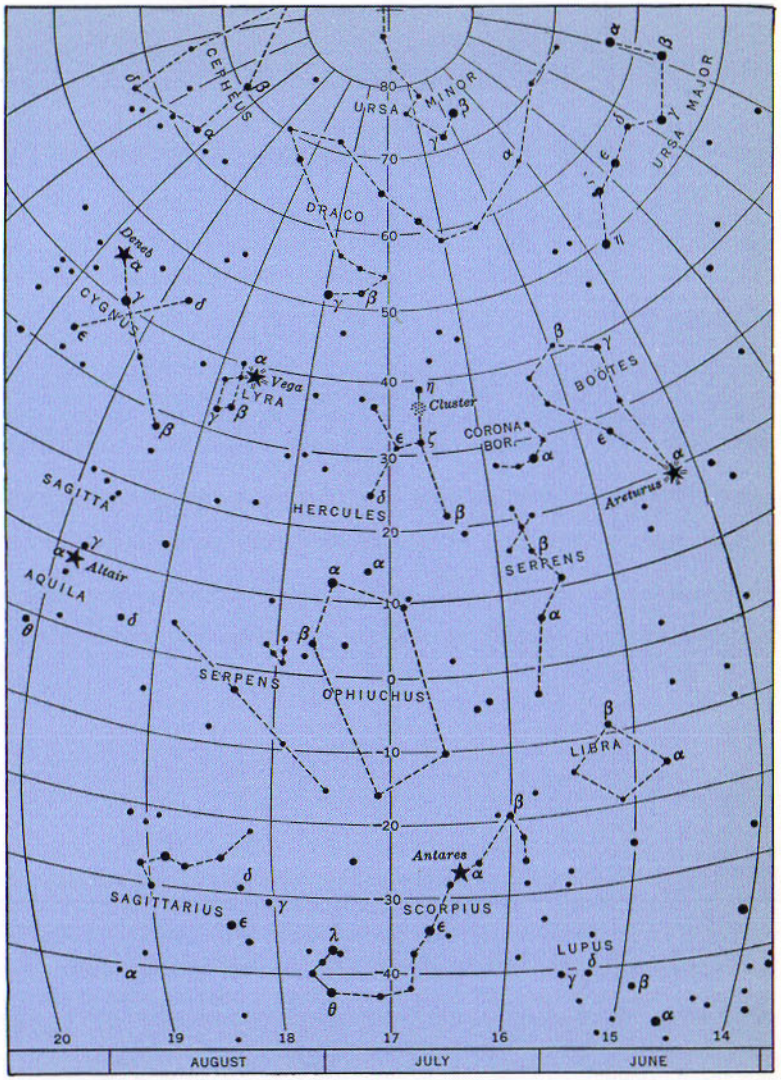


Map 6
The southern constellations.

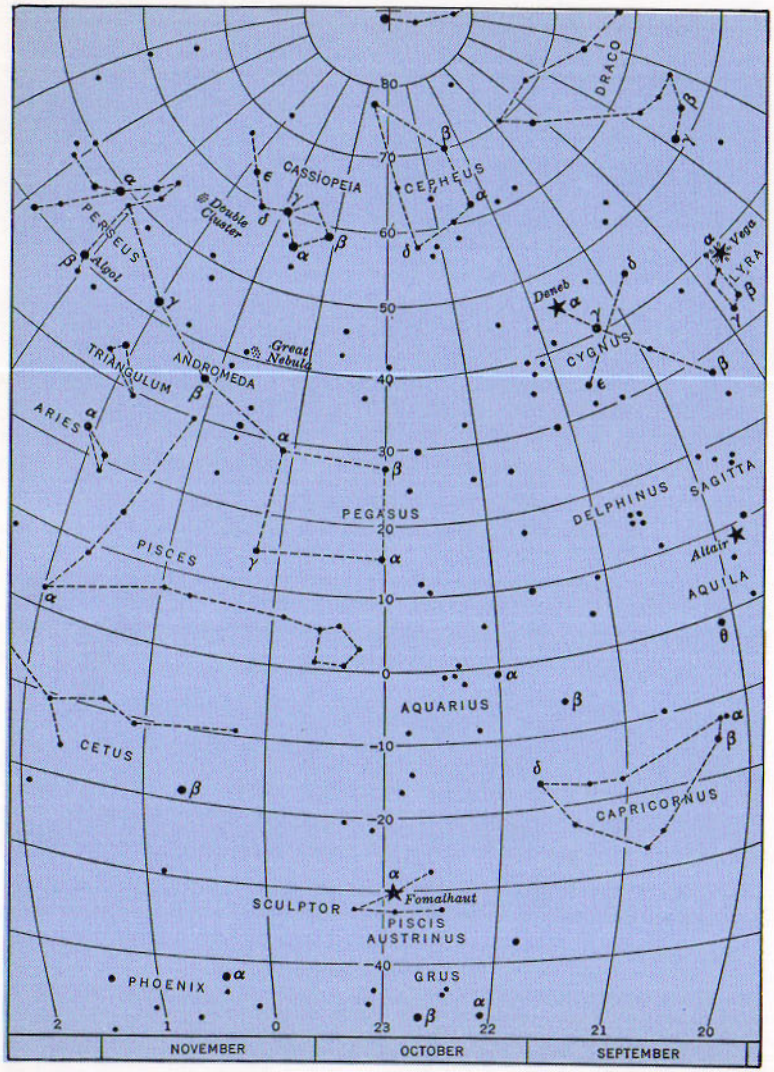




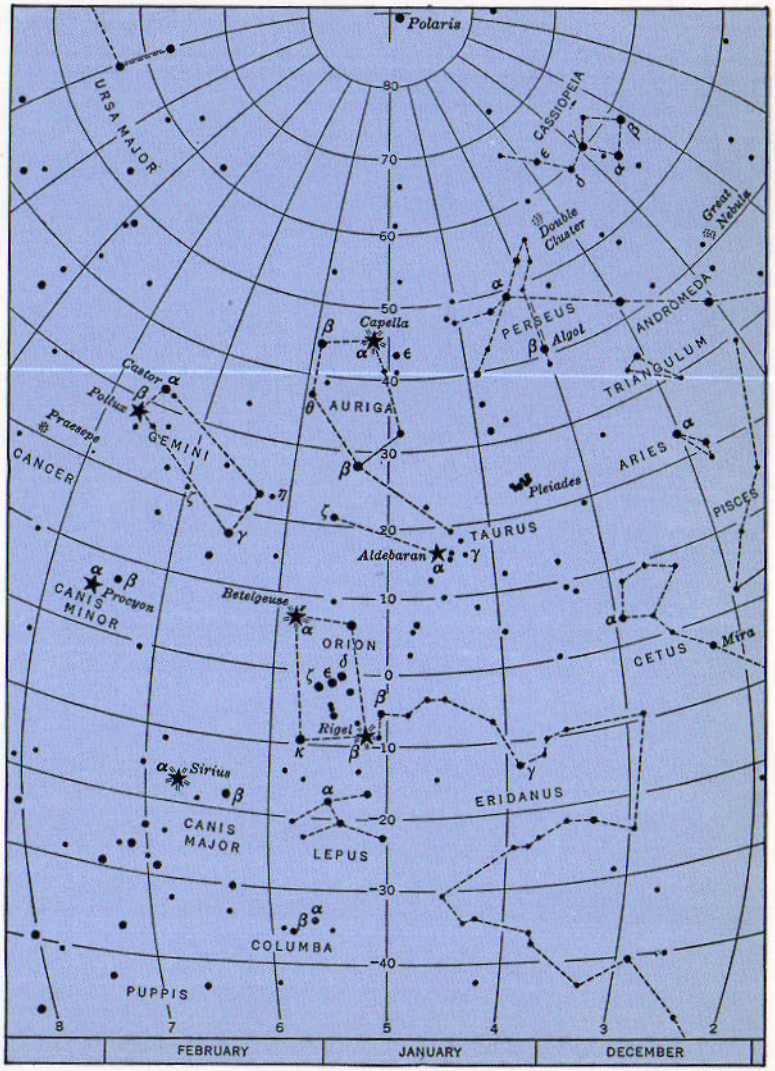
Map 2
The spring constellations.



Map 3
The summer constellations.



Map 4
The autumn constellations.



Map 5
The winter constellations.

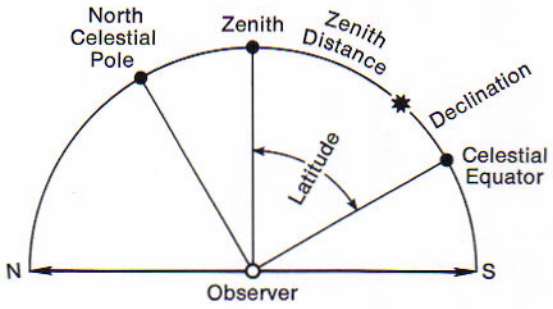


Figure 1.28
The latitude of a place equals the zenith distance of a celestial body at its upper transit of the meridian of the place plus its declination at that time.

1.28 Examples of the Use of the Star Maps

1. On what date is the bowl of the Great Dipper (Map 1) directly above the celestial pole at 9 P.M., standard time?
   Answer: May 1.
2. Read from Map 1 the right ascension and declination of δ Ursae Majoris (where the handle and bowl of the Dipper join).
   Answer: Right ascension $12^h\ 12^m$, declination $+57°$.
3. On what date is Antares (Map 3) at upper transit at 9 P.M., standard time? What is its zenith distance at that time in latitude 40° N.?
   Answer: July 13. The star is 66° south of the zenith.
4. Locate with respect to the constellations (Map 5) a planet in right ascension $5^h\ 30^m$ and declination $+24°$.
   Answer: The planet is midway between β and ζ Tauri.
5. At what time is Orion (Map 5) directly in the south on March 1?
   Answer: Orion is at upper transit at 7 P.M., standard time, on March 1.
6. When Orion has been recognized, how can its stars be used for finding Sirius and Procyon?
   Answer: The line of Orion's belt (δ, ε, and ζ Orionis) leads to Sirius. Procyon completes an equilateral triangle with Sirius and Betelgeuse in Orion.
7. How far south must we be in order to view Canopus, Crux, and the Large Magellanic Cloud (Map 6)?
   Answer: At least as far south as latitudes 37°, 30°, and 21° N., respectively.

1.29 The Planetarium

This is a remarkably successful device for showing a replica of the heavens indoors where many people may view it. The word *planetarium* refers either to the projection apparatus or to the structure that houses it. The Adler Planetarium in Chicago, opened to the public in 1930, first employed a projector of the Zeiss type in America. Others of this type are now in operation in many major cities throughout the United States and the world. In addition, comparable large instruments such as the Korkosz and the Spitz Model B are in international use. Many planetar-

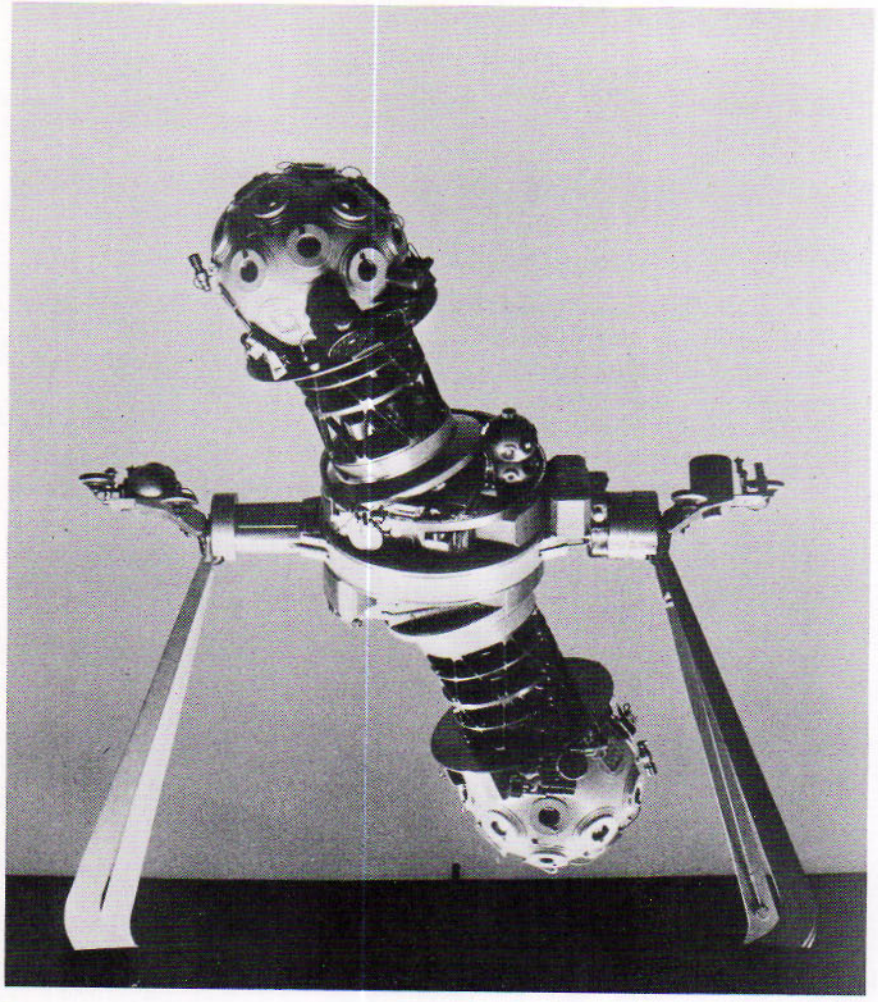


Figure 1.29
West German Zeiss planetarium projector of the Hayden Planetarium, New York City. *(Courtesy of Carl Zeiss, Inc., New York)*

iums, such as the American Museum-Hayden Planetarium in New York City, present four or five shows daily and change their subject matter on a monthly schedule. Planetariums serve as a focus for public education in astronomy by conducting courses and seminars from the most introductory to very advanced levels.

More than 400 smaller planetariums, for example, the Spitz series and

the Minolta projectors, are employed for instruction in astronomy in colleges and museums in the United States alone. These planetariums are often available to the general public on a limited basis.

The value of the planetarium as a teaching tool in spherical geometry and trigonometry, celestial navigation and practical astronomy is only now being realized and exploited. When used in practical astronomy, a fuller use of cloudy nights can be made leaving the clear nights for the use of telescopes and other equipment.

REVIEW QUESTIONS

1. Give brief definitions for:

Horizon	Zenith	Nadir
Vertical circles	Celestial meridian	Prime vertical

2. What is the conventional definition of azimuth as used by astronomers? What coordinate system uses azimuth as one of its coordinates?
3. Where does the star Regulus never rise? Where does the star Regulus never set?
4. At what northernmost latitude can one just see the Large Magellanic Cloud?
5. Describe the astronomical coordinate system of right ascension and declination.
6. Describe a simple method for determining one's latitude by astronomical means.
7. What is the brightest star in Cygnus? On what approximate date does this star transit your meridian at midnight?
8. What object is near the south pole of the ecliptic?
9. What is the name of the sun's apparent path in the sky? Explain the coordinate system based upon this path.
10. If Capella were the pole star, what would be the approximate declination of the star called Polaris?
11. Star maps 2 and 4 do not seem to have as many stars as maps 3 and 5; do your own observations confirm this? What feature in the sky does this fact seem related to?

REFERENCES

Allen, Richard H., *Star Names and Their Meanings.* New York: Dover Publications, 1963.

Baker, Robert H., *Introducing the Constellations,* revised ed. New York: Viking Press, 1957.

Norton, Arthur P. *A Star Atlas,* 14th ed. Edinburgh: Gall & Inglis, 1959.

Shapley, Harlow, ed., *Source Book in Astronomy,* Cambridge: Harvard University Press, 1960.

Shaw, R. William and Samuel L. Boothroyd, *Manual of Astronomy,* 3rd ed. New York: F. S. Crofts, 1947.

Webb, Thomas W., *Celestial Objects for Common Telescopes,* revised by Margaret W. Mayall, New York: Dover Publications, Inc., 1963.

Burnham, Robert, *Celestial Handbook.* Flagstaff, Arizona: Northland Press, 1966.
Davidson, Martin, *Elements of Mathematical Astronomy,* 3rd ed. New York: The Macmillan Co., 1962.
Smart, William M., *Spherical Astronomy.* New York: Cambridge University Press, 1962. (also available as a paperback from the same publisher)

FOR FURTHER STUDY

Big Bear Lake Solar Observatory of the California Institute of Technology *(Photograph from the Hale Observatories)*

2

THE EARTH IN MOTION

THE EARTH — THE EARTH'S ROTATION — THE EARTH'S REVOLUTION — THE EARTH'S PRECESSION — THE SEASONS

The earth is a globe about 12756 kilometers in diameter, slightly flattened at the poles and bulged at the equator. Its irregular surface is 70 per cent covered with water, and it is enveloped by an atmosphere to a height of several hundred kilometers. In addition to principal features of the earth itself, three motions of the earth and some of their consequences are described in this chapter. These are: (1) the earth's daily rotation on its axis; (2) the earth's annual revolution around the sun; (3) the earth's precession, which resembles the gyration of a spinning top.

The Earth from Apollo 11, at a distance of 281,000 km. *(NASA photograph)*

THE EARTH

The earth's size and shape have long been determined by astronomical techniques. The present method employs earth satellites referenced to the stars. Satellites measure the earth's gravitational field and its magnetic field.

2.1 The Earth's Globular Form

If the earth were a perfect sphere, after mountains and depressions are smoothed, all meridians would be circles, so that 1° of latitude would have the same value in kilometers (km) wherever the degree is measured. Because the latitude of a place (1.12) equals the altitude of the north celestial pole at that place, the length of 1° of latitude is the distance we must go along the meridian in order to have the pole rise or drop 1°; this distance may be measured by the appropriate method of surveying and is corrected to average sea level. Many such measurements, usually over longer arcs, show that 1° of latitude is everywhere nearly equal to 111 km. The results are:

At the equator, latitude	0°, 1° of latitude =	110.6 km
At latitude	20°	110.7
	40°	111.0
	60°	111.4
At the poles, latitude	90°	111.7

The *statute mile,* commonly used for measuring land distances in the United States, is seldom used to express distances by astronomers. Astronomers generally prefer to use *kilometers* which is also the distance unit used by most of the civilized world except in marine, air and space navigation where the nautical mile* is used.

The *nautical mile,* employed in marine navigation, is the length of 1′ of a great circle of the earth, regarded as a sphere for this purpose. It

	km	mile	n. mile
kilometer	1	0.621	0.540
statute mile	1.609	1	0.869
nautical mile	1.852	1.151	1.

* Not to be confused with the *knot,* a unit of *speed,* not length. One knot equals one nautical mile *per hour.*

equals about 6080 feet. Because of its relationship to circular measure on the earth, the nautical mile is widely used in marine and air navigation.

2.2 The Earth as an Oblate Spheroid

Although the length of a degree of latitude is everywhere about 111 kilometers, the steady increase in its value from the equator to the poles is significant. The greater length near the poles shows that the meridians curve less rapidly there than at the equator. The meridians are ellipses, and the earth itself approximates an oblate spheroid. For our purposes we will consider the earth's figure to be that of an ellipsoid of revolution, generated by the rotation of an ellipse around its minor axis, which coincides with the earth's polar axis. A detailed study of geodetic satellite results reveals that the earth has a slight bulge in the southern hemisphere and geodesists describe the true figure of the earth as pear shaped.

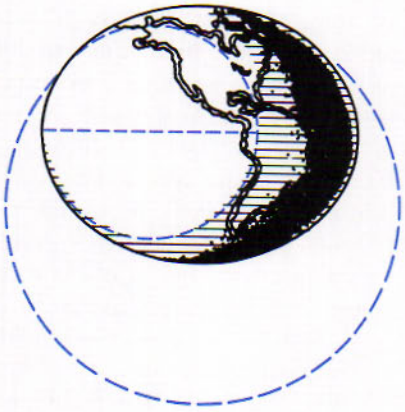


Figure 2.2
Curvature of a meridian at the pole and equator. The greater length of 1° of latitude at the pole, which is exaggerated in the diagram, shows that the meridian is there a part of a larger circle.

2.3 Dimensions of the Earth; Its Oblateness

The dimensions of the earth are considered to be the dimensions of the regular spheroid having a surface that most nearly fits the irregular surface of the earth. The dimensions of the *international spheroid* that follow were calculated in 1909 by the U.S. Coast and Geodetic Survey. This remains the standard spheroid of reference, although later surveys have given slightly different values. The dimensions of this spheroid are:

Equatorial diameter	=	12756.8 km
Polar diameter	=	12713.8
Difference	=	43.0 km

The longest arc of a meridian ever surveyed extends from Finland to the southern end of South Africa, a distance of 10,700 km. By the calculations of the U.S. Army Map Service from these data, completed in

1954, the earth's equatorial diameter is found to be 12,756.8 km, so that the circumference at the equator would be nearly 40,075.5 km.

The *oblateness,* or *ellipticity,* of the spheroid is found by dividing the difference between the equatorial and polar diameters by the equatorial diameter. It is the conventional way of denoting the degree of flattening at the poles. The small value of this fraction, $\frac{1}{297}$, for the earth shows that its flattening is slight. If the earth is represented by a globe 50 centimeters (nearly 20 inches) in diameter, the radius at the poles is only about 1.7 mm less than the equatorial radius, and the highest mountain is less than 0.3 mm above sea level. It has been said that the earth is more nearly spherical than are most of the balls in a bowling alley.

2.4 The Interior of the Earth

The earth's mass is 5.98×10^{27} grams, or 6.6×10^{21} short tons. This value is calculated from the acceleration of gravity at the surface by use of an appropriate formula (7.21). Dividing the mass in grams by the volume, which is 1.083×10^{27} cm^3, we find that the average density of the earth is 5½ times the density of water. This is one of the few data of observation we have concerning the earth's interior. Other information comes from an analysis of various satellite orbits and the transmission of earthquake waves at different depths. Variations in the earth's gravitational field are revealed by periodic harmonic terms in formulae representing a satellite's orbit. For example, the vast iron deposits in the Masabi range are clearly revealed by studies of satellite orbits. The very innermost regions of the earth can be studied by earthquake analysis as we shall briefly discuss later.

Aside from its atmosphere and hydrosphere, the earth consists mainly of two parts: the *mantle,* extending 2,900 kilometers below the crust, and the *core.* The *crust* is the relatively thin layer from the solid surface to the mantle. The continental crust has an average thickness of 32 kilometers, but the subocean crust may be as thin as 4 kilometers in some places. The crust is composed of igneous rocks such as granite and basalt, overlain with sedimentary rocks such as sandstone and limestone, all together about three times as dense as water. It is interesting that this is about the same density as the moon. The mantle is believed to be composed mainly of silicates of magnesium and iron.

The outer 1600 kilometers of the core behaves like a liquid; it does not transmit earthquake waves of the type resembling light waves. The inner core, 18 times as dense as water and supposedly very hot, is frequently said to consist mainly of nickel-iron such as occurs in many meteorites. This impression goes along with the idea that the heavier materials sank to the bottom when the earth was entirely molten.

Much of the increase in the density of the earth toward its center is caused by increasing compression. Near the center the pressure of the

overlying material rises to the order of 4×10^{12} dynes/cm^2 (58 million pounds/in^2), or more than 3 million atmospheres. W. H. Ramsey has suggested that the very dense core may not differ greatly in composition from the regions above it. The higher density may begin abruptly at the distance from the center where the pressure becomes great enough to collapse the molecules.

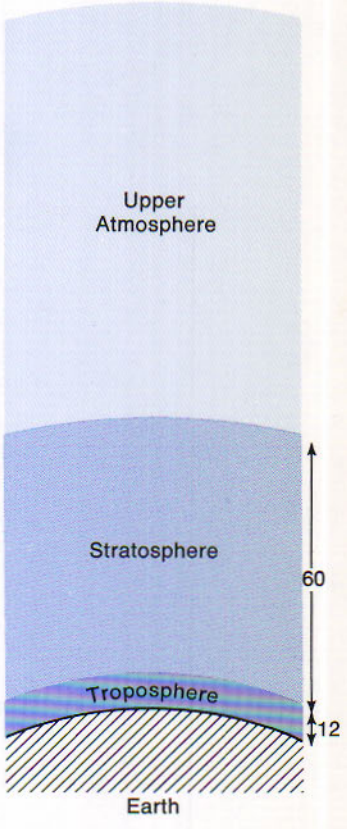
2.5 The Lower Atmosphere

The earth's atmosphere is a mixture of gases surrounding the earth's surface to a height of several hundred kilometers. From its average pressure of 1.013×10^6 dynes/cm^2 ($\sim$ 1 kg/cm^2 or 14.7 psi) at sea level, the mass of the entire atmosphere is calculated to be 5.2×10^{21} grams, or somewhat less than a millionth the mass of the earth itself. The air becomes rarefied with increasing altitude so rapidly that half of it by weight is within 5.6 kilometers from sea level. The lower atmosphere is divided into two layers: the troposphere and the stratosphere.

The *troposphere* extends to heights ranging from 16 kilometers at the equator to 8 kilometers at the poles. It is the region of rising, falling, and swirling currents, and of clouds that seem too frequently to obstruct our view of the heavens. The turbulent air makes the stars twinkle and often causes serious blurring of the view with the telescope. Its chief constituents are nitrogen and oxygen in the proportion of 4 parts to 1 by volume, carbon dioxide, and water vapor; there are other gases in relatively small amounts and dust in variable quantity. By their strong absorption of infrared radiations, water vapor and carbon dioxide are especially useful in preventing rapid escape of heat from the ground.

The *stratosphere* rises above the troposphere to an altitude of 72 kilometers. Here the currents are mainly horizontal, and only a little water vapor remains. Ozone, having molecules composed of 3 instead of 2 atoms of oxygen, is formed by action of the sun's extreme ultraviolet radiations on ordinary oxygen molecules, mainly in the lower stratosphere. The molecules are thereby dissociated into atoms, which by collision with unaffected oxygen molecules produce ozone. The ozone is subsequently dissociated by the radiations. By these and other processes the sun's destructive ultraviolet radiations are attenuated before they can reach the ground. These radiations are now studied spectroscopically in photographs from above the absorbing levels of the atmosphere to determine what information they bring about the sun itself.

The stratosphere contains one fifth of the entire air mass. To its upper limit it is still dense enough to make some contribution to twilight, as is known from the duration of this effect. *Twilight* is sunlight diffused by the air onto a region of the earth's surface where the sun has already set or has not yet risen. Astronomical twilight ends in the evening and begins in the morning when the sun's center is 18° below the horizon;



the fainter stars are then visible overhead in a clear sky. Civil twilight begins in the morning and ends in the evening when the sun is 6° below the horizon. This definition applies to the use of automobile headlights, street lighting, etc., for legal purposes on a clear day. Either twilight may persist for longer or shorter periods of time depending upon the inclination of the ecliptic, as discussed in 1.19. The times of sunset and sunrise and of the end and beginning of astronomical twilight can be found in some of the almanacs for any date and various latitudes.

2.6 The Upper Atmosphere

The *ionosphere* is the region of the atmosphere that is most affected by impacts of high-frequency radiations and high-speed particles from outside. Here the molecules of the gases are largely reduced to separate atoms, and the atoms themselves are shattered into ions and electrons. The ionosphere is often regarded as the layer between altitudes of 72 and 320 kilometers, where the ionized gases are more abundant, but the designation may be extended outward to include the entire upper atmosphere.

1 Hertz (Hz) = 1 cycle/sec.
k (kilo) = thousand
M (mega) = million

Three or more fluctuating ionized (12.6) layers of the upper atmosphere occur at successively higher altitudes. The D layer at the height of 70 to 90 kilometers is attributed to the sun's Lyman-alpha radiation. The E layer, at 90 to 130 kilometers, is ionized mainly by X-rays from the sun's corona. The F region, at 200 to 240 kilometers, is attributed to far ultraviolet radiations by helium in the sun. By repeated reflections from these layers and the ground, radio waves longer than 30 meters emitted from the ground can travel considerable distances over the earth before they are dissipated. The shortest of these are reflected from the top layer; they are employed for long-distance communications, which are interrupted when this layer is disturbed during a severe geomagnetic storm. Radio waves having lengths shorter than 30 meters generally penetrate all layers and escape. Conversely these also can come through from outside the earth to be received by radio telescopes.

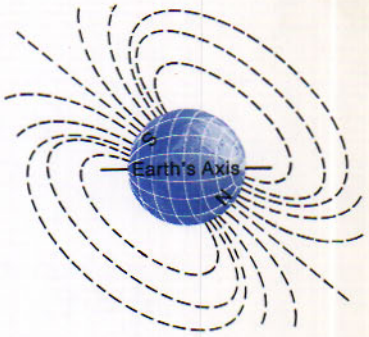
The influx of electrified particles from the sun illuminates the gases of the upper atmosphere in the airglow and the aurora (10.35). In the lower ionosphere the resistance of the denser air to the swift flights of meteors heats these bodies to incandescence, so that they make bright trails across the sky.

2.7 The Earth's Magnetic Field

The magnetic field of the earth resembles that of a bar magnet and hence is referred to as a dipole field. The axis of the field is inclined at a considerable angle with respect to the earth's pole of rotation and it passes very nearly through the center of the earth. Anomalies occur in the field strength above the surface as well as at the surface and even under the surface which may be attributed to large masses of electrically

conductive materials or to deposits of materials having a high magnetic permeability.

The earth's magnetic field at large distances above the earth (the magnetosphere) has been studied theoretically in great detail by C. Stormer. These studies predicted many of the features since confirmed by rocket and satellite-borne equipment. While Størmer predicted that ions with the proper velocities would be trapped by the earth's magnetic field, it was somewhat of a surprise when J. A. van Allen and his associates discovered two belts of very high density (Fig. 2.7), one centered about 3200 kilometers above the earth and the other at an approximate altitude of 16,000 kms. These studies, which began with the International Geophysical Year during 1957 and 1958 have continued and have been extended to the interaction of the earth's magnetic field with interplanetary space and the solar wind.



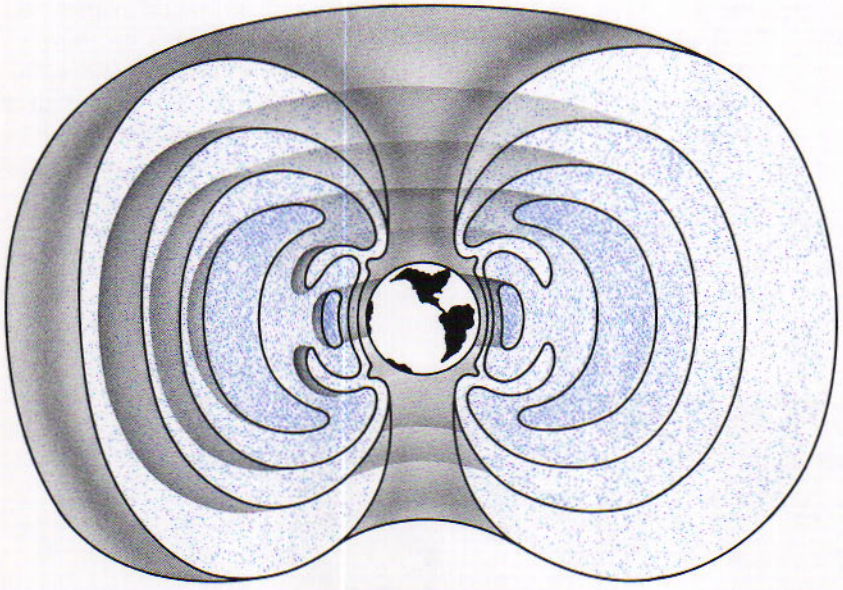


Figure 2.7
Radiation intensity contours around the earth showing the concentration in the regions referred to as the Van Allen belts.

Particles, mostly electrons and protons, trapped in the earth's magnetic field spiral back and forth along the lines of force as predicted by the laws of basic physics. Where the field lines dip down toward the magnetic poles the particles penetrate the atmosphere exciting various atoms and molecules causing auroras and night glow. When large amounts of particles are emitted during a period of solar activity, violent auroral displays can occur accompanied by discontinuous radio communications, violently oscillating compass needles, etc.

At one point in the south Atlantic Ocean off the coast of southern Brazil and Argentina, the zone of charged particles comes abnormally close to the surface of the earth. This region is referred to as the *South Atlantic Anomaly* and was discovered as a result of space exploration and research. It is presumably due to a large deposit of magnetically conducting ores.

2.8 The Earth as a Celestial Body

Because we live here instead of somewhere else in the planetary system, we do not view the earth in its proper relation to the rest of our community. From the nearest planets the earth would look like a brilliant star moving through the constellations; from Mars it would be a fine evening and morning star. From the outer planets it would be lost in the glare of the sun. From the nearest star all the planets would be invisible with the largest telescope, and the sun itself would be only one of the brighter stars.

Through the telescope the observer from a neighboring planet views the earth as a disk. He sees it marked by bright regions of snowfields and clouds, and by dark blue areas of water. He can determine the direction and period of the earth's rotation from the movement of surface markings and notes twilight zones between the day and night hemispheres. The observer would probably be unable to detect optical evidence of human life here. He could only consider, as we do for the other planets, whether conditions on the earth seem suitable for people like himself. Certainly from the excellent Mercury, Gemini and Apollo orbital photographs it is impossible to deduce unequivocally the existence of human life.

Figure 2.8
The Earth from about 281,000 km. Most of Africa, the Arabian Peninsula, the Mediterranean and parts of Europe and Central Asia can be easily identified in this spectacular photograph taken from Apollo 11. *(NASA photograph).*

2.9 The Earth Environment

The direct probing of space around the earth began effectively in the late 1940's with the launching of single stage rockets. According to the laws of mechanics (Chapter 7), if these rockets could achieve sufficient horizontal velocity above the very dense atmosphere they could then orbit the earth without further expenditures of fuel and return direct information about the earth's environment via radio transmissions.

The first artificial earth satellite was launched from a multistage vehicle by Russian scientists on October 4, 1957 and was called *Sputnik I.* Since that time, literally hundreds of earth orbiting spacecraft have been launched including dozens that were manned by astronauts and cosmonauts. In addition, numerous spacecraft have been sent to the moon, Venus, Mars and on solar orbits. Probably the most spectacular results from such spacecraft were those of the Mariner IV, VI and VII missions to Mars.

We have already discussed some results from space probes (2.7). Others have been synoptic weather coverage, radio and television relay stations, etc. Some of the information relayed by unmanned spacecraft, and particularly from the first space telescopes, will be discussed throughout this book.

THE EARTH'S ROTATION

Rotation *is turning on an axis, whereas* revolution *is motion in an orbit. Thus the earth rotates daily and revolves yearly around the sun. The earth rotates from west to east on an axis joining its north and south poles. Among the effects of the rotation are the behavior of the Foucault pendulum, the directions of prevailing winds, the spinning of cyclones, and the oblateness of the earth.*

2.10 Absence of Earlier Proof of the Rotation

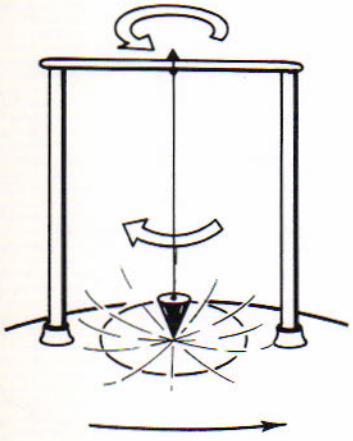
Although the early Greek scholars cited evidence that the earth is a globe, they believed with few exceptions that it was motionless. As late as the time of Copernicus and in fact beyond it, no convincing proof was available that the earth had any motion at all. The alternation of day and night and the rising and setting of the stars could mean either that the heavens are turning daily from east to west or that the earth is rotating

from west to east; but the second interpretation was generally dismissed as unreasonable.

Copernicus favored the earth's rotation because it seemed to him more probable that the smaller earth would be turning rather than the larger celestial sphere. His conviction that the earth also revolves around the sun was likewise unsupported by rigorous proof. In more recent times, many conspicuous effects of the earth's rotation have become known.

2.11 The Foucault Pendulum

A freely swinging pendulum affords a simple and effective demonstration of the earth's rotation. The experiment was first performed for the public by the French physicist J. B. L. Foucault in 1851 under the dome of the Pantheon in Paris. This pendulum consisted of a large iron ball freely suspended from the center of the dome by a wire 60 meters long and set swinging along a line marking the meridian. Those who watched the oscillating pendulum saw its plane turn slowly in the clockwise direction. They were observing in fact the progressive change in the direction of the meridian caused by the earth's rotation under the invariable swing of the pendulum.



The demonstration was widely acclaimed as convincing proof of the earth's rotation, a fact that had not been universally accepted even at that late date. Similar demonstrations are often given today in planetariums, museums, and elsewhere.

The behavior of the Foucault pendulum is most easily understood if we imagine it suspended directly above the pole. A meridian there is turned completely around in a day with respect to the unvarying swing of the pendulum, always toward and away from the same star. In general, the observed deviation Δ of the pendulum from the meridian in 1 hour is 15° times the sine of the latitude (l). At the pole the rate is 15° an hour. In the latitude of Chicago it is 10° an hour, so that the plane of the swing seems to go around there once in 36 hours. The deviation is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. At the equator there is no deviation at all, because the meridian also keeps the same direction as the earth rotates.

$\Delta = 15 \sin l$

2.12 Deflection of Objects Moving Horizontally

Because all parts of the earth's surface rotate in the same period, the linear speed of the rotation varies with the latitude; it is greatest at the equator and diminishes toward the poles.

In its flight toward the target, a projectile retains the speed of the eastward rotation at the place from which it started, aside from air resistance. Fired northward in the northern hemisphere, the projectile moves toward a place of slower rotation; it is therefore deflected ahead of, or to the east of, the target. If it is fired southward instead, the projectile moves toward a place of faster rotation; it now falls behind, or to the west of the target.

In either event the deflection is to the right when the observer faces in the direction of the flight. If the projective is fired in the southern hemisphere, the deflection will evidently be to the left.

In general, objects moving horizontally above the earth's surface are deflected to the right in the northern hemisphere and to the left in the southern hemisphere. The deflection is relative to a meridian, which is skewed meanwhile by the earth's rotation. Although this consequence of the rotation is not conspicuous in the case of short-range projectiles, convincing deflections are found in prevailing winds, cyclones, and long range ballistic missiles.

2.13 Deflection of Surface Winds

The general circulation of the atmosphere shows the effect of the earth's rotation on the transport of heat from the equatorial to the polar regions. The warmer air near the equator rises and flows toward the poles. Cooled at the higher altitudes, the currents descend, especially around latitudes 30°, and flow over the surface. Thus we have the easterly trade winds of the tropics and the westerlies of the temperate zones (Fig. 2.13) as these currents are deflected by the earth's eastward rotation. In addition, there are easterly surface winds in the polar regions, where the colder air moves toward the equator.

By this account the prevailing currents in the mid-latitudes, where the temperature gradients are steepest, are frequently dynamically unstable and break into large eddies. These are the large-scale cyclonic and anticyclonic vortices of the temperate zones.

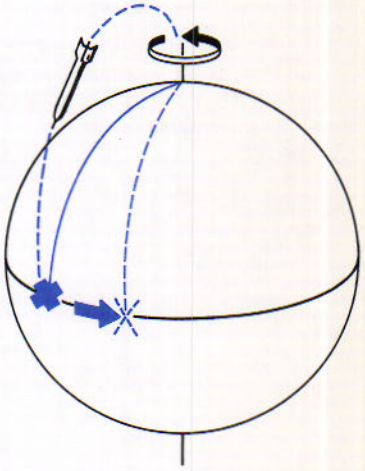


Figure 2.13
Prevailing surface winds. The moving air is deflected to the right in the northern hemisphere and to the left in the southern hemisphere.

2.14 Cyclones Show the Deflection Effect

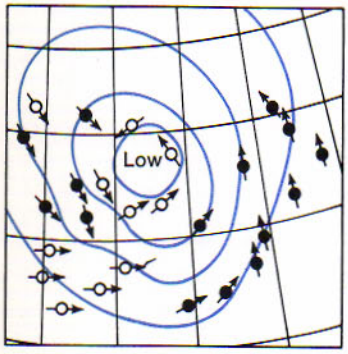


Figure 2.14
A cyclone in the northern hemisphere. The whirl is counterclockwise, as the wind direction arrows show.

The *cyclones* of the temperate zones are great vortices in the atmosphere averaging 2500 kilometers in diameter, which migrate eastward and are likely to bring stormy weather. Marked "low" in the weather map, they are areas of low barometric pressure into which the surface air is moving. The inflowing currents are deflected like projectiles by the earth's rotation, so that they spiral inward in the counterclockwise direction in the northern hemisphere and clockwise in the southern hemisphere.

Anticyclones, marked "high" in the weather map, are areas from which the surface air is moving. The outflowing currents are deflected by the earth's rotation so that they spiral outward, clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. Thus the vortex motions of cyclones and anticyclones are consequences of the earth's rotation.

This deflection of moving objects relative to the surface of the rotating earth is known as the *Coriolis effect* after the French engineer who first discussed it a century ago. Its importance extends beyond its meteorological consequences. Correction for the effect is required in the control of long-range missiles and in the use of the level bubble, as in the bubble octant employed on a moving plane.

Another effect of the earth's rotation is found in the behavior of the gyrocompass, which brings the axis of its rotor into the plane of the geographical meridian and thus shows the direction of true north. Still another effect is the bulging of the earth's equator.

2.15 Cause of the Earth's Oblateness

$$F = \frac{Wv^2}{gr}$$

The effort of a stone to escape when it is whirled around at the end of a string is an example of the centrifugal tendency of whirling bodies. Similarly, all parts of the rotating earth tend to move away from the earth's axis; the action is greatest at the equator and diminishes to zero at the poles. This effect at any place may be regarded as the resultant of two effects operating at right angles to each other:

1. *The lifting effect* (F) in grams, caused by the earth's linear velocity (v) at a point r from the axis, is opposed to the earth's attraction and therefore diminishes the weight (W) of an object at that place. This would cause an object weighed on a spring balance to weigh less at the equator than at the poles by 1 gram in 289 if the earth were a sphere. The actual reduction in weight is 1 gram in 190. An object at the equator also weighs less than at the poles because it is farther from the center of the earth. Here we have additional evidence of the earth's oblateness.
2. *The sliding effect* of the earth's rotation is directed along the surface toward the equator. Yet things that are free to move—the water of the oceans, for example—have not assembled around the equator,

as they would have done if the earth were a sphere. The centrifugal effect of the earth's rotation has produced enough oblateness of the earth itself to compensate this sliding effect.

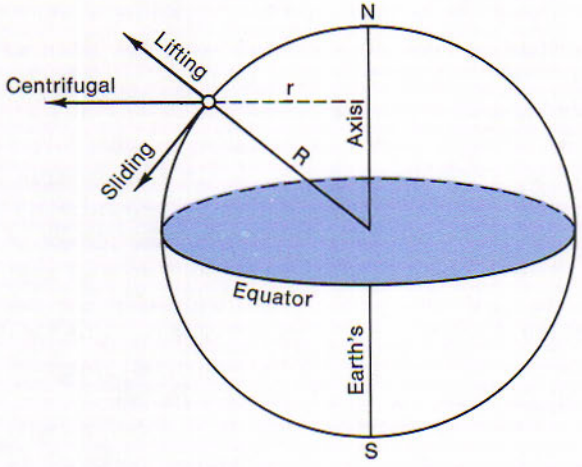


Figure 2.15
Effect of the earth's rotation on a body at its surface. The centrifugal effect directed away from the earth's axis is resolved into two effects at right angles. One diminishes the weight of the body; the other urges it toward the equator.

2.16 Gravity at the Earth's Surface

Gravity is the result of the earth's attraction (gravitation) directed nearly toward its center and diminished by the lifting effect of the earth's rotation. The acceleration of gravity, g, is the rate at which a falling body picks up speed; its value at sea level increases from 978.039 cm/sec^2 at the equator to 983.217 cm/sec^2 at the poles. The weight of an object, which equals its mass multiplied by g, is therefore less at the equator than at the poles by 5.178/983.217, or 1 gram in 190.

Values of the acceleration of gravity are precisely determined by timing the swing of a pendulum at different places. For the simple pendulum, $g = 4\pi^2 l/t^2$, where l is the length of the pendulum in centimeters and t is the time in seconds of a complete oscillation. It must be pointed out that the pendulum is a method for determining g. The acceleration of gravity (often called the superficial gravity) actually depends directly on the mass of the attracting body and inversely on the square of its radius.

A more fundamental constant is the universal gravitational constant G which has a value of 6.67×10^{-8} dyn cm^2 gm^{-2}. We shall see (7.21) that the determination of the values of g and G allow us to determine the mass of the earth, which in turn leads to finding the mass of the moon, which in turn leads to the mass of the sun. Several experiments, presently in progress, promise to improve the values of g and G by a factor of 100. It is surprising to most scientists that G is not known to better than 1 part in 500. This is because gravity is a weak force making it difficult to measure accurately.

1 gram = 0.035274 ounces
1 pound = 453.59 grams

$$g \propto \frac{m}{r^2}$$

2.17 The Variation of Latitude

The declination of the zenith of a place on the earth and therefore the latitude of the place (1.12) can be determined with remarkable accuracy by observations of the transits of stars with a photographic zenith telescope. The difference between the latitudes of two widely separated stations can be measured by this means within 0".01, or about 30 centimeters on the earth's surface. When the latitudes of two places on opposite sides of the earth are measured repeatedly, the values prove to be continuously varying; if one latitude is increasing at a particular time, the other is decreasing. Because latitude is reckoned from the equator, which is midway between the north and south poles, it follows that the poles are not stationary points on the earth's surface.

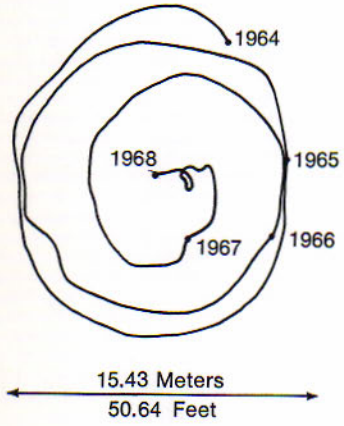


Figure 2.17
Wandering of the north pole, 1964 to 1968. The path of the pole on the earth's surface is shown, with its position at the beginning of each year.

By international cooperation the latitudes of stations at about the same distances north and south of the equator and in different longitudes have been determined persistently over the past 60 years. The records show how the poles have oscillated during that interval. Each pole describes a complex path, which is always confined within an area smaller than that of a baseball diamond.

The wandering of each pole may be resolved mainly into two nearly circular motions. The first, having a period of 12 months and a diameter of about 6 meters, is ascribed to seasonal variations in the distribution of ice, snow and air masses. The second, having a period around 14 months and a diameter that has varied from 3 to 15 meters, has a cause less well agreed upon but is apparently related to the geological history of the earth. The effects of both are slight shiftings of the earth with respect to its axis, which keeps a constant direction in space as far as these are concerned.

There seems to be no possibility of wider migrations of the poles in the past that might account for the marked variations of climates in geologic times.

2.18 Variations in the Earth's Rotation

The earth's rotation has long provided the master clock by which all terrestrial and celestial events have been timed. It has been known for some time, however, that the rate of the rotation is not precisely uniform. This conclusion is based on studies of periodic celestial motions that are independent of the earth's rotation, particularly the monthly revolution of the moon around the earth. If the moon as timed by the earth-clock is forging ahead of its prescribed schedule, the earth is presumably running slow. Periodic variations in the rotation have also been detected with clocks of high precision. The variations in the rate of the earth's rotation are classified as periodic, irregular, and secular.

1. *Periodic variations* are mainly annual and semiannual; they appear to be caused in large part by winds and tides. The period of the rotation becomes $0^s.001$ longer near the vernal equinox and $0^s.001$

shorter near the autumnal equinox than the average for the year. The earth-clock, as compared with crystal-controlled clocks, becomes as much as $0^s.03$ slow in the former season and the same amount fast in the latter.

2. *Irregular variations.* During the past 200 years, the error of the earth-clock, as determined from studies of the motion of the moon, has accumulated to 30 seconds, first in one direction and then in the other. These variations, according to D. Brouwer, are caused by small, cumulative random changes in the rate of the earth's rotation.
3. *Secular variations.* Moon-raised tides in the oceans and in the earth itself should act as brakes to reduce the speed of the earth's rotation and thus to lengthen the day. Recorded observed times of early eclipses have seemed to show that the earth-clock has run slow by 3¼ hours during the past 20 centuries as compared with a clock having a uniform rate.

THE EARTH'S REVOLUTION

It was thought that if the earth revolved around the sun the stars could be detected oscillating back and forth. In an effort to detect this a much larger and equally convincing proof of the earth's revolution was discovered. The average radius of the earth's orbit of revolution is called the astronomical unit.

2.19 Evidence of the Earth's Revolution

The sun's annual motion among the constellations is not a proof of the earth's revolution around the sun, for by itself it might leave us in doubt as to whether the sun or the earth is moving. With the aid of a telescope, other annually periodic phenomena are observed, which provide conclusive evidence of the earth's revolution. Among the effects of this kind are the following.

1. *The annual parallaxes of the stars* are the periodic changes in the alignments of the nearer stars relative to the more distant ones. This effect, which is described in Chapter 11 in connection with the distances of the stars, is so minute that it was not detected until the year 1837. The failure to observe it had contributed greatly to

the persistence of the idea that the earth was stationary, until another effect became known, which serves just as well as a proof of the earth's revolution and is much easier to observe.

2. *The aberration of starlight* was discovered by the English astronomer James Bradley, in 1727. It is also an annually periodic change in the directions of the stars.

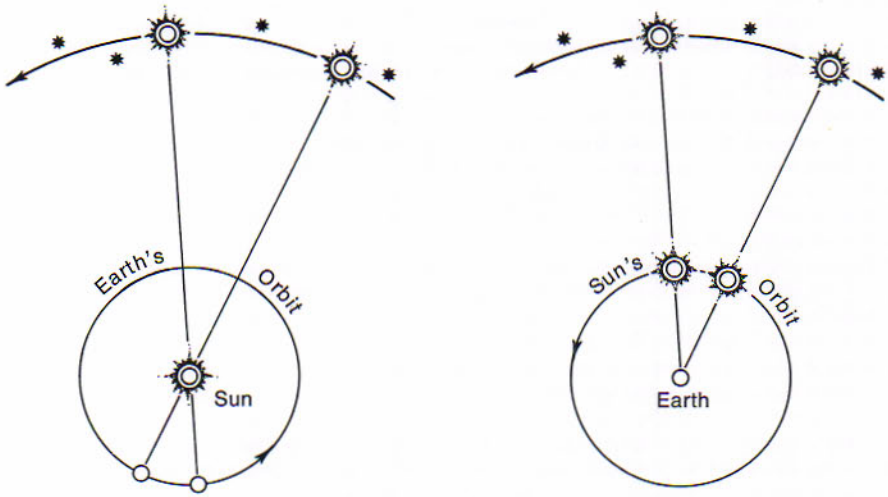


Figure 2.19
The sun's apparent motion among the constellations is not a conclusive proof of the earth's revolution. A similar effect would be observed if the sun revolved around the earth.

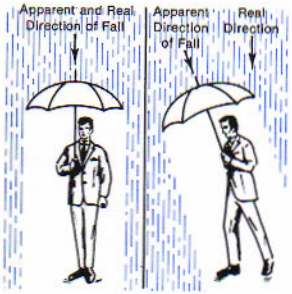
2.20 Aberration of Starlight

Raindrops descending vertically on a calm day strike the face of the pedestrian. Whatever direction he takes, the source of the raindrops seems to be displaced from overhead in that direction. If he runs instead, the apparent slanting direction of the rain becomes more noticeable; and if he drives rapidly, the direction may seem to be almost horizontal. This is a familiar example of aberration.

Aberration of starlight is the apparent displacement of a star in the direction the earth is moving. The amount of the displacement depends on three factors: (1) It is directly proportional to the speed of the earth (v_e). (2) It is inversely proportional to the speed of light (c). Whereas the moderate speed of the pedestrian in the rain causes a considerable displacement of the source of the raindrop, very swift movement such as that of the revolving earth is required to produce an appreciable change in the direction of a star. (3) The displacement is greatest when the earth moves at right angles to the star's direction and becomes zero if the earth moves toward or away from the star.

If the earth were motionless, there would be no aberration of starlight. If it had only uniform motion in a straight line, the displacement of the star would be always the same and might therefore be unnoticed.

Figure 2.20
Aberration of raindrops. The source of the raindrops is apparently displaced in the direction of the observer's motion.



If the earth revolves, the changing direction of its motion would cause the star's displacement to change direction as well, always keeping ahead of us, so that the star would seem to describe a small orbit. This is precisely what the telescope shows. The aberration of starlight is convincing evidence of the earth's revolution around the sun.

2.21 Aberration Orbits of the Stars' the Constant of Aberration

A star at either pole of the ecliptic has a circular apparent orbit, because the earth's motion is always perpendicular to the star's direction. Because the earth's orbit is an ellipse, the true direction of the star is not precisely at the center of the circle. A star on the ecliptic oscillates in a straight line. Between the ecliptic and its poles the aberration orbit is an ellipse. We view these orbits (Fig. 2.21A) flatwise at the ecliptic pole, edgewise on the ecliptic, and at various angles in between.

The *constant of aberration* is the apparent displacement of a star when the earth is revolving at average speed at right angles to the star's direction. Its value is the same for all stars regardless of their distance or direction; it is the radius of the circle at the ecliptic pole, half the major axis of the ellipse, and half the length of the straight line on the ecliptic. The value of the constant of aberration is about 20″.5. It is nearly 30 times as great for all stars as the parallax effect for even the nearest star and was accordingly the earlier of the two to be detected.

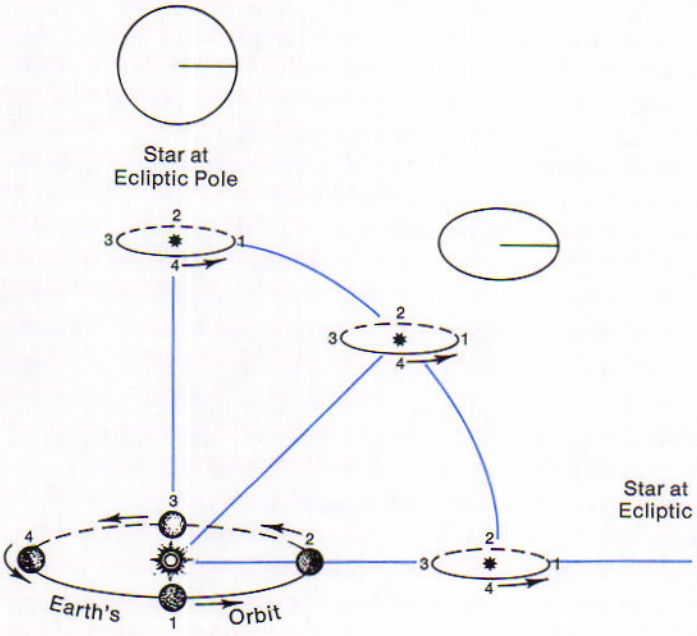


Figure 2.21A
Aberration orbits of the stars. The numbers mark corresponding positions of the earth in its orbit and of the stars in their apparent aberration orbits. The outer figures show the observed forms of the aberration orbits.

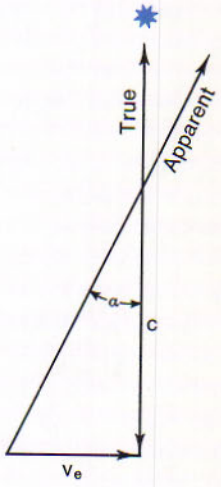


Figure 2.21B
Aberration of starlight.

The situation is represented by the right triangle of Fig. 2.21B. The side v_e is the earth's average speed in its orbit, the side c is the speed of light, and the angle a is the aberration constant. If the earth did not revolve, so that v_e would be zero, a would also be zero. If the light were propagated instantly, so that c would be infinite, a would again be zero. The observed aberration of starlight demonstrates both the earth's revolution and the finite speed of light.

2.22 The Earth's Orbit

It is an ellipse of small eccentricity with the sun at one focus. It is the path the earth follows in its revolution around the sun and is not to be confused with the ecliptic, the great circle that the sun seems to describe annually on the celestial sphere. The plane of the earth's orbit is also the plane of the ecliptic. Because the orbits of celestial bodies are ellipses, the following definitions will be useful here and elsewhere.

The *ellipse* is a plane curve such that the sum of the distances from any point on its circumference to two fixed points within, the *foci*, is constant and equal to the major axis of the ellipse (Fig. 2.22).

The *eccentricity* of the ellipse is half the distance between the foci divided by half the major axis. It is the conventional way of denoting the degree of flattening of the ellipse. The eccentricity may have any value between 0, when the figure is a circle, and 1, when it becomes a parabola. The eccentricity of the earth's orbit is about 0.017, or $\frac{1}{60}$.

2.23 The Earth's Distance from the Sun

The earth's *mean distance* from the sun is 149,598,000 km.; it is half the length of the major axis of the orbit, or the average between the least and greatest distances from the sun. This distance is known as the *astronomical unit* (a.u. or A.U.), because it is frequently taken as the unit in stating the distances of the nearer celestial bodies. The distances of the planets from the sun are given in Table 1 (Appendix) in astronomical units as well as in kilometers.

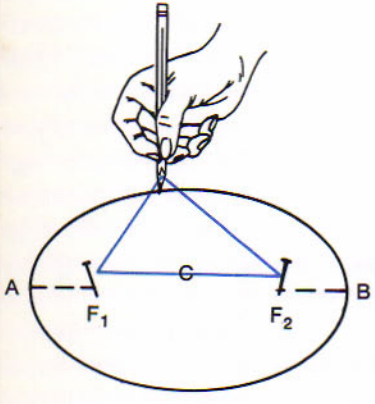


Figure 2.22
The ellipse. The sum of the distances from the foci, F_1 and F_2, is constant and equal to the major axis, AB. The eccentricity of the ellipse is the fraction F_1C/AC.

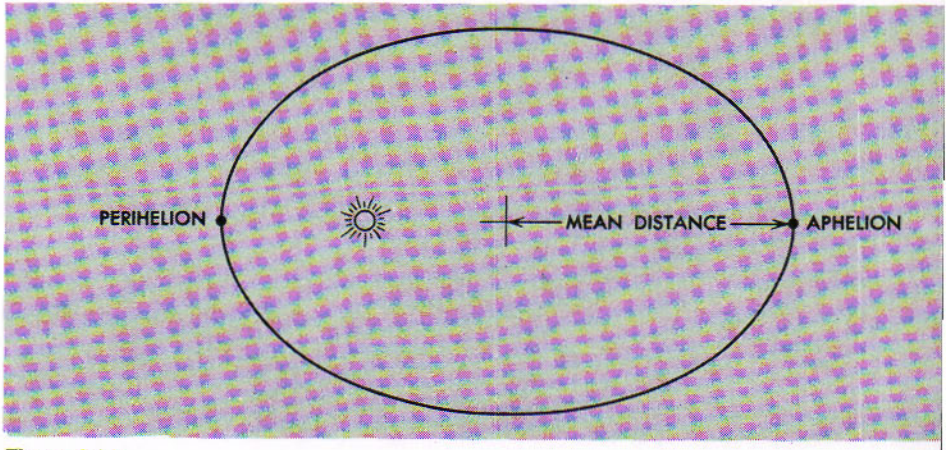


Figure 2.23
The earth's orbit. It is an ellipse of small eccentricity (much exaggerated in the diagram), having the sun at one focus.

Perihelion and *aphelion* are the two points on an orbit respectively nearest and fartherst from the sun; they are the extremities of the major axis. The earth is at perihelion early in January, when its distance from the sun is 1.7 per cent, or about 2.5×10^6 kilometers, less than the mean. It is at aphelion early in July, when its distance is the same amount greater than the mean. The earth is at its mean distance from the sun early in April and October, when it is at the extremities of the minor axis of the orbit.

This preliminary description of the earth's motion relative to the sun neglects effects of the attractions of the earth by neighboring bodies and, of course, the motion of the sun itself.

THE EARTH'S PRECESSION

The direction of the earth's axis is not fixed in space. It precesses in a circle around the pole of the earth's orbit, hence the present pole star will not be near the pole 5000 years from now.

2.24
Conical Motion of the Earth's Axis

The axis of a spinning top describes the surface of a cone around a line perpendicular to the floor. When it stops spinning, the top falls over. While it continues to spin, the action of gravity does not tip the axis, but causes instead the conical motion we observe in the same direction as that of the rotation. This is the precession of the top.

Just as the axis of the spinning top is likely to lean away from the vertical, so the earth's axis is inclined 23½° from the perpendicular to the plane of its orbit, and its equator is inclined to this plane by the same amount. The attractions of the moon and sun, both nearly in the ecliptic plane, for the earth's equatorial bulge try to bring the equator into the plane of the ecliptic. Because of the earth's rotation, however, the inclination is not much affected. Again, as in the case of the top, there is a conical motion of the axis, but in the opposite direction with respect to the rotation.

The *earth's precession* is a slow conical movement of the earth's axis around a line joining the ecliptic poles, having a period of 25,780 years.

The effect we are considering is a change in the axis relative to the stars. It is unlike the wandering of the terrestrial poles (2.17), which is caused by a shifting of the earth upon its axis.

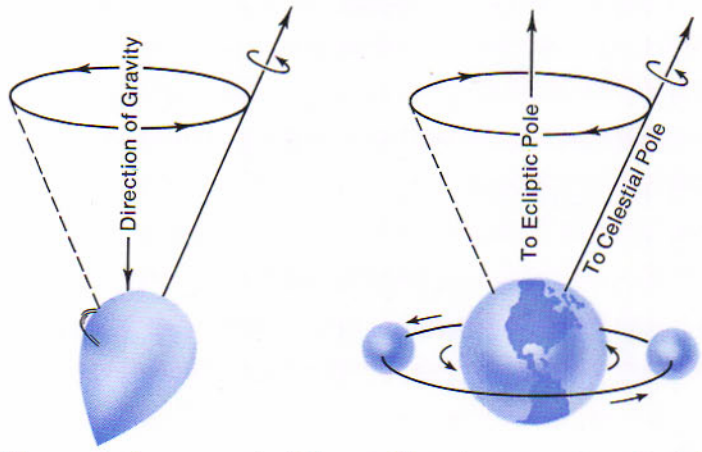


Figure 2.24
The earth resembles a spinning top. The effort of the moon, and of the sun, to bring the plane of the earth's equator into the ecliptic plane combines with the earth's rotation to produce the conical precession of the earth's axis.

2.25 Precessional Paths of the Celestial Poles

The conical movement of the earth's axis causes the celestial poles, toward which the axis is directed, slowly to describe circles around the ecliptic poles; the radii of the two circles are the same and equal to 23½°. This is a movement of the poles among the constellations.

As one faces north, the precessional motion of the north celestial pole is counterclockwise. This pole is now about 1° from the star Polaris, which it will continue to approach until the least distance of slightly less than half a degree is reached, about the year 2100. Thereafter, the diurnal circle of Polaris will grow larger. For those who live in the year 7000,

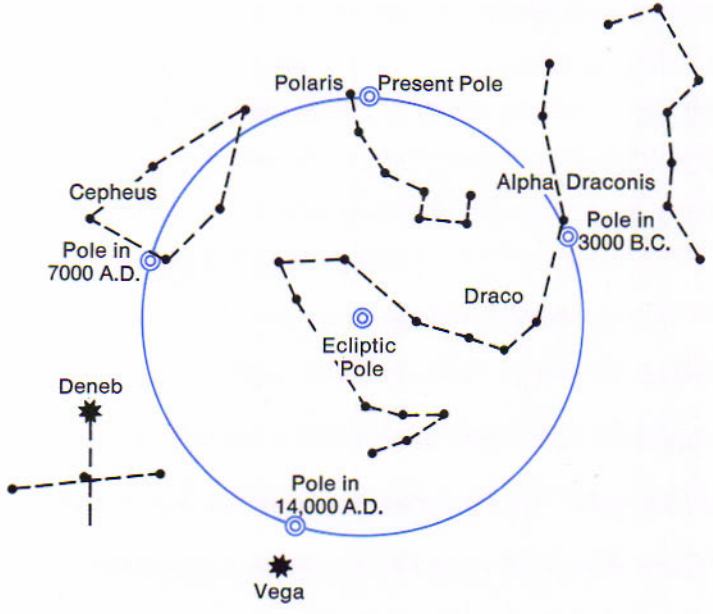


Figure 2.25
Precessional path of the north celestial pole. The celestial pole describes a circle of 23½° radius around the ecliptic pole.

Alpha Cephei will be the nearly invariable pole star, and Polaris will circle daily around it 28° away.

Because the celestial poles are the centers of regions where the stars never set or never rise, the precessional motion shifts the constellations relative to these regions, out of them or into them. The Southern Cross, which rose and set 6000 years ago throughout the United States, is now visible only from the extreme southern part of this country.

2.26 The General Precession

It is the *lunisolar precession* that has been described. The sun's attraction contributes to this effect as well as the moon's attraction, but in smaller amount. *Planetary precession* is the effect of other planets on the plane of the equator, so that its intersection with the ecliptic shifts slowly toward the east along the celestial equator. The result of the two precessions is the *general precession.*

A complete account of precession involves additional factors. Because the inclination of the moon's path to the plane of the earth's equator varies in a period of 18.6 years (5.11), the celestial pole describes a small ellipse in this period around its mean position in the precessional path. The semimajor axis of the ellipse is 9″.2 in the direction of the ecliptic pole. This is the chief term in *nutation,* the nodding of the pole. Thus the precessional path of the celestial pole is irregular; it is not exactly circular and is not precisely the same from one cycle to the next.

2.27 Precession of the Equinoxes

The earth's precession has been defined as a conical movement of the axis. It may also be regarded as a corresponding gyration of the earth's equator and of the celestial equator in the same plane. The intersection of the ecliptic with the celestial equator slides westward on the ecliptic, keeping about the same angle between the two. The equinoxes, where

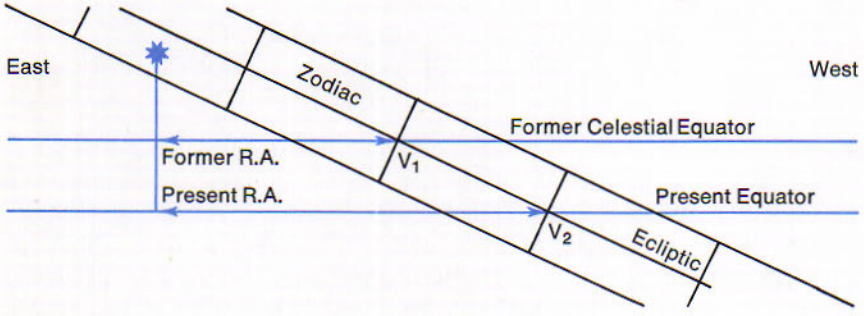


Figure 2.27
Precession of the equinox as seen from the earth. The westward motion of the vernal equinox, from V_1 to V_2, causes the signs of the zodiac (the 12 equal divisions marked off from the equinox) to shift westward away from the corresponding constellations. Right ascensions and declinations of the stars are altered by precession.

the two circles intersect, accordingly shift westward along the ecliptic; they move in the general precession at the rate of 50″.26 in celestial longitude in a year. This is the *precession of the equinoxes.*

The annual displacement of the vernal equinox in right ascension is now 46″.09, or $3^s.07$, and in declination is 20″.05. Thus the equatorial coordinates of the stars on the celestial sphere, which are measured from the vernal equinox, are continuously changing. Accurate catalogs give the right ascensions and declinations of the stars at a stated time and the annual variations of these positions caused by precession as well as by the motions of the stars themselves.

Two other effects of the precession of the vernal equinox are described in the two following sections. These are the displacement of the signs of the zodiac relative to the constellations of the same names and the shortening of the year of the seasons.

2.28 Signs and Constellations of the Zodiac

The *zodiac* is the band of the celestial sphere, 16° in width, through which the ecliptic runs centrally. It contains at all times the sun and moon, and the principal planets, with the exceptions of Venus and Pluto; these two and many asteroids are not confined within its limits.

The *signs of the zodiac* are the 12 equal divisions, each 30° long, which are marked off eastward beginning with the vernal equinox. The signs are named from the 12 *constellations of the zodiac* situated in the respective divisions over 2000 years ago. The names of the signs and the seasons in the northern hemisphere in which the sun is passing through them are as follows:

Aries	Spring	Libra	Autumn
Taurus		Scorpius	
Gemini		Sagittarius	
Cancer	Summer	Capricornus	Winter
Leo		Aquarius	
Virgo		Pisces	

Because of the precession of the equinoxes, the vernal equinox has moved westward about 30°, and the signs have moved along with it, away from the constellations after which they were named. Thus the signs and constellations of the zodiac of the same names no longer have the same positions. When the sun, on March 21, arrives at the vernal equinox and therefore enters the sign Aries, it is near the western border of the constellation Pisces and will not enter the constellation Aries until the latter part of April.

THE SEASONS

The passing of the seasons is a natural measure of time. The seasons result from the inclination of the earth's axis to the plane of its orbit. The earth's atmosphere causes the seasons to lag behind what would be expected if solar radiation alone was considered.

2.29 The Year of the Seasons

The year is the period of the earth's revolution, or of the sun's apparent motion in the ecliptic. The kind of year depends on the point in the sky to which the sun's motion is referred, whether this point is fixed or is itself in motion. Just as the day in common use is not the true period of the earth's rotation, so the year of the seasons is not the true period of its revolution. Two kinds of year have the greatest use.

The *sidereal year* is the interval of time in which the sun apparently performs a complete revolution with reference to a fixed point on the celestial sphere. Its length is $365^d\ 6^h\ 9^m\ 10^s$ ($365^d.25636$) of mean solar time, which is now increasing at the rate of $0^s.01$ a century, in addition to any change caused by variations in the rate of the earth's rotation. The sidereal year is the true period of the earth's revolution.

The *tropical year* is the interval between two successive returns of the sun to the vernal equinox. Its length is $365^d\ 5^h\ 48^m\ 46^s$ ($365^d.24220$) of mean solar time and is now diminishing at the rate of $0^s.53$ a century. It is the year of the seasons, the year to which the calendar conforms as nearly as possible. Because of the westward precession of the equinox, the sun returns to the equinox before it has gone completely around the ecliptic. The year of the seasons is shorter than the sidereal year by the fraction $50''.26/360°$ of 365.25636 days, or a little more than 20 minutes.

2.30 Cause of the Seasons

The change of seasons takes place because the earth's equator is inclined 23½° to the plane of its orbit. It keeps nearly the same direction in space during a complete revolution, each pole is presented to the sun for part of the year and is turned away from it for the remainder of the year.

The amount of the inclination determines the boundaries of the climatic zones. The *frigid zones* are the regions within 23½° from the poles, in which the sun becomes circumpolar and where the seasons are accordingly extreme. The *torrid zone* has as its boundaries the tropics of Cancer and Capricorn, 23½° from the equator. Here the sun is overhead at noon at least once a year; the durations of sunlight and darkness never differ greatly, and temperature changes during the year are not extreme.

In the *temperate zones* the sun never appears in the zenith, nor does it become circumpolar, aside from effects of atmospheric refraction (4.3) and the considerable size of the sun's disk.

The inclination of the earth's equator to its orbit causes the sun's annual migration in declination. When the sun is farthest north, at the summer solstice, its altitude at noon is the greatest for our northern latitudes and the duration of sunlight is the longest here for the year. At the winter solstice we have the other extreme, namely, the lowest sun at noon and the shortest duration of sunlight.

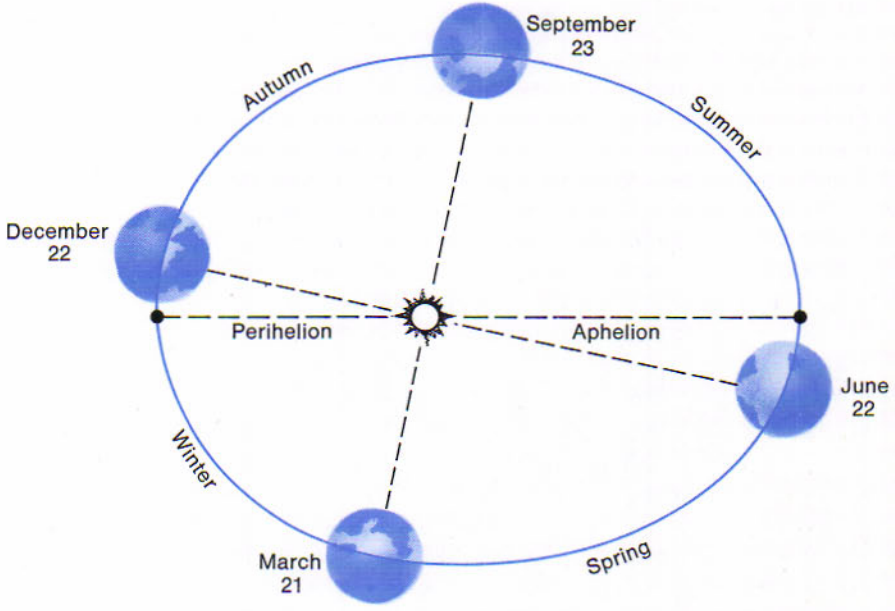


Figure 2.30
The seasons in the northern hemisphere. This hemisphere is inclined farthest toward the sun at the summer solstice (June 22) and farthest away at the winter solstice (December 22). The earth is nearest the sun early in January.

2.31 Seasonal Changes in Temperature

Temperature changes are produced mainly by differences in the *insolation*, or exposure to sunshine of the regions of the earth's surface. The daily amount of the insolation depends on the intensity of the radiation that is received and the duration of the sunshine.

When the sun is higher in the sky, so that its rays fall more directly on the ground, the radiation is more concentrated. When the sun is lower, so that its rays fall more obliquely, a given amount of radiation is spread over more territory and is less effective in heating any part of it. The rays from the lower sun also have to penetrate a greater thickness of the atmosphere and are subject to more absorption. Summer with us is a warmer season than winter because the two factors conspire to produce higher temperatures; the sun's altitude becomes greater and the daily duration of sunshine is longer.

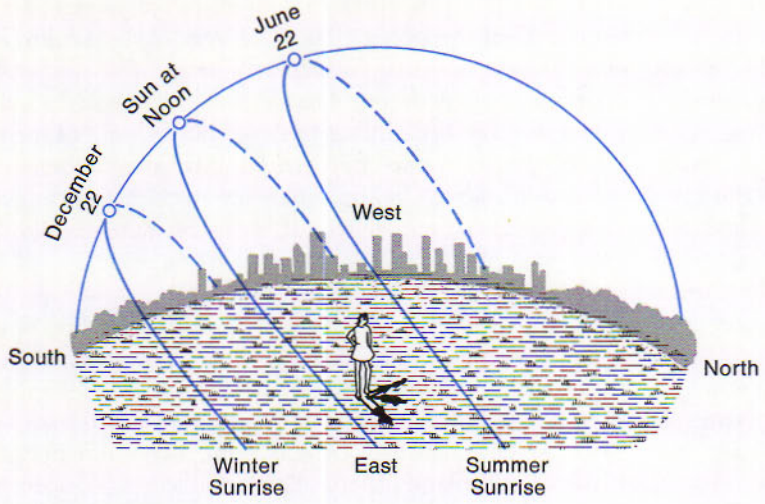


Figure 2.31
Diurnal circles of the sun in different seasons. The daily duration of sunshine is longer in the summer, and the sun is higher at noon.

At the time of the summer solstice the sun is higher at noon in the latitude of New York than it is at the equator, and it is visible for a longer time, so that the amount of heat delivered in a day is fully 25 per cent greater. Even at the north pole at that time the daily insolation exceeds that at the equator. The uninterrupted radiation from the circumpolar sun compensates its lower altitude; but the temperature is lower at the pole because much of the heat is taken up in the melting of ice.

2.32 Lag of the Seasons

If the temperature depended on insolation alone, the warmest days in the United States and Canada should come around June 22 and the coldest part of the winter about December 22. Our experience, however, is that the highest and lowest temperatures of the year are likely to be delayed several weeks after the times of the solstices. The reason is found in the conservation of heat by the earth and its atmospheric blanket.

It is the balance of heat on hand that determines the temperature. As with one's bank balance, the quantity increases as long as the deposits of heat exceed the withdrawals. On June 22 we receive the maximum amount of radiation. Afterward, as the sun moves southward, the receipts grow less, but for a time they exceed the amounts the earth returns into space. As soon as the rate of heating falls below the rate of cooling, the temperature begins to drop. In the winter the sun's altitude at noon and the daily duration of sunshine increase after December 22. It is not until considerably later in the season that the rate of heating overtakes the rate of cooling, so that the temperature begins to rise.

2.33 The Seasons in the Southern Hemisphere

These differ from those in the northern hemisphere, of course, in that a particular season occurs at the opposite time of year. Another difference might be introduced by the eccentricity of the earth's orbit. Summer in the southern hemisphere begins at the date of the winter solstice, which is only a little while before the earth arrives at perihelion, nearest the sun. It might be supposed that the southern summer would be warmer than the northern summer, which begins when the earth is near aphelion, and similarly that the southern winter would be colder than the northern winter.

The earth's distance from the sun at perihelion, however, is only about 3 per cent less than the distance at aphelion; and the slightly greater extremes of temperature that might otherwise be experienced in the southern hemisphere are modified by the greater extent of the oceans in that hemisphere. It will be noted later that the conditions that might produce more extreme temperatures in our southern hemisphere are repeated in the case of the planet Mars. There the effect is observable owing to the greater eccentricity of that planet's orbit.

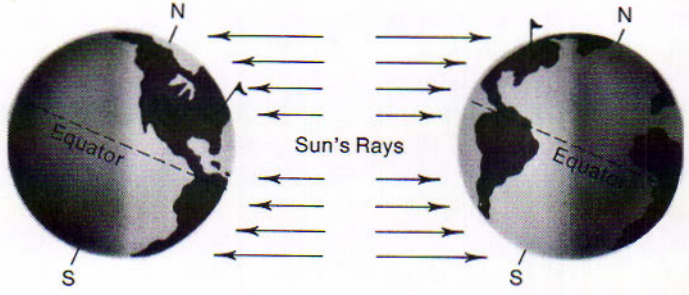


Figure 2.33
Corresponding seasons in the two hemispheres occur at opposite times of the year. Northern summer is represented at the left and southern summer at the right.

REVIEW QUESTIONS

1. What information helps us infer the conditions in the core of the earth?
2. Earth satellites often orbit the earth in orbits 200 kilometers above the earth. What part of the atmosphere is this?
3. What is the cause of the aurora?
4. At what latitude north and south does astronomical twilight last all night on June 21 and December 21?
5. What angle does the celestial equator make with the horizon for an observer at latitude 20°N? At latitude 20°S?
6. Derive the relation for the period of the Foucault pendulum starting with the relation in section 2.11. This relation may be used to determine one's latitude, however, it is almost never used. Why?
7. Explain why anticyclones form around high pressure areas.
8. From information given in section 2.15, show that the equatorial radius of the earth is larger than the radius of a sphere by the ratio $^{289}/_{190}$.

9. Mars has a mass $\frac{1}{10}$ that of the earth and a radius $\frac{1}{2}$ that of the earth, show that the value of g on the surface of Mars is about 391cm/sec^2.
10. Give two ways of finding variations in the earth's superficial gravity.
11. What is the variation of latitude having a one year period ascribed to?
12. Give three specific causes for variations in the earth's rotation.
13. Why was the aberration of starlight discovered before the parallax effect (motion of the stars due to the earth's revolution around the sun)?
14. What is the cause of precession?

REFERENCES

Namowitz, Samuel N. and Donald B. Stone, *Earth Science,* 3rd ed. Princeton: D. Van Nostrand Co., 1965.

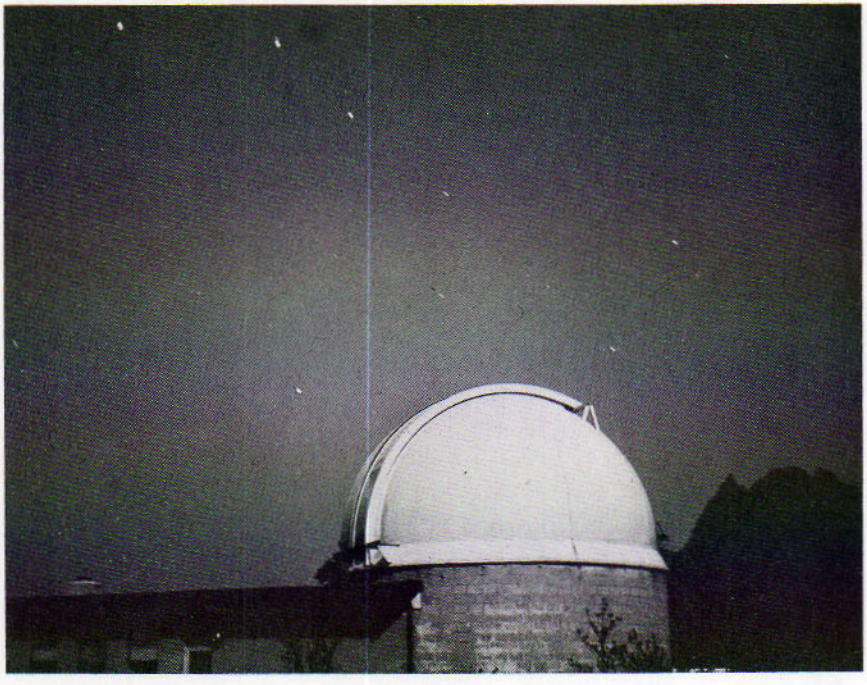
Orr, Clyde, Jr., *Between Earth and Space.* New York: The MacMillan Co., 1959.

Rosen, Sydney, Robert Siegfried and John M. Dennison, *Concepts in Physical Science.* New York: Harper & Row, 1965.

FOR FURTHER STUDY

Kuiper, Gerard P., ed., *The Earth as a Planet.* Chicago: University of Chicago Press, 1955.

Mueller, Ivan I., *Spherical and Practical Astronomy.* New York: Frederick Ungar Publishing Co., 1969.



The moonlit dome of the Flower and Cook Observatory, University of Pennsylvania. *(Time exposure by L. Fredrick)*

3

TIMEKEEPING

THE TIME OF DAY—THE CALENDAR

The natural units of time that we find most suitable for our activites are provided by two motions of the earth. The day is the period of the earth's rotation, or of the resulting apparent rotation of the heavens; it is divided arbitrarily into smaller units: the hour, minute, and second. The year is the period of the earth's revolution around the sun; it is divided naturally into the four seasons. The two main units and some arbitrary ones are combined conventionally in the calendar.

The standard time of day in common use is more readily explained if we first consider the kinds of time of day the astronomers read from the clock in the sky in order to derive the correct standard time.

Tuning-fork wrist timepiece. *(Courtesy of the Bulova Watch Co.)*

THE TIME OF DAY

Telling time requires general agreement on a zero or reference point. Historically this has been the moment the sun crosses the meridian. This defines the day. Different types of time are used for different purposes, some as a matter of convenience (zone time, universal time, etc.) others as a matter of necessity (sidereal time, atomic time, etc.).

3.1 Time Reckoning

Two features on the face of a clock are required for showing the time of day: first, a time reckoner, the hour hand; second, a reference line, the line joining the noon mark to the center of the dial. The angle between the hour hand and the reference line, when it is converted from degrees to hours and minutes, denotes the time of day. Divisions and numerals around the dial are added for convenience, and interpolating devices, the minute and second hands, add accuracy to the reading of the time. In most ordinary clocks the hour hand goes around twice instead of once in a day, as it does in the case of the celestial clock.

Figure 3.1
A tuning-fork wrist timepiece. A minute tuning-fork vibrates 360 times each second, permitting an accuracy of 60 seconds per month and a chartable rate to one or two seconds per month. The fork is driven electromagnetically. *(Courtesy of the Bulova Watch Co.)*

To observe the time of day from the master clock in the sky, a point on the celestial sphere is chosen as the *time reckoner*. The part of the hour circle joining that point to the celestial pole may be regarded as the hour hand; it circles westward around the pole once in a day. The *reference line* is the observer's celestial meridian.

It is *noon* when the time reckoner is at upper transit (1.11), and *midnight* when it is at lower transit of the meridian. A *day* is the interval between two successive upper or lower transits by the time reckoner. *Time of day* is either the hour angle of the time reckoner if the day begins at noon (which is true of the sidereal day), or it is the hour angle of the time reckoner plus 12 hours if the day begins at midnight. These definitions apply to any kind of local time.

The three time reckoners in use are the vernal equinox, the apparent sun, and the mean sun. The corresponding kinds of time are sidereal time, apparent solar time, and mean solar time.

3.2 The Sidereal Day Is Shorter Than the Solar Day

In the upper position of the earth, in Fig. 3.2, it is both sidereal and solar noon about March 21 for the observer at *0*. The vernal equinox and the sun are both at upper transit over his meridian. By the time the earth has made a complete rotation relative to the equinox, so that this meridian is very nearly parallel to its original direction, the earth in its revolution around the sun has moved to its lower position in the diagram. In the new position it is zero hours sidereal time of the following day; the vernal

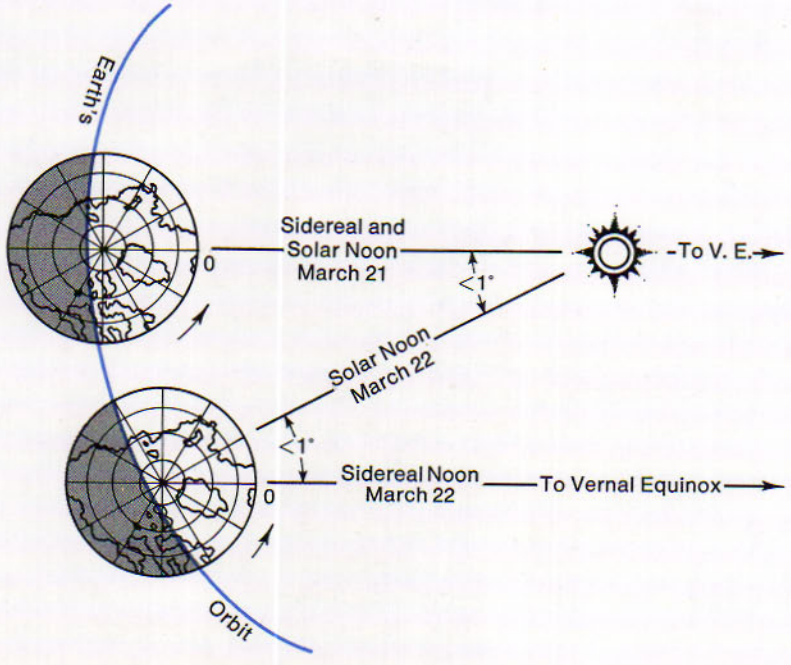


Figure 3.2
The sidereal day is shorter than the solar day. Because it is also revolving around the sun, the earth must rotate farther after completing the sidereal day before the solar day is ended.

equinox on the remote celestial sphere is again at upper transit, so that the sidereal day is completed. Because of its revolution, however, the earth must rotate still farther before one solar day has passed.

Now the angle through which the earth revolves in 1 day averages 360°/365.25, or a little less than 1°. It is evident from the figure that this is also the angle through which the earth must rotate after completing the sidereal day before the ending of the solar day. Because the earth rotates at the rate of 15° an hour, or 1° in 4 minutes, the sidereal day is about 4 minutes shorter than the solar day.

More exactly, the difference is $3^m\ 55^s.909$, so that the length of the sidereal day is $23^h\ 56^m\ 4^s.091$ of mean solar time. Owing to the precession of the equinox (2.27), the sidereal day is actually slightly shorter than the true period of the earth's rotation, which is $23^h\ 56^m\ 4^s.099$ of mean solar time.

3.3 Sidereal Time

This is the local hour angle of the vernal equinox and might have been more correctly called equinoctial time. The sidereal day begins with the upper transit of the equinox and is reckoned through 24 hours to the next upper transit. Sidereal time agrees with our solar time about September 21. It gains on solar time thereafter at the rate of $3^m\ 56^s$ a day, which accumulates to 2 hours in a month and to a whole day in the course of a year. Evidently the earth rotates once more in a year than the number of solar days in the year.

Figure 3.3
A pendulum sidereal clock. The perfection of the pendulum clock was an early major advance in timekeeping. The model here was built by Parkinson and Frodsham. *(Photograph by H. Bluemel)*

By our watches a star rises, or crosses the celestial meridian, 4 minutes earlier from night to night, or 2 hours earlier from month to month. Thus a star that rises at 10 o'clock in the evening on November 1 will rise at 8 o'clock on December 1. The westward march of the constellations with the advancing seasons (1.17) is caused by the difference in length between the sidereal and the solar day.

3.4 Determining the Sidereal Time

Sidereal time is kept in the observatories by sidereal clocks, which run about 4 minutes faster than ordinary clocks in the course of a day. The sidereal clocks at the Naval Observatory and elsewhere are corrected frequently by observations of stars. Corrections are required not only for errors in the running of the clocks themselves, but also because of unpredictable irregularities in the earth's rotation. When the correct sidereal time is known, the corresponding standard time clocks may be kept right as well.

Sidereal time is determined by observing transits of stars across the celestial meridian. The rule employed, which can be verified by reference to Fig. 3.4A, is as follows: The sidereal time at any instant equals the right ascension of a star that is at upper transit at that instant. In order to apply the rule it is necessary to know precisely when the star is at upper transit. This can be observed by means of a meridian transit instrument.

The *meridian transit instrument* in its simplest form is a rather small telescope mounted on a single horizontal axis, which is set east and west so that the telescope may be pointed only along the observer's celestial meridian. Having directed the telescope to the place where the star will

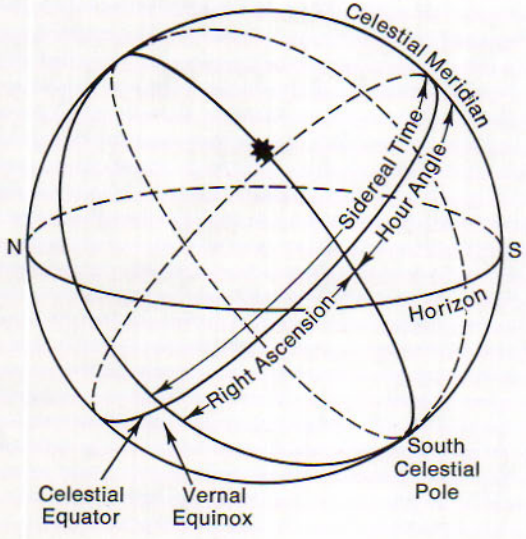
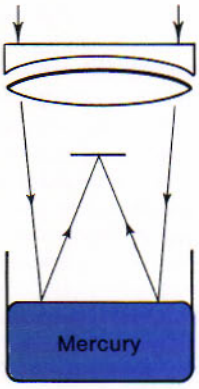


Figure 3.4A
Relation between the sidereal time and the right ascension and hour angle of a star. Sidereal time equals the right ascension of a star plus its hour angle. When the star is at upper transit (hour angle zero), the sidereal time equals the star's right ascension.

soon transit, the observer looks into the eyepiece and watches the star move toward the middle of the field of view. A vertical wire through the middle of the field marks the celestial meridian. At the instant that the star's image is bisected by the wire, the reading of the sidereal clock is recorded.

Suppose that the clock reading is $6^h\ 40^m\ 17^s.2$ and that the star's right ascension is given in an almanac as $6^h\ 40^m\ 15^s.7$. The clock is accordingly $1^s.5$ fast, because at the instant of its upper transit the star's right ascension is the correct sidereal time.

Recording devices are employed to time the transits of stars more accurately than can be done by visual observations. The *photographic zenith tube* (Fig. 3.4A) has replaced the simple transit instrument at the Naval Observatory and in other places. This is a fixed vertical telescope, with which the stars are photographed as they cross the celestial meridian



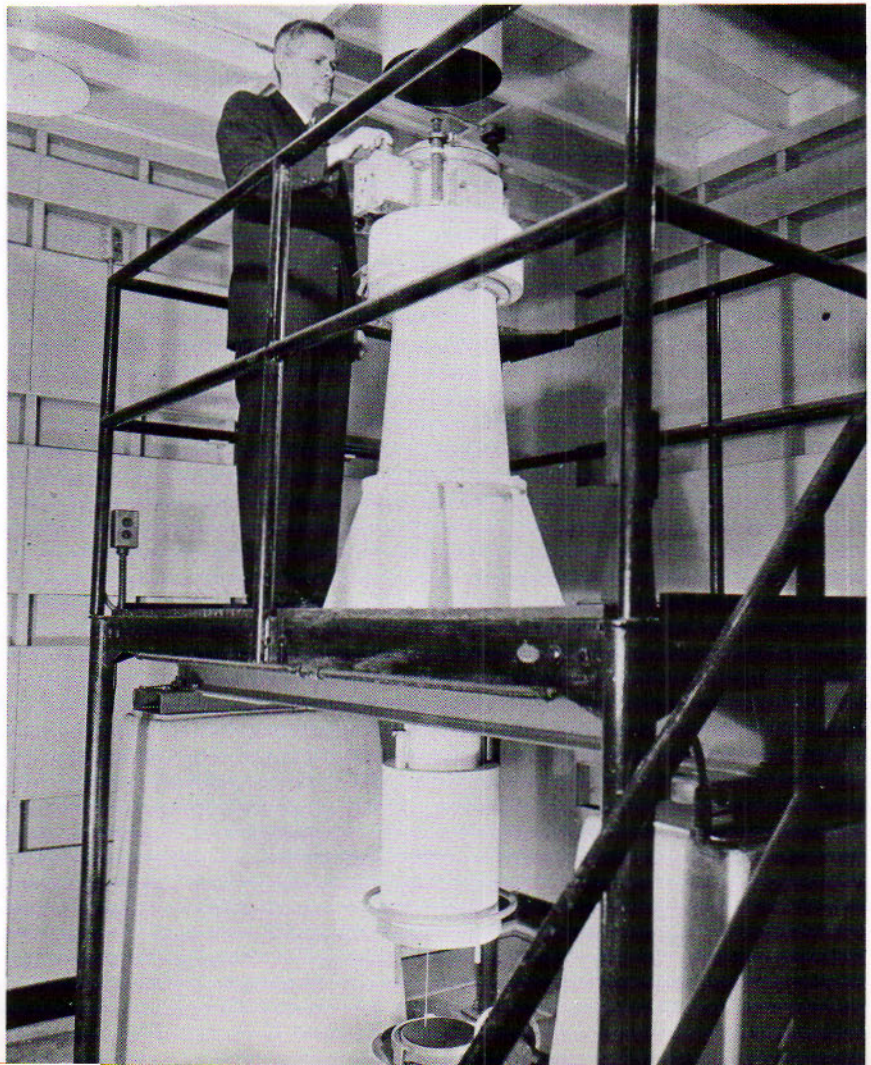


Figure 3.4B
The 8-inch photographic zenith tube at the U. S. Naval Observatory. *(Official U.S. Navy photograph)*

nearly overhead. The cone of starlight from the objective is reflected from a mercury surface before coming to focus on a small photographic plate below the lens. With this instrument an error of only about $0^s.003$ is obtained in a time determination from a set of 18 stars.

3.5 Apparent Solar Time

Although sidereal time is suited to certain activities of the observatories, it is not useful for civil purposes, because our daily affairs are governed by the sun and not by the vernal equinox. Sidereal noon, for example, comes at night during half of the year.

The *apparent sun* is the sun we see. The local hour angle of its center plus 12 hours is the *apparent solar time*, or simply, apparent time. The apparent solar day begins at apparent midnight and is reckoned through 24 hours continuously. However, the sun itself is not a uniform timekeeper; it runs fast or slow of a regular schedule, at times nearly half a minute a day. The sundial is the only timepiece adapted to the sun's erratic behavior.

Two factors contribute chiefly to the irregularity in the length of the apparent solar day: (1) The variable speed of the earth's revolution, because of the eccentricity of its orbit; and (2) the inclination of the ecliptic to the celestial equator. We are not considering here the further irregularity caused by variations in the rate of the earth's rotation.

(Length of Day, 1971)

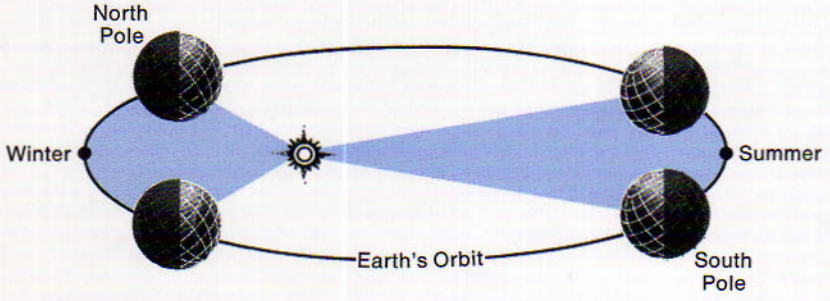
1 Jan	24^h	00^m	28^s
1 Feb	24	00	8
1 Mar	23	59	48
1 Apr	23	59	42
1 May	23	59	53
1 Jun	24	00	9
1 Jul	24	00	12
1 Aug	23	59	56
1 Sep	23	59	41
1 Oct	23	59	41
1 Nov	23	59	58
1 Dec	24	00	22

3.6 Effect of the Earth's Variable Revolution

Because the earth's orbit is an ellipse, the speed of the earth in its revolution around the sun is not uniform (7.25). The closer it is to the sun, the faster the earth revolves. This relation is expressed precisely as a particular case of Kepler's law of areas (7.15): The line joining the earth to the sun sweeps over equal areas in equal intervals of time.

When the earth is near perihelion, early in January, it revolves farthest in a day. Because the difference in length between the solar day and the

Figure 3.6
The earth's variable revolution. Since, by Kepler's law of areas, the line joining the earth and sun sweeps over the same area every day, the earth must revolve farther in a day when it is nearer the sun.



more nearly constant sidereal day depends on the daily progress of the earth's orbital motion (3.2), apparent solar days would then have the maximum length. When the earth is near aphelion, early in July, it makes the smallest advance in its orbit in a day. Apparent solar days would then have the minimum length if this were the only cause of their variation.

This effect can also be considered by shifting the attention from the earth's revolution around the sun to the consequent apparent motion of the sun along the ecliptic. It is this motion that delays the sun's return to the meridian from day to day. The farther the earth revolves in a day, the greater is the sun's displacement along the ecliptic and the greater for this reason would be the length of the apparent solar day.

3.7 Effect of the Obliquity of the Ecliptic

Even if the sun's motion in the ecliptic were uniform, the length of the apparent solar day would still be variable, because the ecliptic is inclined to the celestial equator. It is the projection of the sun's eastward motion upon the celestial equator that determines the delay in completing the apparent solar day.

At the equinoxes, a considerable part of the sun's motion in the ecliptic is north or south, which does not delay the sun's return. At the solstices, where the ecliptic is parallel to the equator, the entire motion is eastward; moreover, the hour circles are closer together along the ecliptic here. Therefore, as far as the obliquity of the ecliptic is concerned, apparent solar days would be shortest at the equinoxes and longest at the solstices. Both factors act to make the apparent day longest in winter.

3.8 Mean Solar Time

The *mean solar day*, as its name implies, is the average apparent solar day; its length varies slightly because the earth's rotation is not precisely uniform (2.18). The mean solar day is the interval between two successive upper or lower transits of the mean sun. The *mean sun* is an imaginary point that moves eastward in the celestial equator, completing its circuit of the heavens in the same period as that of the apparent sun in the ecliptic. *Mean solar time* is time by the mean sun. The mean solar second, which is 1/86,400 part of the mean solar day, was formerly the basic unit for all measurements of time.

Mean solar time is reckoned from the beginning of the day at local midnight through 24 hours to the following midnight. The local mean time at a particular place is 12 hours plus the local hour angle of the mean sun. The present usage was adopted in 1925 for astronomical reckoning, which had previously begun the day at noon.

Mean time is commonly reckoned in two 12-hour divisions of the day with the designations A.M. and P.M., but the preference for its continuous reckoning through 24 hours is found in various places. In astronomical

practice, 6:30 P.M., is often recorded as $18^h\ 30^m$, and in the operation of ships and airplanes as 1830. Very often times are given in terms of the mean time at the zero meridian (Greenwich), such time is referred to as Universal Time. We shall also see later that the motion of the earth in its orbit and the finite velocity of light cause systematic shifts in the time of events occurring outside the solar system. To circumvent this, times of observation are reduced to the center of the sun and are referred to as *Heliocentric time.* In the astronomical literature, times that are given, unless specifically indicated to the contrary, may safely be assumed to be Heliocentric Universal Time.

3.9 The Equation of Time

The difference at any instant between apparent and mean solar time is called the equation of time, thus the equation of time equals apparent solar time minus mean solar time. Four times a year the two agree. At other times the apparent sun is either fast or slow in the westward diurnal motion. Early in November the sundial is more than a quarter of an hour ahead of local mean time. Table 3.I gives the value of the equation of time at midnight at Greenwich in 1973 or, very nearly, in any other year.

TABLE 3.I
Equation of Time (1973)
(Apparent time faster or slower than local mean time)

Jan.	0*	2m	54s	slow	July	0	3m 23s	slow
	10	7	23	slow		10	5 08	slow
	20	10	57	slow		20	6 12	slow
Feb.	0	13	27	slow	Aug.	0	6 18	slow
	10	14	19	slow		10	5 21	slow
	20	13	53	slow		20	3 28	slow
Mar.	0	12	44	slow	Sep.	0	0 28	slow
	10	10	32	slow		10	2 49	fast
	20	7	44	slow		20	6 21	fast
Apr.	0	4	24	slow	Oct.	0	9 48	fast
	10	1	30	slow		10	12 48	fast
	20	0	58	fast		20	15 03	fast
May	0	2	45	fast	Nov.	0	16 18	fast
	10	3	39	fast		10	16 05	fast
	20	3	37	fast		20	14 28	fast
Jun.	0	2	33	fast	Dec.	0	11 30	fast
	10	0	50	fast		10	7 26	fast
	20	1	15	slow		20	2 39	fast
						30	2 18	slow

* Zero date is an astronomical convention for the last day of the previous month.

The rapid change in the equation of time near the beginning of the year has an effect that can be noticed by everyone. At this time of year the earth is nearest the sun and is accordingly revolving fastest. The sun is then moving eastward along the ecliptic fastest, delaying its rising and setting as timed by the mean sun. For this reason the sun does not begin to rise earlier in the morning by our watches until about 2 weeks after the date of the winter solstice, although it begins to set later in the evening 2 weeks before that date.

3.10 Universal Time and Ephemeris Time

Universal time is the local mean solar time at the meridian of Greenwich; it is based on the earth's rotation. This is the kind of time in ordinary use and will remain so, because the observatory clocks that distribute this time will continue to be corrected for irregularities in the earth's rotation by frequent sights on the stars. For the foretelling of celestial events, however, the irregular rate of the rotation makes it impossible to predict with very high precision what the universal times of these events will be.

Beginning with the issues for the year 1960, the *American Ephemeris and Nautical Almanac* and the British *Astronomical Ephemeris*, which now conform in other respects as well, tabulate the fundamental positions of the sun, moon, and planets at intervals of ephemeris time.

Ephemeris time runs on uniformly; its constant arbitrary unit equals the length of the tropical year at the beginning of the year 1900 divided by 31,556,925.9747, which was the number of seconds and fraction in the year at that epoch. This invariable unit of time is adopted as the fundamental unit by the International Committee of Weights and Measures. Corrections for converting ephemeris time are determined frequently by observing the universal times when certain celestial bodies arrive at positions in the sky predicted on ephemeris time. In actual practice, observations of the moon's positions among the stars (5.12) seem to be the most suitable. In the present century, ephemeris time has been gaining on universal time and in 1970 was ahead by 40 seconds.

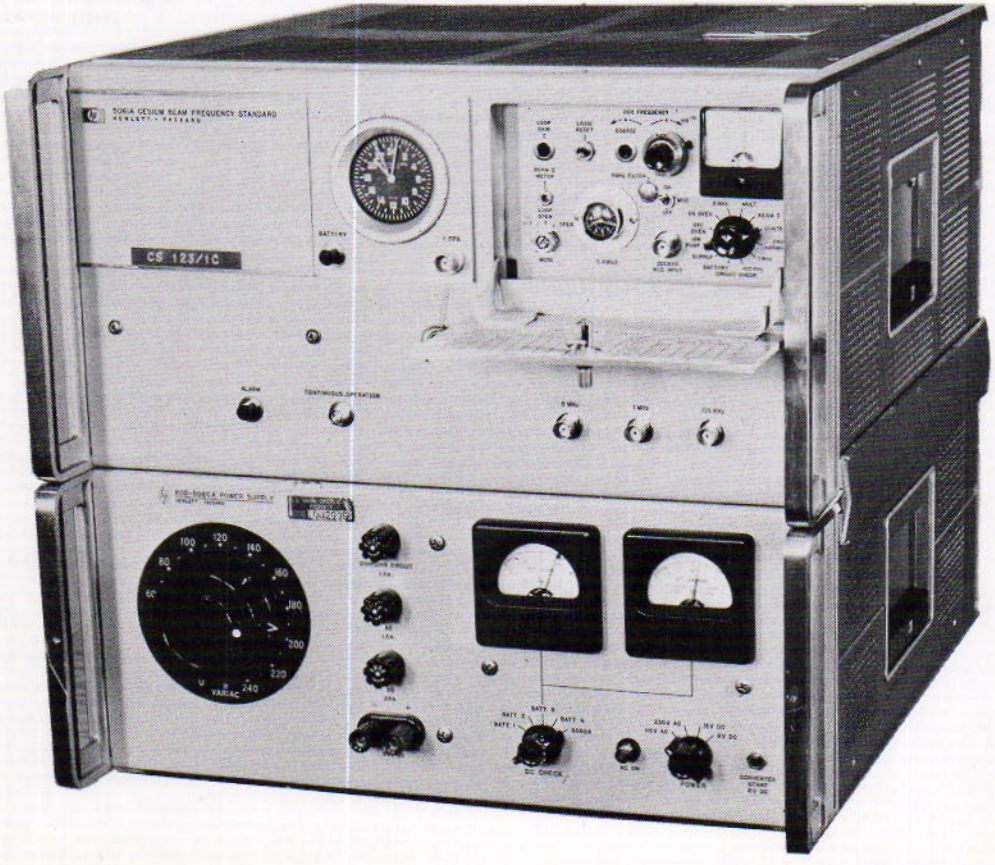
3.11 Atomic Time

The household electric clock using a synchronous electric motor is an example of the replacement of the pendulum and escapement mechanisms of keeping time by a simple and reliable control of the rate. In this case it is the 60Hz alternating current carefully controlled at the generating station that provides the control.

More accurate measurement of intervals of time can be achieved by means of the quartz-crystal clock. In such a clock a quartz crystal vibrates at its natural frequency, the resulting electrical effect is amplified and the ensuing power is used to run a synchronous motor with a small amount of power being used to restore energy to the crystal to keep it vibrating. Such clocks have errors on the order of 1 second per decade.

Even more accurate clocks are the atomic clocks, two kinds of which are in use; one uses the natural frequency of the cesium atom and the other a natural frequency of the ammonia molecule. In the latter type, the ammonia molecules will absorb microwave radiation over a very narrow range in frequency. A microwave system with a tunable generator (oscillator) at this frequency radiates a container of ammonia molecules with a receiver at the other side. When the energy received is at a minimum the generator is exactly on the ammonia frequency. If it drifts off, the molecules absorb less and the receiver observes an increase in energy which causes the generator to vary its frequency back to the minimum. The generator then is at a known frequency and makes a good clock. These clocks achieve such an accuracy that their error is only 1 second in 50 years.

Figure 3.11
A cesium beam atomic clock. With an accuracy of one millionth of a second over a period of three weeks, this clock is a recent major advance in timekeeping. *(Official U.S. Naval Observatory photograph)*



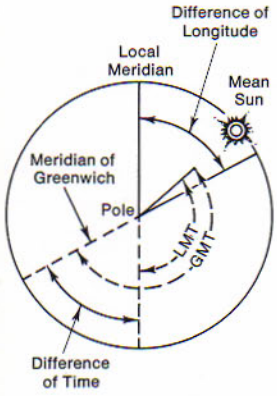


Figure 3.12
Difference of local time between two places equals their difference of longitude.

The cesium clock is even more accurate, having an error of only about 1 sec. in 1000 years. This clock makes use of the fact that the cesium atom acts like a dipole magnet. When the atoms pass through a magnetic field their poles align with the field and point up or down. Then they pass through two spaced magnets whose fields are changed opposite to each other by an alternating oscillator. If the frequency of the oscillator is exactly right (making it the clock) the alignment of the atoms is exactly the same as before they entered the oscillating fields and a fixed magnet can now sort them out in such a way as to give a maximum signal on a detector. If the frequency of the oscillator drifts, the detected signal drops and causes the oscillator to retune to the maximum signal again.

At first, atomic clocks were large cumbersome laboratory devices. Now, clocks accurate to 1 second in 50 years are of such a size as to be portable by commercial aircraft. Because of this, two or more clocks can be synchronized at one point and flown thousands of miles for timing events. Such developments have made long baseline interferometry possible as we shall see in the next chapter and have contributed to the mounting excitement pervading modern astronomy.

3.12 Difference of Local Time Equals Difference of Longitude

In any one of the three kinds of local time, a day of 24 hours is completed when the earth has made one rotation, through 360°, relative to the point in the sky that serves as the time reckoner. Twenty-four hours of that kind of time equal 360° or 24 hours of longitude on the earth; and there is a difference of 15° or 1 hour between the local times of two places that differ by 15° or 1 hour in longitude.

Thus the difference of local times of the same kind between two places equals the difference of their longitudes. Transposed, this is the basis of longitude determinations.

When the local time at one place is given and the corresponding local time at another place is required, add the difference of their longitudes if the second place is east of the first; subtract if it is west. In Fig. 3.12 the observer is in longitude 60° W., or 4^h W. The local mean time (LMT) is 9^h, and the Greenwich mean time (GMT), or universal time, is 13^h, so that the difference of 4 hours equals the difference between the observer's longitude and that of Greenwich.

3.13 Zone Time

Because difference of local time equals difference of longitude, local mean time becomes progressively later with increasing distance east of us and earlier west of us. The inconvenience of continually resetting our watches as we travel east or west is avoided by the use of zone time at sea and standard time on land. These are conventional forms of mean time.

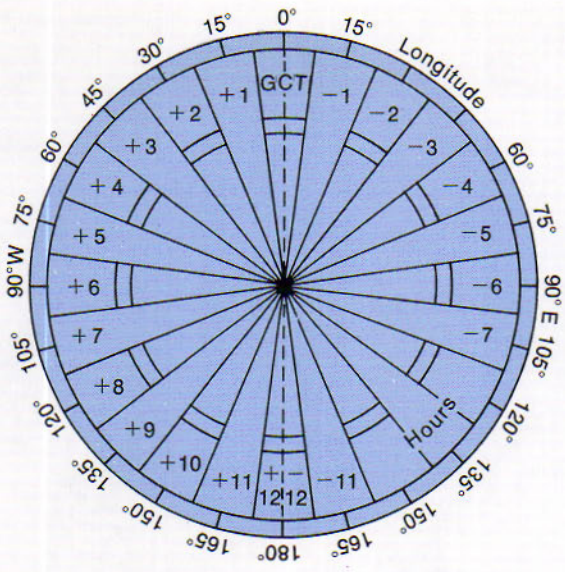


Figure 3.13
Time zone diagram. The numbers outside the circle are the longitudes of the standard meridians. The numbers inside are the corrections in hours from zone time to universal time.

Standard meridians are marked on the earth at intervals of 15° or 1 hour in both directions from the meridian of Greenwich. The local mean time of the standard meridian is the time to be kept at sea by timepieces in the entire zone within 7½° east and west of that meridian if the plan is followed. *Zone time* for any place is accordingly the local mean time of the standard meridian nearest the place. Thus the earth is divided into 24 zones, in which the time differs from universal time by whole hours. There are certain exceptions to the rule; a time zone is occasionally modified near land to correspond with the time kept ashore.

3.14 Standard Time

Standard time on land conforms in a general way with zone time at sea, but its divisions are less uniform. The boundaries of the standard time zones have been determined by local preference and may be altered by legislative action, subject in the United States to approval by the Interstate Commerce Commission. Moreover, the plan is not completely adopted in all parts of the world; the legal times in certain land areas differ from those in adjacent divisions by a fraction of an hour. Occasionally clocks are set ahead by one or even two hours. Such times are called "saving" times, though it is not clear what is being saved. For astronomers, the middle of the night does not change, so they are forced more or less to remain on standard time. One wonders why working hours aren't adjusted rather than setting clocks ahead and back every year.

Figure 3.14A
Time zones in the Western Hemisphere. The time for each zone at a hypothetically given hour is shown on top, the numbers below show the corrections in hours from zone time to universal time. Note also zone adjustments.

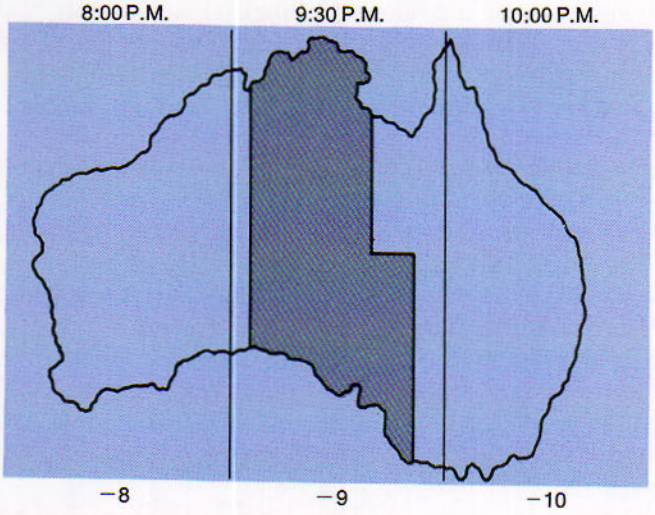


Figure 3.14B
Time zones in Australia.

All time divisions are shown in the U. S. Oceanographic Office Time Zone Chart of the World. It is available from the U. S. Government Printing Office for a nominal fee. This chart shows how civilization adjusts its idealized time system to suit its own convenience. The most striking example of this perhaps is seen in Australia, which spans three time zones (−8 to −10) but has readjusted the zones to suit its population pattern. Western Australia is eight hours east of Greenwich and uses this as its zone time. Eastern Australia is ten hours east of Greenwich and also uses this as its zone time, but both zones are extended slightly into the intermediate or central area, which would normally be the ninth-hour zone and which is very sparsely populated. This is not the only change, however. In order to make the transition from time in the central area to the eastern area (where most of Australia's population resides) less abrupt, Central Australia is considered as a 9½-hour zone.

Four standard times are employed in the continental parts of the United States and Canada, namely, Eastern, Central, Mountain, and Pacific standard times; they are respectively the local mean times of the meridians 75°, 90°, 105°, and 120° west of the Greenwich meridian and are therefore 5, 6, 7, and 8 hours slow as compared with universal time. Newfoundland standard time is 3½ instead of 4 hours slow. Yukon standard time is 9 hours slow, and Alaska standard time is 10 hours slow of universal time.

Certain changes for Canada are shown in the Dominion Operations Map of 1958. Apart from Quebec, Ontario, and the Northwest Territories, each division of Canada has adopted a single standard time.

3.15 The International Date Line

In successive time zones west of the zone including Greenwich, the time becomes progressively earlier than universal time. In successive zones east of the Greenwich zone the time becomes later. Finally, in the zone including the 180° meridian the western half differs by a whole day from the eastern half. The rule for ships and for airplanes over the ocean is to change the date at the 180° meridian. At the westward crossing of this meridian the date is advanced a day, and at the eastward crossing it is set back a day. Thus if the eastward crossing is made on Tuesday, June 7, the date becomes Monday, June 6.

For the land areas in this vicinity the boundary between the earlier and later dates is the *international date line.* This line departs in places from the 180° meridian so as not to divide politically associated areas; it bends to the east of the 180° meridian around Siberia and to the west around the Aleutian Islands.

3.16 Radio Time Service

Time signals based on Naval Observatory time determinations are transmitted from 4 naval radio stations and 2 stations of the National Bureau of Standards. The signals are seconds of universal time, corrected for predicted seasonal variations in the speed of the earth's rotation.

The naval radio stations at Annapolis (NSS) and elsewhere transmit the time signals during the last 5 minutes preceding hours scheduled in the Appendix to *U.S. Naval Observatory Circular* Number 49 (1954). The schedule is subject to change. The signals are on a variety of frequencies and represent the seconds beats of the crystal-controlled transmitting clock. These are on continuous waves and can be heard only with receivers suited to code reception. The signals are omitted at certain seconds so that the listener can readily identify the minute and second of each signal. The beginning of a long dash following the longest break announces the precise beginning of the hour.

The seconds signals from the standard frequency station WWV at Fort Collins, Colorado and WWVH in Hawaii, are transmitted continuously day and night on frequencies of 2.5, 5, 10, 15, 20, and 25 MHz. They are heard as clicks superposed on standard audio frequencies; they employ modulated waves and can therefore be heard with ordinary shortwave radiophone receivers. The seconds clicks are interrupted at the 59th second of each minute. At 2 minutes before the beginning of the hour and at each multiple of 5 minutes thereafter, the audio tone is interrupted for exactly 2 minutes for making announcements. In the United States the signal is available by telephone [303-447-1192].

The Dominion Observatory at Ottawa supplies continuous time service from three different transmitters on frequencies of 3330, 7335, and 14,670 KHz, respectively, adjacent to the amateur bands. The seconds signals are omitted each minute on the 29th second and from the 51st to the 59th second inclusive, when a voice announcement of the time is given. More than 60 broadcast stations across Canada transmit the Dominion Observatory time signal at 1 P.M., EST each day. These are examples of the distribution of the correct time by observatories in various parts of the world.

THE CALENDAR

Calendars have been employed since the beginnings of civilizations as registers on which to record past events and to predict future ones that occur regularly. They have undertaken to combine two or three natural measures of time, the solar day, the lunar month, and the year of the seasons, in the most convenient ways; and they have frequently encountered difficulties, because these measures do not fit evenly one into another. Calendars have generally been of 3 types: lunar, lunisolar, and solar.

3.17
Three Types of Calendars

The lunar calendar is the simplest of the 3 types and was the earliest used by almost all nations. Each month began originally with the "new moon," the first appearance for the month of the crescent moon after sunset. Long controlled only by observation of the crescent, this calendar was eventually operated by fixed rules. In the fixed lunar calendar the 12 months of the lunar year are alternately 30 and 29 days long, making 354 days in all and thus having no fixed relation to the year of the seasons. The Mohammedan calendar is a survivor of this type.

The lunisolar calendar tries to keep in step with the moon's phases and the seasons and is the most complex of the 3 types. It began by occasionally adding a 13th month to the short lunar year to round out the year of the seasons. The extra month was later inserted by fixed rules. The Jewish calendar is the principal survivor of this type.

The solar calendar makes the calendar year conform as nearly as possible to the year of the seasons and neglects the moon's phases; its 12

months are generally longer than the lunar month. Only a few early nations, notably the Egyptians and eventually the Romans, adopted this simple type.

3.18 The Early Roman Calendar

This calendar dates formally from the traditional founding of Rome in 753 B.C. It was originally a lunar calendar of a sort, beginning in the spring and having 10 months. The names of those months, if we use mainly our own style instead of the Latin, were: March, April, May, June, Quintilis, Sextilis, September, October, November, and December. The years for many centuries thereafter were counted from 753 and were designated A.U.C., in the year of the founding of the City. Two months, January and February, were added later and were eventually placed at the beginning, so that the number months have since then appeared in the calendar out of their proper order.

In the 12-month form the Roman calendar was of the lunisolar type. The day began at midnight instead of at sunset as it did with most early people. An occasional extra month was added to keep the calendar in step with the seasons. The calendar was managed so unwisely, however, by those in authority that it fell into confusion; its dates drifted back into seasons different from the ones they were supposed to represent.

When Julius Caesar became the ruler of Rome, he was disturbed by the bad condition of the calendar and took steps to correct it. He particularly wished to discard the lunisolar form with its troublesome extra months. Caesar was impressed with the simplicity of the solar calendar the Egyptians were using, and he knew of their discovery that the length of the tropical year is very nearly 365¼ days. The Egyptian calendar dating from 4236 B.C. used a 360-day year by having 12 months of 30 days each and then adding 5 feast days to the end of the year. Caesar accordingly formulated his reform with the advice of the astronomer Sosigenes of Alexandria. In preparation for the new calendar the "year of confusion," 46 B.C., was made 445 days long in order to correct the accumulated error of the old one. The date of the vernal equinox was supposed to have been brought thereby to March 25. The Julian calendar began on January 1, 45 B.C.

It is interesting to note that the Mayas had a similar calendar having 18 months of 20 days each with a five day year end week. The Aztecs realized the year was actually 365¼ days and brought the Maya calendar into line with the seasons by making an intercalation of 12½ days every 52 years, a feat the Egyptians never accomplished.

3.19 The Julian Calendar

This calendar was of the solar type, and so neglected the moon's phases. Its chief feature was the adoption of 365¼ days as the average length of the calendar year. This was accomplished conveniently by the plan of leap years. Three common years of 365 days were followed by a fourth

year containing 366 days; this *leap year* in our era has a number evenly divisible by 4.

In lengthening the calendar year from 355 days of the old lunisolar plan to the common year of 365 days, Caesar distributed the additional 10 days among the months. With further changes made in the reign of Augustus, the months assumed their present lengths. After Caesar's death, in 44 B.C., the month Quintilis was renamed July in honor of the founder of the new calendar. The month Sextilis was later renamed August in honor of Augustus.

Because its average year of $365^{d}\ 6^{h}$ was $11^{m}\ 14^{s}$ longer than the tropical year, the Julian calendar fell behind with respect to the seasons about 3 days in 400 years. When the council of churchmen convened at Nicaea in A.D. 325, the vernal equinox had fallen back to about March 21. It was at this convention that previous confusion about the date of Easter was ended.

3.20 Easter

Easter was originally celebrated by some churches on whatever day the Passover began and, by others, on the Sunday included in the Passover week. The Council of Nicaea decided in favor of the Sunday observance and left it to the church at Alexandria to formulate the rule, which is as follows:

Easter is the first Sunday after the 14th day of the ecclesiastical moon (nearly the full moon) that occurs on or immediately after March 21. Thus if the 14th day of the moon occurs on Sunday, Easter is observed one week later. Unlike Christmas, Easter is a movable feast because it depends on the moon's phases; its date can range from March 22 to April 25 and in the epoch year of 2000 it will fall on April 23. There is some interest expressed by church authorities to make Easter a fixed holiday, a view that is enthusiastically supported by businessmen. It would almost be a pity to lose the variety that the moving of Easter gains, a reason perhaps behind the general resistance to calendar reform as well (3.22).

Dates of Easter Sunday

1970 Mar 29	1980 Apr 6	1990 Apr 15
1971 Apr 11	1981 Apr 19	1991 Mar 31
1972 Apr 2	1982 Apr 11	1992 Apr 19
1973 Apr 22	1983 Apr 3	1993 Apr 11
1974 Apr 14	1984 Apr 22	1994 Apr 3
1975 Mar 30	1985 Apr 7	1995 Apr 16
1976 Apr 18	1986 Mar 30	1996 Apr 7
1977 Apr 10	1987 Apr 19	1997 Mar 30
1978 Mar 26	1988 Apr 3	1998 Apr 12
1979 Apr 15	1989 Mar 26	1999 Apr 4

The week of 7 days was introduced in the Roman calendar in the year 321 by the emperor Constantine. The Christian Era for the recording of dates forward and back from about the time of the birth of Christ is said to have been introduced in the 6th century, but did not replace the Roman plan generally until several centuries later.

3.21 The Gregorian Calendar

As the date of the vernal equinox fell back in the calendar, March 21 and Easter, which is reckoned from it, came later and later in the season. Toward the end of the 16th century the equinox had retreated to March 11. Another reform of the calendar was proposed formally by Pope Gregory XIII.

Two rather obvious corrections were made in the Gregorian reform. First, 10 days were suppressed from the calendar of that year; the day following October 4, 1582, became October 15 for those who wished to adopt the new plan. The date of the vernal equinox was restored in this way to March 21. The second correction made the average length of the calendar year more nearly equal to the tropical year, so that the calendar would not again get seriously out of step with the seasons. Evidently the thing to do was to omit the 3 days in 400 years by which the Julian calendar year was too long. This was done conveniently by making common years of the century years having numbers not evenly divisible by 400.

Thus the years 1700, 1800, and 1900 became common years of 365 days instead of leap years of 366 days, whereas the year 2000 will remain a leap year as in the former calendar. The average year of the new calendar is still too long by 26 seconds, which is hardly enough to be troublesome for a long time to come. It amounts to almost one day in 3300 years.

The Gregorian calendar was gradually adopted, until it is now in use, at least for civil purposes, in practically all nations. England and its colonies including America made the change in 1752. By that time there were 11 days to be suppressed, for that century year was a leap year in the old calendar and a common one in the new calendar. September 2 was followed by September 14. The countries of eastern Europe were the latest to make the change, when the difference had become 13 days. Russia briefly adopted the Gregorian Calendar in 1918 and then discarded it until 1940.

3.22 Suggested Calendar Reform

Irregularities in our present calendar are frequently cited as reason for reforming it further. The calendar year is not evenly divisible into quarters; the months range in length from 28 to 31 days, and their beginnings and ending occur on all days of the week; the weeks are split between months. Some people believe that the irregularities should be

corrected. Others are not sure that the improvement would be great enough to offset the confusion in our records that the change might bring.

Recent proposals for calendar reform are based on the period of 364 days, a number evenly divisible by 4, 7, and also 13, and the addition of two stabilizing days. An extra day might be added each year in such a way as not to disturb the sequence of weekdays, and another might be added every four years in the same manner except in the century years not evenly divisible by 400. Two proposed calendars have been the 13-month perpetual calendar and the 12-month perpetual calendar known as the World Calendar.

The first plan divides the year into 13 months of 28 days each. In this plan a calendar for one month would serve for every other month forever if it is remembered when to add the two stabilizing days. This proposal met with considerable approval for a time, but lost favor because it seemed too drastic a change from the present calendar.

The second plan divides the 364-day period into 12 months. The four equal quarters of the year remain the same forever. Each quarter begins on Sunday and ends on Saturday. Its first month has 31 days, and its second and third months have 30 days each. One stabilizing day is added each year at the end of the fourth quarter; it is called Year-End Day, December Y, and is an extra Saturday. The second extra day is added every fourth year, with the usual exceptions, at the end of the second quarter; it is called Leap-Year Day, June L, and is also an extra Saturday. This plan is a more moderate change from the present calendar. It was disapproved by the United Nations in 1956.

3.23 Julian Day

Many astronomical observations are recorded in decimals of a day and it is convenient to have a whole night's observations recorded for the same day rather than split between two days. The Julian Day is the number of days that have elapsed since the beginning of an arbitrary zero day at noon Greenwich mean time on January 1, 4713 B.C. on the Julian Calendar. The term is thus derived from this calendar but is not named after Julius Caesar. When the interval between two events is required, especially where they are widely separated in time, it is easier to take the difference between the Julian dates than to refer to a calendar and worry about leap years, etc. The Julian Day numbers are tabulated for each year in the American Ephemeris and Nautical Almanac. J.D. 2,440,587.5 corresponds to January 1, 1970 at Greenwich mean midnight.

REVIEW QUESTIONS

1. Show why the solar day is longer than the sidereal day.
2. What is meant by the apparent sun?
3. Give the three kinds of local time.

4. What type of time is related by a constant to atomic time?
5. Fomalhaut, α, and β Pegasus have very nearly the same right ascension, this makes the line joining them a useful and easy time reckoner. When this line defines the meridian on September 21 what is your approximate local time?
6. If the length of days differ at most by ½ a minute, how is it that the sundial can lag as much as 16 minutes behind local mean time?
7. What factors cause the equation of time?
8. How does the diurnal motion of the sun differ from that of the stars?
9. Where does a given date first appear on the earth?
10. What is the chief advantage in the use of Julian days and decimals thereof?
11. Why is the Gregorian Calendar such an improvement over the Julian Calendar?
12. In this chapter we have, by implication, discussed two types of time, epoch and interval. Give two examples of each.

REFERENCES

Cowan, Harrison J., *Time and its Measurement.* New York: World Publishing Co., 1958.

Glasstone, Samuel, *Sourcebook on the Space Sciences.* Princeton: D. van Nostrand Co., 1965.

U.S. Government, Naval Observatory, *The Air Almanac,* annual. Washington, D.C.: Superintendent of Documents, U.S. Government Printing Office.

The Nautical Almanac, annual. Same as above.

The American Ephemeris and Nautical Almanac, annual. Same as above.

Wilson, Philip W. *The Romance of the Calendar.* New York: Norton Co., 1937.

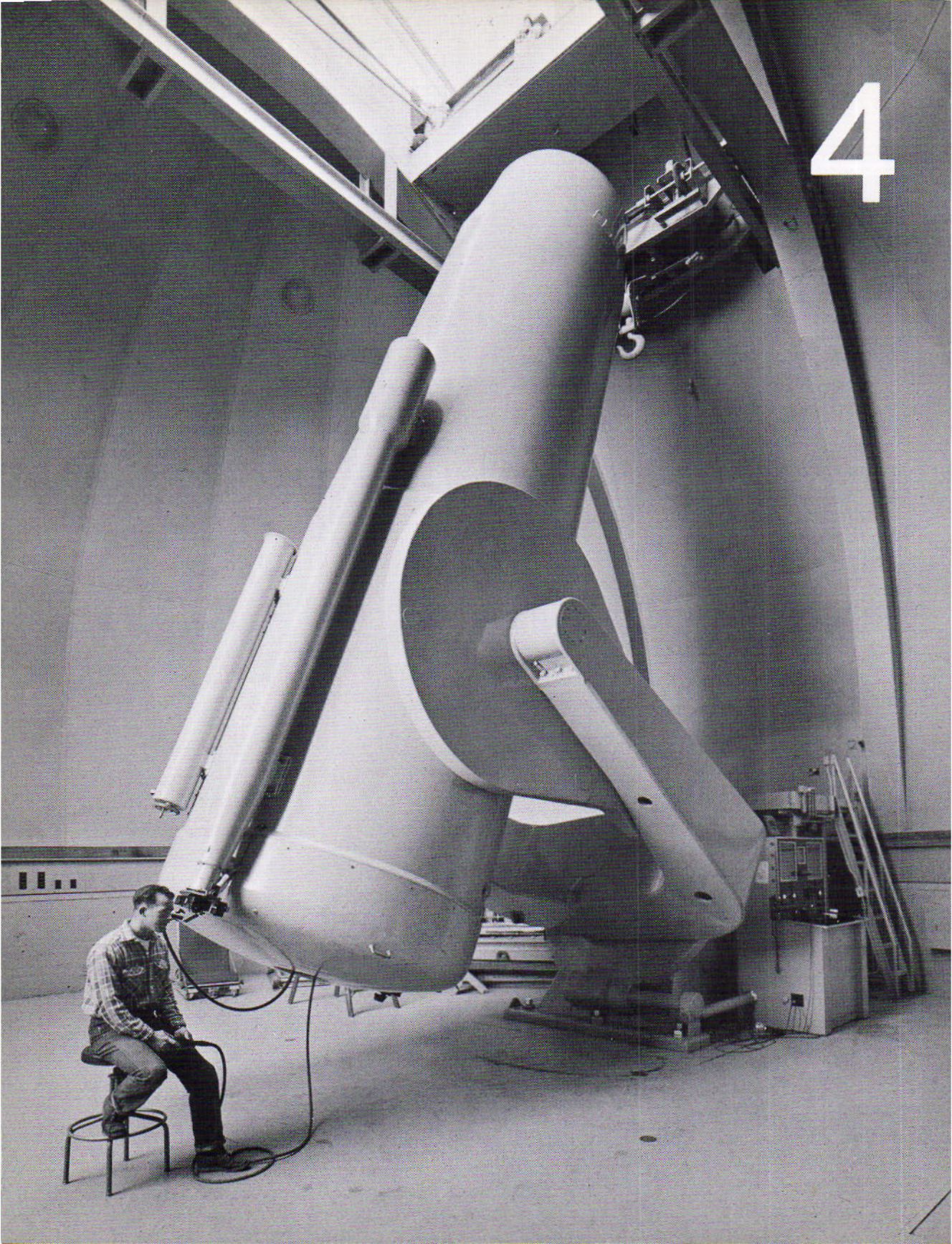
FOR FURTHER STUDY

Brower, Richard and Gerald Clemence, *Methods of Celestial Mechanics,* New York, Academic Press, 1961.

Smart, William M., *Text-book on Spherical Astronomy,* 5th ed. New York: Cambridge University Press, 1962. (also available in paperback).

Woolard, Edgar W. and Gerald M. Clemence, *Spherical Astronomy,* New York: Academic Press, 1966.

Panoramic view of Mount Wilson Observatory site. *(Photograph from the Hale Observatories)*


4

RADIATION AND THE TELESCOPE

REFRACTION OF LIGHT — DISPERSION OF LIGHT; THE SPECTRUM — THE REFRACTING TELESCOPE — THE REFLECTING TELESCOPE — THE RADIO TELESCOPE

The principal feature of the optical telescope is its objective, which receives the light of a celestial body and focuses the light to form an image of the object. The image may be formed either by refraction of the light by a lens or by its reflection from a curved mirror. Optical telescopes are accordingly of two general types: refracting telescopes and reflecting telescopes. In the radio telescopes the radiations in radio wavelengths from a celestial source are collected by an antenna, from which they are conveyed to receiving and recording apparatus.

This chapter begins with some features of light, particularly its refraction and dispersion into a spectrum.

48-inch (112 cm) Schmidt telescope at Mount Palomar. *(Photograph from the Hale Observatories)*

REFRACTION OF LIGHT

Radiation crossing the boundary between two media of different densities at an angle is refracted *or bent. This characteristic introduces problems in the interpretation of the information obtained from any given radiation after it passes through the ionosphere, the atmosphere, and the planetary and interstellar medium. By the same token, it is the characteristic that allows us to build some of the telescopes and instruments for collecting and analyzing the radiation.*

4.1 Light

Light is propagated in waves which emerge from a source such as a lamp or the sun not unlike the way that ripples spread over the surface of a pond when a stone is dropped into the water. When light waves of appropriate lengths enter the eye, they produce the sensation we call light in its different colors from violet to red. The visual effect is caused by a limited range of lengths in the waves that come to us.

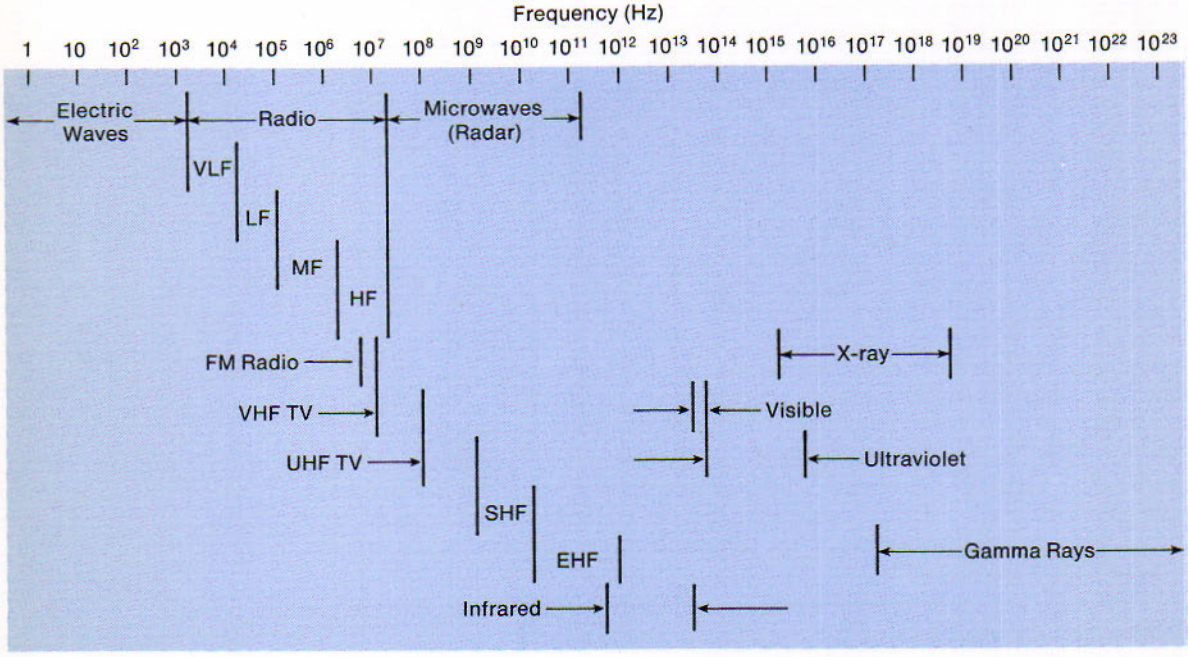


Figure 4.1 Radiation in a variety of wavelengths (exponential scale).

Radiation in general emerges from a source in waves of a wide variety of lengths, from gamma rays having billions of wave crests to the centimeter to radio waves as long as several kilometers from crest to crest. All radiation has the constant "velocity of light" in empty space.

The *wavelength* λ of the radiation is the distance between the same phase of successive waves, as from crest to crest. The lengths to which the eye is sensitive range from 4×10^{-5} cm, for extreme violet light to nearly twice that length for the reddest light we can see. The photographic range extends into the ultraviolet and infrared.

The *velocity of light* c is about 300,000 km/sec. (186,300 miles/sec.). This is the speed in a vacuum.* The speed is reduced in a medium such as air or glass, depending on the density of the medium; and in the same medium the reduction is greater for shorter wavelenghts than for longer ones.

The *frequency* μ of the radiation is the number of waves emitted by the source in a second; it equals the velocity of light divided by the wavelength. Thus the frequency of violet light having a wavelength of 4×10^{-5} cm is 7.5×10^{14} Hz.

1 Hz = 1 cycle/second

4.2 Refraction of Light

A *ray of light* denotes the direction in which any portion of the wave system is moving. It is often convenient to picture rays of light as radiating in all directions from a source and continuing always in straight lines as long as they remain in the same homogeneous medium. Thus light is said to travel in straight lines.

When a ray of light passes from a rarer to a denser medium, as from air into glass, it proceeds through the denser medium with reduced speed. If the ray falls obliquely on the surface of the glass, the part of each wave front on one side of the "ray" enters the glass and has its speed reduced before the other side enters. The front is therefore swung around and the ray changes direction (Fig. 4.2). The parallel lines of the figure represent the progress of the wave after equal intervals of time.

Refraction of light is the change in the direction of a ray of light when it passes from one medium into another of different density. The change is toward the perpendicular to the boundary if the second medium (having an index n') is the denser, and away from the perpendicular if it is the less dense (having an index n). When the ray enters the second medium at right angles to its surface, there is no refraction, for all parts of the wave front enter and are retarded together.

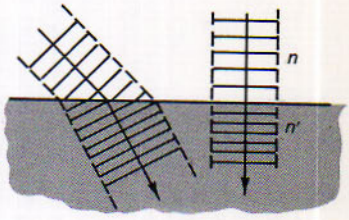


Figure 4.2
Refraction of light. A ray of light passing from a rarer into a denser medium is refracted toward the perpendicular to the boundary between them.

* We are using speed and velocity here as more or less interchangable terms. Speed is a distance covered in a given interval of time while technically velocity is a distance covered in a given interval of time in a specified direction.

4.3 Refraction Increases the Altitude of a Star

When a ray of starlight enters the atmosphere, it is refracted downward according to the rule we have noted; and the bending continues until the earth's surface is reached, because the density of the air increases downward. The point in the sky from which the light appears to come is therefore above the star's true direction. Atmospheric refraction elevates the celestial bodies by amounts which depend on the distance from the zenith. The relations are as follows:

Zenith Distance	Refraction	Zenith Distance	Refraction
0°	0′00″	85°	9′45″
20	0 21	86	11 37
40	0 48	87	14 13
60	1 40	88	18 06
70	2 37	89	24 22
80	5 16	90	34 50

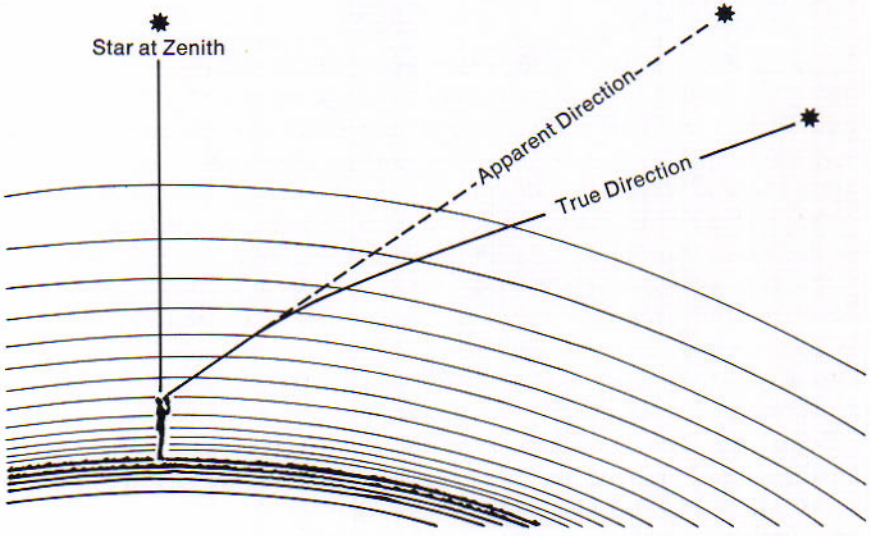


Figure 4.3
Atmospheric refraction increases the star's altitude. As the ray of starlight is bent down in passing through the air, the star is apparently elevated. A star in the zenith is not displaced by refraction.

These values for average conditions are somewhat altered by variations in the temperature and pressure of the air.

A star directly overhead is not displaced by refraction, because its rays are perpendicular to the atmospheric layers. The amount of the refraction increases as the distance from the zenith increases, but so slowly at first that for considerably more than halfway to the horizon the effect on the star's direction is appreciable only when observed with the telescope. As

the horizon is approached, the effect becomes more noticeable. A star at the horizon is raised by refraction more than half a degree, or slightly more than the apparent diameter of the sun, or moon.

4.4 Refraction Effects Near the Horizon

Owing to atmospheric refraction the sun comes fully into view in the morning before any part of it would otherwise appear above the horizon, and it remains visible in the evening after it would otherwise have passed below the horizon. Thus refraction lengthens the daily duration of sunshine. Similarly, the risings of the moon and stars are hastened and their settings are delayed. Refraction also increases by more than half a degree the radius of the region around the north celestial pole where stars never set, and diminishes by the same amount the opposite region where stars never rise (1.15).

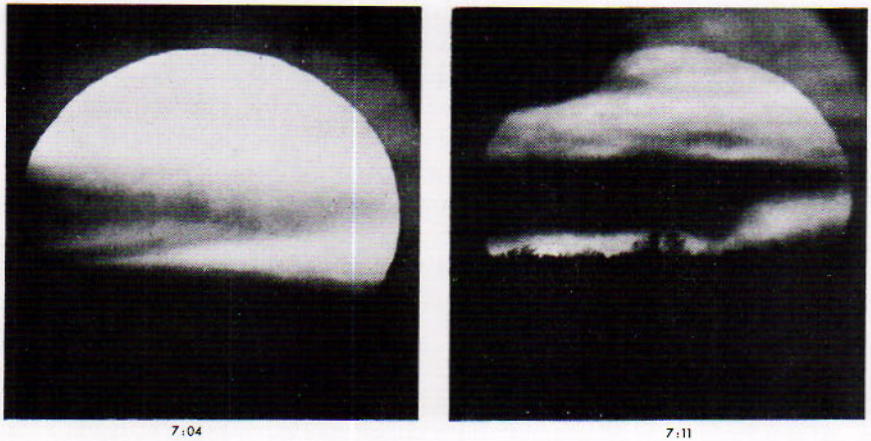


Figure 4.4
Apparent flattening of the setting sun by refraction. *(Yerkes Observatory photographs)*

Because atmospheric refraction increases with distance from the zenith, the lower edge of the sun's disk is raised more than the upper edge. This apparent vertical contraction becomes noticeable near the horizon, where the amount of the refraction increases most rapidly with increasing zenith distance. Thus the sun near its setting and soon after its rising sometimes appears conspicuously oval.

Another effect near the horizon has no connection with refraction. It is the well-known illusion that the sun, moon, and constellation figures seem magnified there.

4.5 Twinkling of the Stars

The twinkling, or *scintillation*, of the stars, that is, their rapid fluctuations in apparent brightness, results from the turbulence of the atmosphere within a few miles of the earth's surface. Currents differing in temperature and water content cause a varying irregularity in the density of the

air through which the rays of starlight pass. By variable refraction and especially by reinforcement and interference of different parts of the wave fronts, the starlight comes to us nonuniformly. Alternate fringes of light and shade cross the line of sight with frequencies up to 1000 Hz according to G. Keller, and may cause the star to vary by as much as 10 per cent of its average brightness.

The larger planets do not ordinarily twinkle; their steady light distinguishes them from neighboring stars. Similarly, the moon does not twinkle, nor does a street light that is close at hand. Like the moon, these planets are luminous disks, although a telescope is required to show them as such. While each point of the disk may be twinkling, the effects are not synchronized; for the rays take slightly different paths through the air and do not encounter the same irregularities.

Astronomical seeing, referring to the distinctness of the view with the optical telescope, is also affected by the drifting of the recurrent turbulent atmospheric pattern across the incoming beam. If the aperture of the telescope is large relative to the length of the pattern, the image of a celestial object is spread and blurred in "bad seeing," but remains fairly steady. With smaller apertures the image is not greatly enlarged in these conditions, but may shift about somewhat in the field of view.

A similar seeing effect is encountered by radio astronomers and is caused by fluxuations in the ionosphere and the interstellar medium. The blurring effect of the interstellar medium will set the limit upon the least resolvable element by interferometric techniques.

DISPERSION OF LIGHT; THE SPECTRUM

Dispersion is the term given to the variation of the velocity of radiation in material substances. This property gives us a powerful tool for studying conditions where the radiation originated. Spreading radiation out in order of wavelength gives the spectrum of the object.

4.6 Dispersion

Whenever light is refracted, it is separated into its constituent colors. An example is seen in the rainbow, the array of colors from violet to red that is formed when sunlight is refracted by drops of water. Refraction, as we have noted, is caused by the change in the speed of light

when it passes from one medium into another of different density. The change in speed is greater as the wavelength of the light is shorter. Thus the amount of the refraction increases with diminishing wavelength. In the visible spectrum violet light is the most changed in direction, red the least, and the other colors intermediate.

Refraction of light is accordingly accompanied by its *dispersion* in order of wavelength into the *spectrum*. The visible spectrum from violet to red comprises only a small part of the whole range of wavelengths radiated by a source such as the sun. The spectrum goes on into the ultraviolet in one direction and into the infrared in the other, where it is recorded by photography and other means.

The *spectroscope* is the instrument with which the spectrum is observed. It is employed in the laboratory and in connection with the telescope.

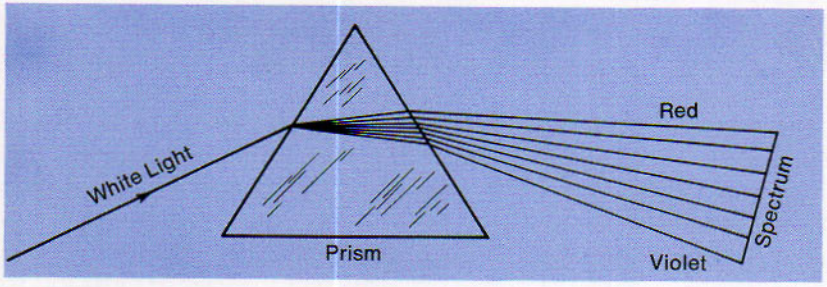


Figure 4.6
Formation of a spectrum by a prism.

4.7 The Spectroscope

A familiar type of spectroscope consists of a *prism* of glass, quartz, or other transparent material, toward which a *collimator* and *view telescope* are directed. The light enters the tube of the collimator through a narrow *slit* between the sharpened parallel edges of two metal plates. The slit is at the focus of the collimator lens; after passing through this lens the rays are accordingly parallel as they enter the prism. The light is refracted by the prism and dispersed into a spectrum, which is brought to focus by the objective of the view telescope and magnified by its eyepiece. When the eyepiece is replaced by a plate holder, the view telescope becomes a camera.

If the light is monochromatic, the spectrum is simply the image of the slit in that particular wavelength; if the light is white, that is, composed of all colors, the visible spectrum is a band, violet at one end and red at the other, which is formed by overlapping images of the slit in the different colors. The absence of a certain wavelength in the otherwise continuous spectrum is detected most easily when the separate images are so narrow that they overlap as little as possible. This is the reason for the narrow slit.

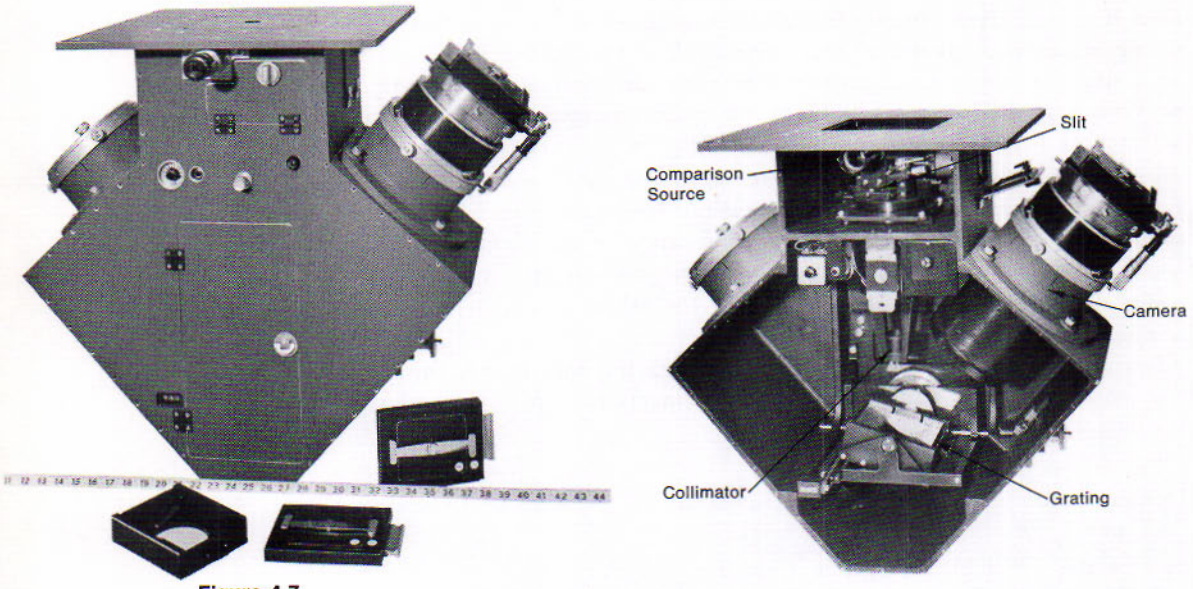
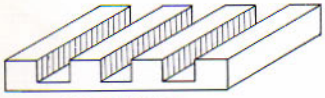
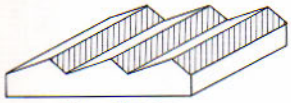


Figure 4.7
A modern reversible grating spectrograph, which is attached to the telescope at the Cassegrain position. *(Astrometrics, Inc.)*

Thus the spectroscope is an instrument for arraying light in order of wavelength. Bright lines in the spectrum represent the presence of certain wavelengths in the light. Dark lines show the absence of certain wavelengths. Integrated light is like a set of books in many volumes piled in disorder, whereas the spectrum of light is like the set arranged in order on a shelf; if a volume is withdrawn, the vacant space bears witness to the fact.

The prism spectroscope just described is simplest to understand because most of us are familiar with the action of lenses and prisms. The prism spectroscope has the great advantage that it puts all of the light into a single spectrum. It has the disadvantage that it compresses the color bands closer and closer together towards the red end of the spectrum.


Normal Grating


Blazed Grating

The scale of the spectrum can be made constant and given almost any value, large or small, by substituting a *grating* for the prism. A grating is a plate on which many parallel grooves are ruled—as many as 10,000 to the inch (about 4,000 to the centimeter). The grating forms spectra by the *diffraction* of light, which occurs when light passes a straight sharp edge or through a very narrow slit. The intensity of the light is redistributed by wavelength and in certain directions only. Because early gratings diffracted light in all allowed directions, they could only be used to study the sun or a few of the brightest stars. Modern gratings are *blazed*, that is, their grooves are cut at a slant, (see drawing on margin) looking like saw teeth rather than troughs, as in the older gratings. This new technique causes as much as 80 per cent of the light to go into a single order,

and nowadays grating spectrographs have essentially replaced prism spectrographs. We should also point out that mirrors often replace the collimator and camera lenses.

There are other two major "spectroscopes" in current use and both use the *interference property of radiation.* When light is divided into two beams, these can be recombined. If one of the two beams traverses a longer path, however, its intensity at a given wavelength may be increased or decreased accordingly for the recombination to take place in phase or out of phase. The Fabry-Perot interferometer is used to look at small regions of the spectrum with very high resolution. The Michelson interferometer, long familiar in optics, can also be used for spectroscopy. In this case there is no spectrograph in the conventional sense and thus spectroscopy done with a Michelson interferometer is called *Fourier transform* spectroscopy.

Spectroscopy is achieved in radio astronomy simply by having multiple receivers tuned to slightly different frequencies. In X-ray astronomy, the spectrum is obtained by sorting out the radiation with devices that respond to different energies.

4.8 Emission and Absorption Spectra

The spectra of luminous sources are mainly of three types:

The *bright-line spectrum,* or *emission-line spectrum,* is an array of bright lines. The source of the emission is a glowing gas, which radiates in a limited number of wavelengths characteristic of the chemical element of which the gas is composed. Each gaseous element in the same conditions emits its peculiar selection of wavelengths and can, therefore, be identified by the pattern of lines of its spectrum. The glowing gas of a neon tube, for example, produces a bright-line spectrum.

The *continuous spectrum* is a continuous emission in all wavelengths. The source is a luminous solid or liquid, or it may be a gas in conditions such that it does not emit selectively. The glowing filament of a lamp produces a continuous spectrum. There may also be emission or absorption continua (12.7) of limited extent in the spectra of certain gases.

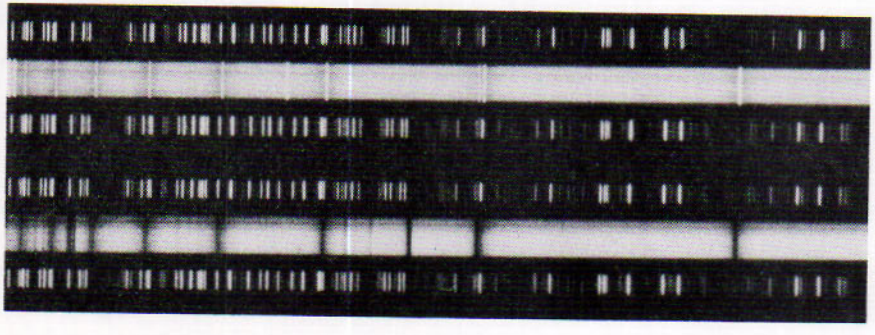


Figure 4.8
Bright-line and dark-line spectra. The spectra of BM Orionis (top) and 17 Leporis (bottom) are excellent examples of stellar emission line and absorption line spectra. The comparison lines are bright-line spectra free of a background continuum. *(Spectrograms by Catharine Garmany)*

The *dark-line spectrum,* or *absorption-line spectrum,* is a continuous spectrum except where it is interrupted by dark lines. Cooler gas intervenes between the source of the continuous spectrum and the observer. The intervening gas is opaque to the wavelengths it emits in the same conditions. The spectrum is therefore the reverse of that of the gas itself; it is a pattern of dark lines, which identifies the composition of the gas. Sunlight, having passed through the atmospheres of both the sun and the earth, produces a dark-line spectrum.

Where the gas consists of molecules such as carbon dioxide or methane, the spectrum shows bands of bright or dark lines characteristic of these molecules.

Our discussion emphasizes the radiation arising in a thermal source. Electrons accelerated in a magnetic field to extremely high velocities (very nearly the velocity of light) radiate a continuum. The study of cosmic non-thermal sources is forming a new and exciting chapter in astronomy.

4.9 The Doppler Effect

When the source from which waves are emitted is approaching the observer, the waves are crowded together, so that the wavelengths are diminished. When the source is receding, the waves are spread farther apart, so that their lengths are increased. A familiar example in sound is the lowered pitch of the whistle as a locomotive passes us. The Austrian physicist C. J. Doppler pointed out, in 1842, that a similar effect is required by the wave theory of light.

The *Doppler effect* as applied to the spectral lines is as follows:

When the source of light is relatively approaching or receding from the observer, the lines in its spectrum are displaced respectively to shorter or longer wavelengths by an amount proportional to the speed of approach or recession.

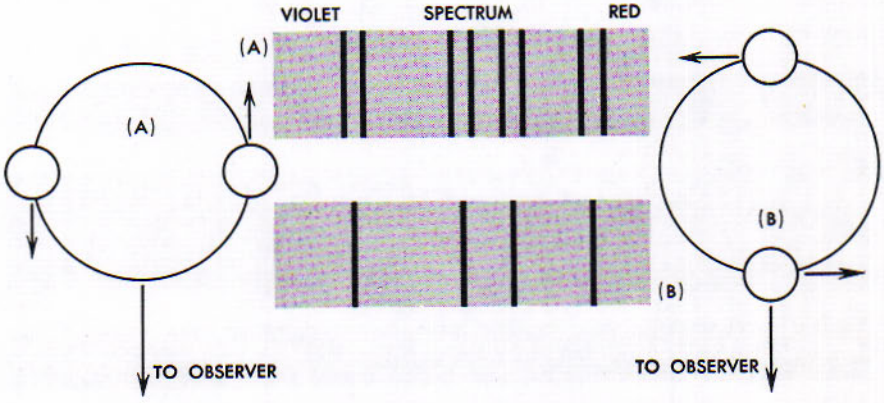


Figure 4.9
The doppler effect. The spectrum lines are displaced to the violet when the source is approaching the observer and to the red when the source is receding from him.

The amount of displacement of a line in the spectrum is related to the relative speed of approach or recession by the formula:

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

$$\frac{\text{change of wavelength } (\Delta\lambda)}{\text{normal wavelength } (\lambda)} = \frac{\text{relative speed } (v)}{\text{speed of light } (c)}$$

If, for example, the source and observer are relatively approaching at the rate of 29.5 km/sec, the lines in the spectrum of the source are displaced shortward nearly a ten-thousandth of their normal wavelengths.

The relation just given is adequate when the velocity involved is small compared to the velocity of light. When the velocity is large, say 30,000 kilometers per second or more, we must use the more general form of the relation derived by applying the special theory of relativity.

4.10 Some Uses of Spectrum Analysis

The spectrum of a luminous celestial body gives information as to the physical state of that body. A bright-line spectrum is produced generally by a tenuous gas, and a dark-line spectrum by a gas intervening in the path of the light. The pattern of lines identifies the chemical composition of the gas producing them.

The selective reflection of sunlight for different parts of the spectrum by the surface of a planet may inform us of the nature of the reflecting surface. If a planet has an atmosphere, any dark bands in the spectrum of the reflected sunlight which do not appear in the spectrum of direct sunlight give evidence of the chemical composition of that atmosphere.

The Doppler effect in the spectrum of a celestial body informs us about its motion in the line of sight, whether it is moving relatively toward or away from the earth and how fast it is moving. Other uses of spectrum analysis in astronomy will be noted later.

THE REFRACTING TELESCOPE

The refraction of light by transparent materials makes possible the refracting telescope. The great telescopes of the 19th century were refractors.

4.11 Refraction by Simple Lenses

Lenses are generally of two kinds: converging lenses, which are thicker at the center than at the edge, and diverging lenses, which are thinner at the center. An example of the former is the double convex lens, two important uses of which are to form a real image of an object and to serve as a magnifying glass.

The *focal length* of the lens is the distance from the center, C, to the *principal focus*, F (Fig. 4.11A), where rays of light parallel to the optical axis are focused by the lens. When the object is farther from the lens than is the principal focus, the lens produces a real inverted image of the object, which may be shown on a screen or photographic plate. As indicated in the figure, rays passing through the center of a thin lens are unchanged in direction.

When the object is nearer the lens than the position of the principal focus, the eye behind the lens sees an erect and enlarged virtual image of the object (Fig. 4.11B). This is the use of the lens as a magnifier. In combination, two double convex lenses can serve as a simple refracting telescope. The first lens is the *objective* of the telescope, which forms an inverted image of the object. The second lens, the *eyepiece*, placed at the proper distance behind that image, permits the eye to view the object, which now appears magnified and still inverted. Two other lenses may be added at the eye end if it is desired to reinvert the image.

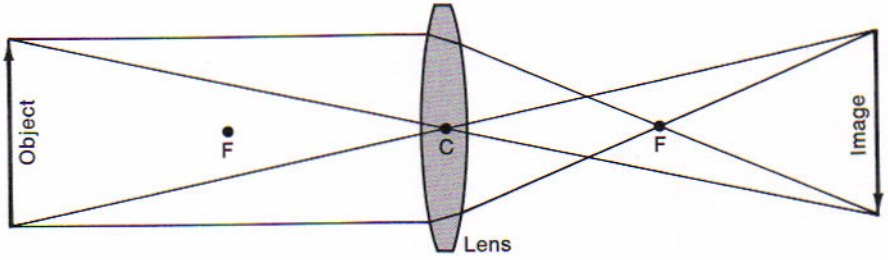


Figure 4.11A
The convex lens as an objective. The lens produces an inverted real image of the object.

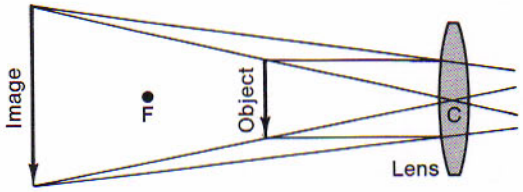


Figure 4.11B
The convex lens as an eyepiece. It produces an erect and enlarged virtual image.

4.12 The Refracting Telescope

The discovery of the principle of the refracting telescope is generally credited to a Dutch spectacle maker. Galileo, in 1609, is credited with being the first to apply this principle in the observation of the celestial bodies. Two of his telescopes are preserved intact in the Galileo Museum in Florence, Italy; the larger one, having a paper tube about 1.2 m long and less than 5.1 cm in diameter, magnifies 32 times. The Galilean telescope has a double concave eyepiece. It has the merit of giving an erect image of the object observed, but its field of view is small.

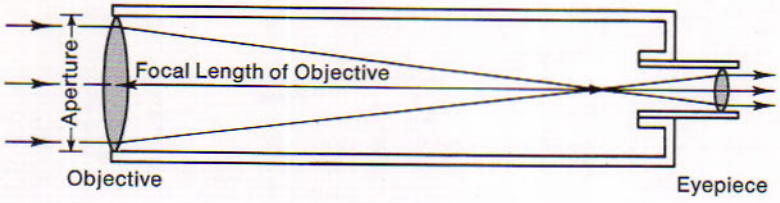


Figure 4.12
The simple refracting telescope. In its simplest form the refracting telescope consists of two convex lenses separated by a distance equal to the sum of their focal lengths. In modern refracting telescopes both objective and eyepiece are compound lenses.

The *simple astronomical telescope,* which is the basis of modern refracting telescopes, contains two double convex lenses at a distance apart equal to the sum of their focal lengths (Fig. 4.12). The objective has the greater *aperture,* or clear diameter, and the longer focal length; it receives the parallel rays from each point of the remote celestial object and forms an inverted image of the object at the principal focus. The eyepiece, set in a sliding tube, serves as a magnifier for viewing the image formed by the objective. As he looks through it, the observer sees an inverted and enlarged image of the celestial object.

The early refracting telescopes proved disappointing, especially when larger ones were constructed. The views of the celestial bodies were blurred. An important cause of the poor definition was the dispersion of light which accompanies its refraction. A single lens focuses the different colors at different distances from the lens, violet light at the least distance and red light farthest away (Fig. 4.13A). The image of a star focused in any particular color is confused with out-of-focus images in the other colors. This is *chromatic aberration.*

As long as the telescope contained only single lenses, the only known way to improve the view was to increase the focal length. Toward the end of the 17th century, refracting telescopes as long as 46 and 61 meters (150 and 200 feet) were attempted, but they were so unwieldy that not much could be done with them.

4.13 The Achromatic Telescope

The new era of the refracting telescope began, in 1758, with the introduction of the achromatic objective. By an appropriate combination of lenses of different curvatures and compositions it is possible to unite a limited range of wavelengths at the same focus. The objectives of refracting telescopes designed for visual purposes are combinations of two lenses. The upper lens is double convex and of crown glass; the lower lens, either cemented to the upper one or separated by an air space, is likely to be plano-concave and of heavier, flint glass. By the use of two

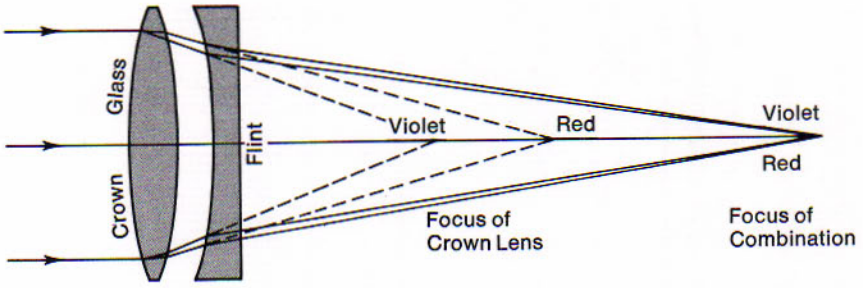


Figure 4.13A
Principle of the achromatic objective. The crown lens focuses the different wavelengths at different distances, so that a clear image of a star is not obtained anywhere. The addition of the flint lens increases the focal length and, since it affects the shorter more than the longer wavelengths, brings the different colors to focus more nearly at the same point.

lenses other difficulties inherent in the single spherical lens are compensated as well.

The objective of the visual refracting telescope brings together the yellow and adjacent colors, to which the eye is especially sensitive; but it cannot focus with them the deep red, or the blue and violet light which most strongly affects the ordinary photographic plate. Evidence of this is seen in the purple fringe around the image of the moon. Thus a refracting telescope giving fine definition visually does not by itself produce clear photographs. When the visual refracting telescope is employed as a camera, the plate holder replaces the eyepiece. A correcting lens is introduced before the plate, or else a yellow filter is used to transmit only the sharply focused light to an appropriate type of plate.

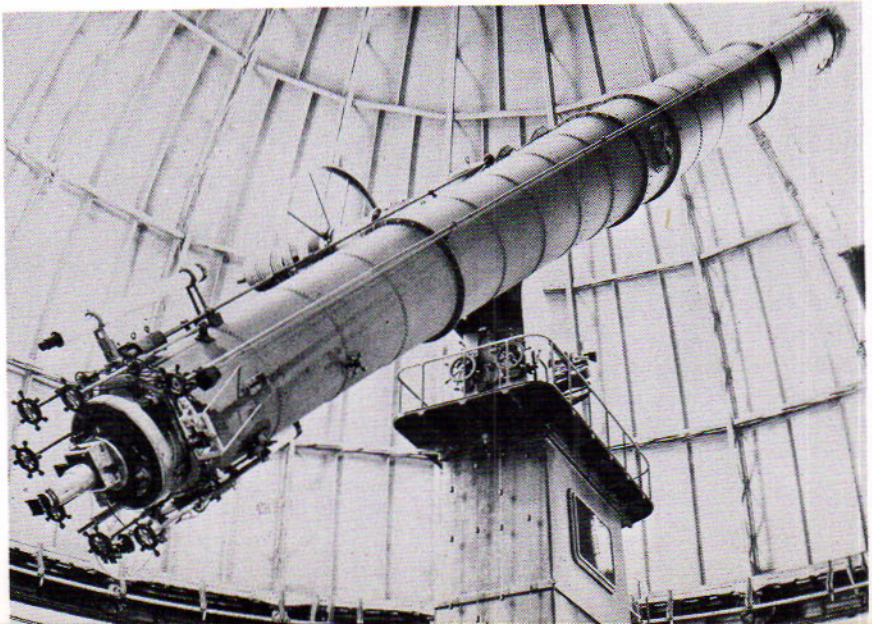


Figure 4.13B
The 40-inch (102 cm) refracting telescope (largest in the world) at Yerkes Observatory, of the University of Chicago, Williams Bay, Wisc. *(Yerkes Observatory Photograph).*

The aperture of the objective is usually stated in denoting the size of a telescope. Thus a 30.5 cm telescope has an objective with a clear diameter of 30.5 cm. In ordinary refracting telescopes the ratio of aperture to focal length is about 1 to 15 or more, so that a 30.5 cm telescope is likely to be about 457 cm (4.57 meters) long. Refractors designed for photography of large areas of the sky have objectives containing more than two lenses and are referred to as astrographs. These permit shorter focal ratios, which increase the diameter of the field and also the speed.

THE REFLECTING TELESCOPE

Highly polished surfaces reflect light. If the surface is carefully made it can focus light and thus serve as an objective for a telescope. All of the great telescopes of the 20th century have been reflectors.

When a ray of light encounters a polished surface which prevents its further progress in the original direction, the ray bounds back from, or is *reflected* by, the surface. If the mirror is appropriately curved, it forms the image of an object, taking the place of a lens. Consider, for example, a concave spherical mirror (Fig. 4.14), having its center of curvature at C with a radius R and its principal focus at F. Of an object beyond C the mirror forms an inverted real image, which may be viewed on a screen or with an eyepiece. Thus the mirror can serve as the objective of a telescope.

The mirror has the advantage over the lens of being perfectly achromatic; there is no dispersion when light is reflected. But the spherical mirror, in greater degree than the spherical lens, introduces *spherical aberration;* the focal point is not the same for different zones of the mirror. This effect is seen in the caustic curve formed on the surface of a cup of coffee by light reflected from the sides of the cup. The perfect remedy for spherical aberration is to make the mirror paraboloidal instead of spherical.

There are other reasons why the objectives of very large telescopes are mirrors rather than lenses. (1) It is easier to make a disk of glass for a mirror, because the light does not go through the glass. Striae and other defects in the disk, which would render it useless as a lens, do

4.14
Reflection from Curved Mirrors

Figure 4.14
Image formed by a concave mirror. The mirror forms an inverted real image of the object that is beyond the center of curvature.

not make it unfit for use as a mirror. (2) The optician has to figure only one surface instead of four. (3) The entire back of the mirror can be supported, whereas the lens is held only at its edge; a large lens may bend slightly under its own weight, affecting its figure and therefore its performance. (4) The focal length of the mirror can conveniently be made shorter than that of the lens (the ratio of aperture to focal length of the mirror is often about 1 to 5 or less), so that the mounting and dome may be smaller, with reduction in the cost of construction.

4.15 The Reflecting Telescope

The objective of the usual type of reflecting telescope is a paraboloidal mirror at the lower end of the tube, which in the largest telescopes is of skeleton construction. The mirror is a circular block of glass having its upper, concave surface coated with a thin film of metal such as aluminum. The light does not pass through the glass, which serves simply to give the required shape to the metal surface. The large mirror reflects the light of the celestial object to focus in the middle of the tube near its upper end and may form an image there. Observations with large telescopes may be made at four different places:

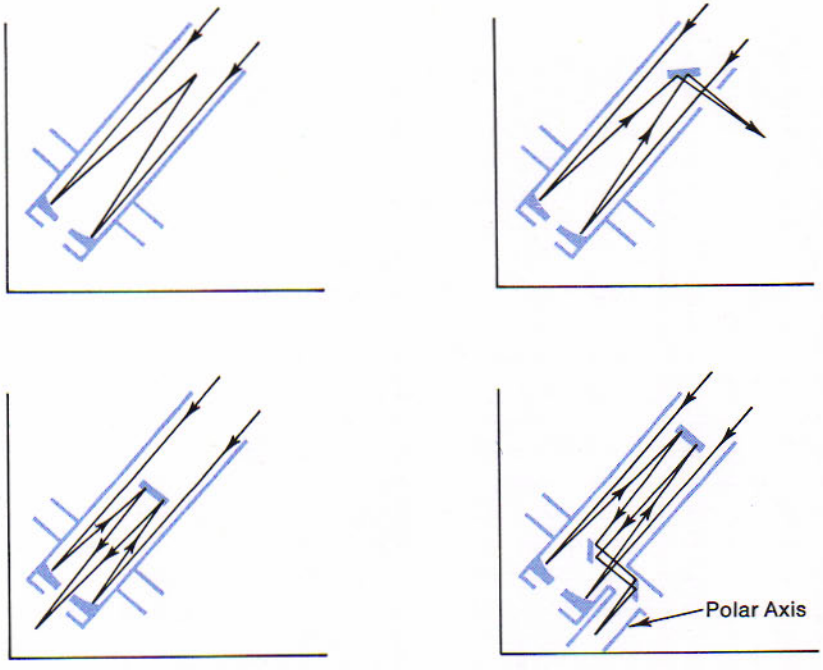


Figure 4.15
Four places of focus with the reflecting telescope. (1) Prime Focus. The focus of the large mirror. (2) Newtonian Focus. The converging beam from the large mirror is diverted to the side of the tube by a small diagonal mirror. (3) Cassegrain Focus. The converging beam is reflected by a small mirror through the central aperture of the large mirror. (4) Coudé Focus. The returning beam is diverted by a diagonal mirror down the polar axis to the spectroscopic room.

1. At the *prime focus.* This is generally not easily accessible to the observer except in the largest telescopes.
2. At the *Newtonian focus.* A smaller plane mirror at an angle of 45° near the top of the tube receives the converging beam from the large mirror and reflects it to focus at the side of the tube. In the visual use the observer looks in at right angles to the direction of the telescope.
3. At the *Cassegrain focus.* A small convex hyperboloidal mirror near the top of the tube reflects the converging beam back through an opening in the center of the large mirror to focus below it. The observer then looks in the direction of the star as with the refracting telescope.
4. At the *coudé focus.* The beam returned by the convex mirror is reflected by a small diagonal plane mirror down the polar axis to focus in the laboratory where spectroscopic analysis may be made conveniently.

Several other focus positions enjoy a certain currency, notably the *Nasmyth,* which is a variation of the Cassegrain; and the *Springfield,* the opposite of the coudé focus. The large mirrors of the Mount Wilson telescopes (Nasmyth focus) have no openings, and the beam returned by the convex mirror is reflected to focus at the side by a plane mirror set diagonally in front of the large one. In the Springfield focus the light is sent up the polar axis, on top of which is the eyepiece. The observer can sit in a fixed position and view the image. This focus, popular among amateurs, is common in smaller telescopes.

4.16 The Schmidt Telescope

The type of reflecting telescope described in the preceding section is admirable for viewing and photographing limited areas of the heavens. The Schmidt telescope is useful for photographing large areas and is employed only as a camera. It is named after its designer, Bernhard Schmidt, who was an optical worker at the Hamburg Observatory in Germany.

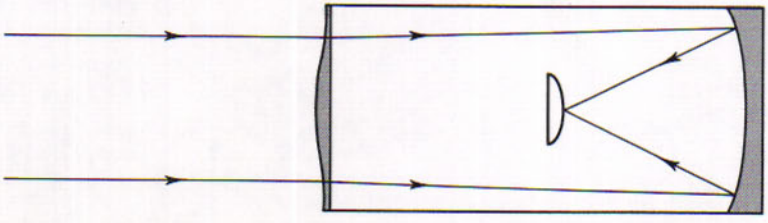


Figure 4.16A
Optical system of the Schmidt telescope. The starlight passes through the correcting plate before being reflected by the spherical mirror to focus on the curved photographic plate.

The objective of the Schmidt telescope is a spherical mirror, which is easy to make but is not by itself suitable for a telescope. Parallel light rays reflected by the middle of a spherical mirror are focused farther away from it than are those reflected from its outer zones. The correction is effected by a thin *correcting plate*, set at the center of curvature of the mirror in the upper part of the telescope tube. The correcting plate is so thin in present telescopes that it introduces no appreciable chromatic aberration. It slightly diverges the outer parts of the incoming beam with respect to the middle, so that the entire beam is brought together by the mirror on a somewhat curved focal surface. The photographic plate, suitably curved by springs in the plate holder, faces the mirror between it and the correcting plate. The size of the telescope is denoted by the diameter of the correcting plate.

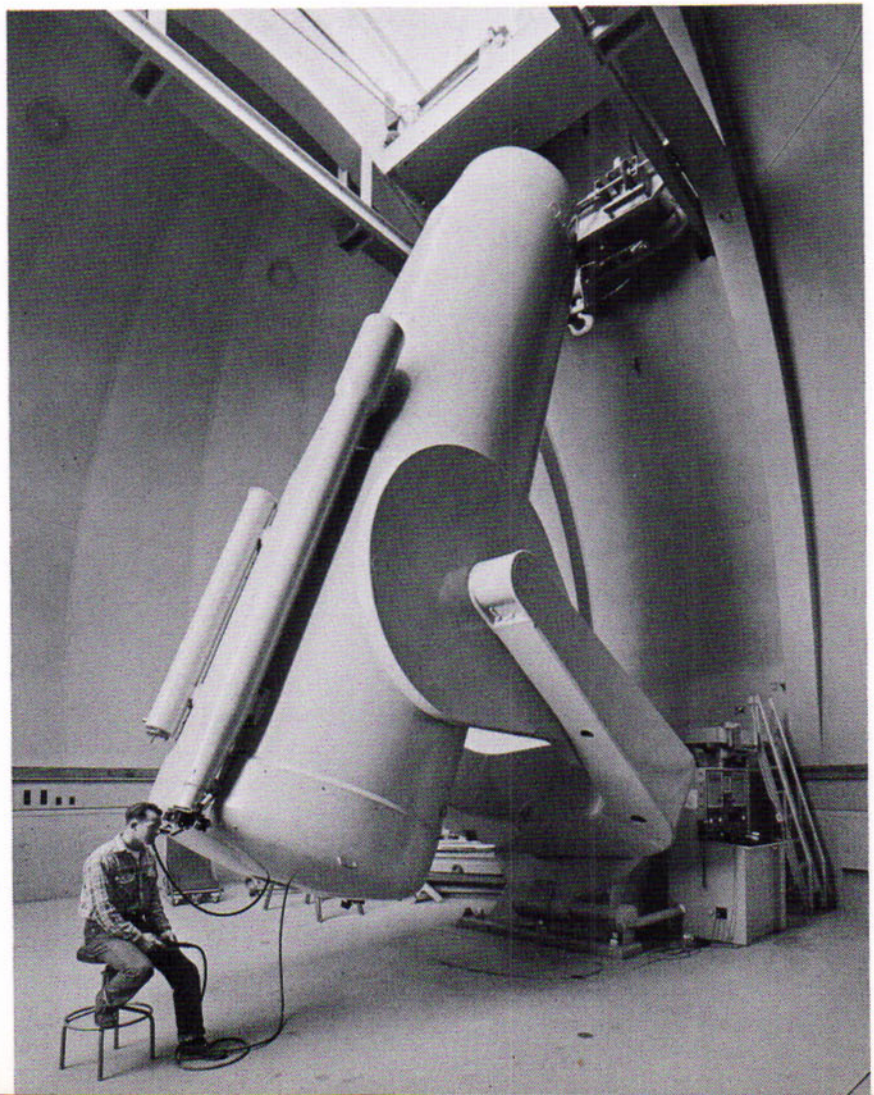


Figure 4.16B
The 48-inch (112 cm) Schmidt telescope at Mount Palomar. *(Photograph from the Hale Observatories)*

A large telescope of this kind is the 48-inch (122 cm) Schmidt telescope of the Palomar Observatory. Its 72-inch (1.83 meters) mirror has a radius of curvature of 20 feet (6.1 meters), which is about the length of the tube. The focal length is 10 feet (3.05 meters). An initial achievement of this telescope is the National Geographic Society—Palomar Observatory Sky Survey, a photographic atlas of the heavens north of declination 27° S. The atlas consists of 879 pairs of negative prints from blue- and red-sensitive plates, each print 35.6 cm square and covering an area 7° on a side. The survey reaches stars of the 20th magnitude and exterior galaxies of the 19.5 magnitude.

Other examples of this newer type are the 24-inch (61 cm) Schmidt telescope of the Warner and Swasey Observatory, the similar telescope of the University of Michigan Observatory, the 26-inch (66 cm) telescope of the Astrophysical Observatory of Mexico, the 80 cm telescope of the Hamburg Observatory, and the one meter telescope of the Burakan Observatory in Armenia. The 200 cm reflecting telescope of the German Academy of Sciences, near Jena, has a spherical mirror that can be operated with a 1.35 meter correcting plate as a Schmidt telescope. Very wide-angle 51 cm Schmidt telescopes having three correcting plates and triaxial mountings have been employed at some stations of the Smithsonian Astrophysical Observatory for tracking spacecraft.

The short focal lengths for a given aperture makes the Schmidt a very fast system. This, along with its large field, makes it an ideal camera for spectrographs and similar applications.

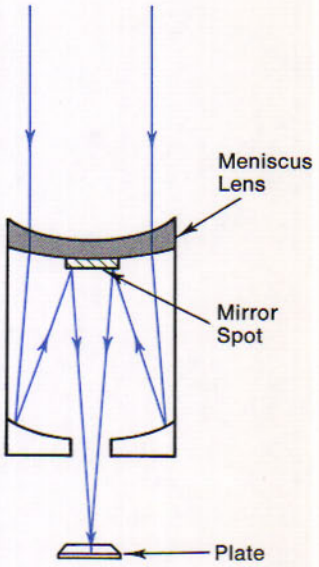


Figure 4.17
The Maksutov-Bouwers telescope.

4.17 The Maksutov-Bouwers Telescope

This system also uses a spherical primary but achieves its correction by having a *concave spherical meniscus lens* in place of a corrector plate. The basic design of the optics was described by A. Bouwers early in the Second World War, but because of circumstances his work was not known until the cessation of hostilities. In the meantime, D. D. Maksutov independently developed the same optical system. Basically this optical design allows a large well corrected field and puts a very long focal length system into a small space. This latter characteristic has made it a popular "small" telescope and telephoto system for photography. The largest Maksutov-Bouwers telescope is in Chile at the southern station of the Soviet Union. It has an aperture of about one meter.

4.18 The Equatorial Mounting

Most telescopes are mounted so that they can turn on two axes to follow the circles of the equatorial system. The *polar axis* is parallel to the earth's axis and is therfore inclined to the horizontal at an angle equal to the latitude of the place. Around this axis the telescope is turned parallel to the celestial equator, and so along the diurnal circle of the star. The *declination axis* is at right angles to the polar axis; around it the telescope is turned along an hour circle from one declination to another.

Each axis carries a graduated circle. The circle on the polar axis denotes the hour angle of the star toward which the telescope is pointing; and there may be a dial on the pier, from which the right ascension of the star can be read directly. The circle on the declination axis is graduated in degrees of declination. With the aid of these circles the telescope can be quickly pointed toward a celestial object of known right ascension and declination. The telescope is then made to follow that object by a driving mechanism, which is operated either mechanically or electrically.

In the standard type of equatorial mounting for refracting telescopes (Fig. 4.13B), the polar axis is at the top of a single pier, and the telescope must be reversed frequently from one side of the pier to the other. In the type used for most large reflecting telescopes, the long polar axis is supported by two piers, between which the telescope can swing from the east to the west horizon without reversal. In the fork type the handle of the fork is the polar axis and the two arms of the fork carry the trunnions of the telescope tube.

The equatorial mounting is mechanically the simplest for tracking celestial objects but for certain purposes an *altitude-azimuth mounting*, an *altitude mounting* or a *zenith mounting* is best. The alt-azimuth mounting reduces engineering problems and is used extensively in large radio telescopes. With the introduction of inexpensive computers this should become a more popular mounting for all wavelengths. The altitude mounting has long been used for transit circles and recently for the very large 91.4 meter (300-foot) telescopes of the National Radio Astronomy Observatory. The zenith mounting is popular for time service observations and just recently has been used for the largest radio telescope: the 305 meter (1000-foot) telescope at Arecibo, Puerto Rico.

4.19 Light-Gathering Power

The brightness of the image of a star increases in direct proportion to the area of the telescope objective, or the square of its diameter. This defines the *light-gathering power* of a telescope, aside from loss in the optical parts. Thus a star appears 25 times as bright with the 200-inch (508 cm) telescope as with the 40-inch (102 cm) Yerkes refractor, and a million times as bright as with the naked eye if the aperture of the lens of the eye is taken to be one fifth of an inch.

The telescope shows stars that are too faint to be visible to the eye alone. A larger telescope permits us to see fainter stars than does a smaller one, and extends our view farther into space for objects of the same intrinsic brightness. Whereas the limit of brightness for the unaided eye is about the 6th magnitude, stars as faint as the 19th magnitude are visible with an eyepiece at the prime focus of the 200-inch (508 cm) telescope; and stars of nearly the 24th magnitude can be detected in photographs made there.

The surface of the daytime sky appears less bright with the telescope than to the eye alone and especially so when higher powers are used. Because the starlight is concentrated by the telescope, the brighter stars and planets are visible with the telescope in the daytime.

4.20 Resolving Power

The image of a star or other point source of radiation is spread by diffraction in the objective of the telescope into a disk surrounded by fainter concentric rings. Two stars which are closer together than the diameter of the brighter part of the diffraction disk cannot be separated by any amount of magnification. This diameter is defined as between points where the intensity of light is one half that at the center of each disk.

The *resolving power* of a telescope, which equals this diameter, is the angular distance between two stars of moderate brightness that can be just separated with the telescope in the best conditions. The least distance, d, in seconds of arc is related to the wavelength, λ, of the light and the aperture, a, of the telescope in the same units by the formula: $d'' = 1.22 \times 206{,}265'' \times \lambda/a$. For visual telescopes the formula becomes: $d'' = 14.''1/a$, where the wavelength is taken to be about 56×10^{-6} cm (5600 Å) and the aperture is in centimeters.

$1\ \text{Å} = 10^{-8}$ cm

Thus with a 10 cm telescope the minimum resolvable separation is about 1.″4. It is slightly better than 0.″2 for the large Yerkes and Lick refracting telescopes and is only 0.″03 for the 200-inch (508 cm) telescope on Palomar Mountain. These values are for visual observations. In photographs with all telescopes the least distance for separation of two stars is made greater by the spreading of the images in the photographic emulsions. We note presently how the longer wavelenghts employed in radio telescopes affect their resolving power.

Aside from its ability to show fainter objects, a larger telescope has the advantage of better resolving power; it can show finer detail that runs together with the smaller instrument. On the other hand, the blurring of the image by atmospheric turbulence is more pronounced with the larger telescope. In order to profit by its greater resolving power, the site of a large telescope must be chosen carefully with respect to the steadiness of the air at that place.

The correctness of the formula for two stars of equal faintness has been demonstrated by visual observations with telescopes of various apertures. For the eye alone the formula does not hold; it gives the resolving power as about 20″ but the least separation the eye can resolve is several times greater. In fact, the eye is said to be a good one if it can separate the two stars of Epsilon Lyrae, which are 207″ apart. The difference is ascribed to the coarse structure of the retina of the eye.

4.21 Magnifying Power

The magnifying power of a visual telescope is the number of times the telescope increases the apparent diameter of an object as compared with the naked-eye view. It equals the focal length of the objective divided by the focal length of the eyepiece that is used.

The *linear size* of the image of a particular object formed by the objective increases directly with the focal length of the objective. The diameter of the image equals the angular diameter of the object times the focal length of the objective divided by the value of the radian, 57°.3. Thus the diameter of the image of the moon, which has an angular diameter of ½°, formed by an objective of 3 meter focal length is ½° × 300 cm/57°.3, or about 2.6 cm, this is its size in the photograph when the telescope is used as a camera.

The *angular size* of the image is greater as the eye is nearer it. The least distance of distinct vision for the unaided normal eye is 25.4 cm. With the eyepiece the eye can be brought closer to the image, which accordingly appears larger. If, for example, the focal length of the objective is 304.8 cm and that of the eyepiece is 1.27 cm, the object is magnified 240 times.

A telescope is usually provided with eyepieces of different focal lengths so that the magnification can be varied as desired. There is little gained by using a power greater than 50 times the aperture in inches because of the spreading of the light by magnification; and the higher powers are useful only when the seeing is rather steady. On the other hand, a power less than 3 times the aperture is unsuitable because the beam of light entering the eye is then larger than the widest opening of the lens of the eye at night. Thus with a 25.4 cm telescope the useful magnifying powers range from 500 down to 30.

Although the sun, moon, and some of the planets appear larger with the telescope, no amount of magnification with most telescopes can show the real disk of a star. This is because the diffraction disk is almost always larger than the angular diameter of the star itself.

4.22 The Telescope as a Camera

Large telescopes and many small ones as well are often employed for photography. The objective becomes the camera lens and the plate holder replaces the eyepiece at the focus. During the time exposure the telescope is turned westward by its driving mechanism to follow the star in the diurnal motion. Any inaccuracies in the following are corrected by slow motion devices, either operated by the observer or automatic, so as to keep a star in the field precisely at the intersection of the cross-wires of the guiding eyepiece.

By cumulative effect of the light during the exposures, the photographs can show features too faint to be visible to the eye at the same telescope. With the addition of a prism or grating they show the spectra of celestial

bodies, where there is much information not available in the direct views. The photographs provide permanent records, which the astronomer can observe leisurely and repeatedly so as to increase the accuracy of his results. They are of course valuable to the student of astronomy as well. The reader of this book has before him the prints from many celestial photographs. He can decide for himself as to the correctness of the descriptions and might even find in a photograph a significant feature the astronomer has overlooked.

A limit to the faintness of stars that can be reached by prolonged exposure is set by the light of the night sky. This illumination, caused partly by starlight diffused in the atmosphere and especially by the airglow (10.35), ultimately fogs the photograph seriously. The limiting exposure time for direct photography with fast emulsions with the 200-inch (508 cm) telescope is 30 minutes. The faintest stars that can be detected in these photographs are of blue magnitude 23.9, a gain of 5 magnitudes over the visual observations with this telescope.

It is difficult to overemphasize the revolutionary effect of the introduction of the photographic plate to astronomy. From the original work of Bond in 1850, photography has been the mainstay of data collection in astronomy as the student can gather for himself from the illustrations in this book. The "information" capacity of even a small photograph is difficult to match by other techniques and its "storage" lifetime is essentially indefinite. Other data collection techniques have caused equally great steps in astronomy and are even beginning to challenge the photographic plate. We will point these techniques out in the course of our study.

4.23 Large Optical Telescopes

The 200-inch (508 cm) Hale telescope on Palomar Mountain in California is the largest optical telescope. Next in order of size are the 120-inch (305 cm) telescope of the Lick Observatory on Mount Hamilton, California, the 107-inch (272 cm) telescope at the McDonald Observatory on Mount Locke, Texas, the 102-inch (259 cm) telescope of the Crimean Astrophysical Observatory, the 100-inch (254 cm) telescope of the Mount Wilson Observatory in California, the 100-inch (254 cm) Isaac Newton telescope at Hermontseux, England, the 90-inch (229 cm) of the Steward Observatory on Kitt Peak, Arizona and the 84-inch (213 cm) of the Kitt Peak National Observatory in Arizona. Half a dozen other telescopes have objectives 1.8 meters or more in diameter. All are reflecting telescopes.

Refracting telescopes have more moderate apertures. The largest are the 40-inch (102 cm) telescope (Fig. 4.13B) of the Yerkes Observatory at Williams Bay, Wisconsin, and the 36-inch (91 cm) telescope of the Lick Observatory.

The construction of larger refractors is not even contemplated, due to

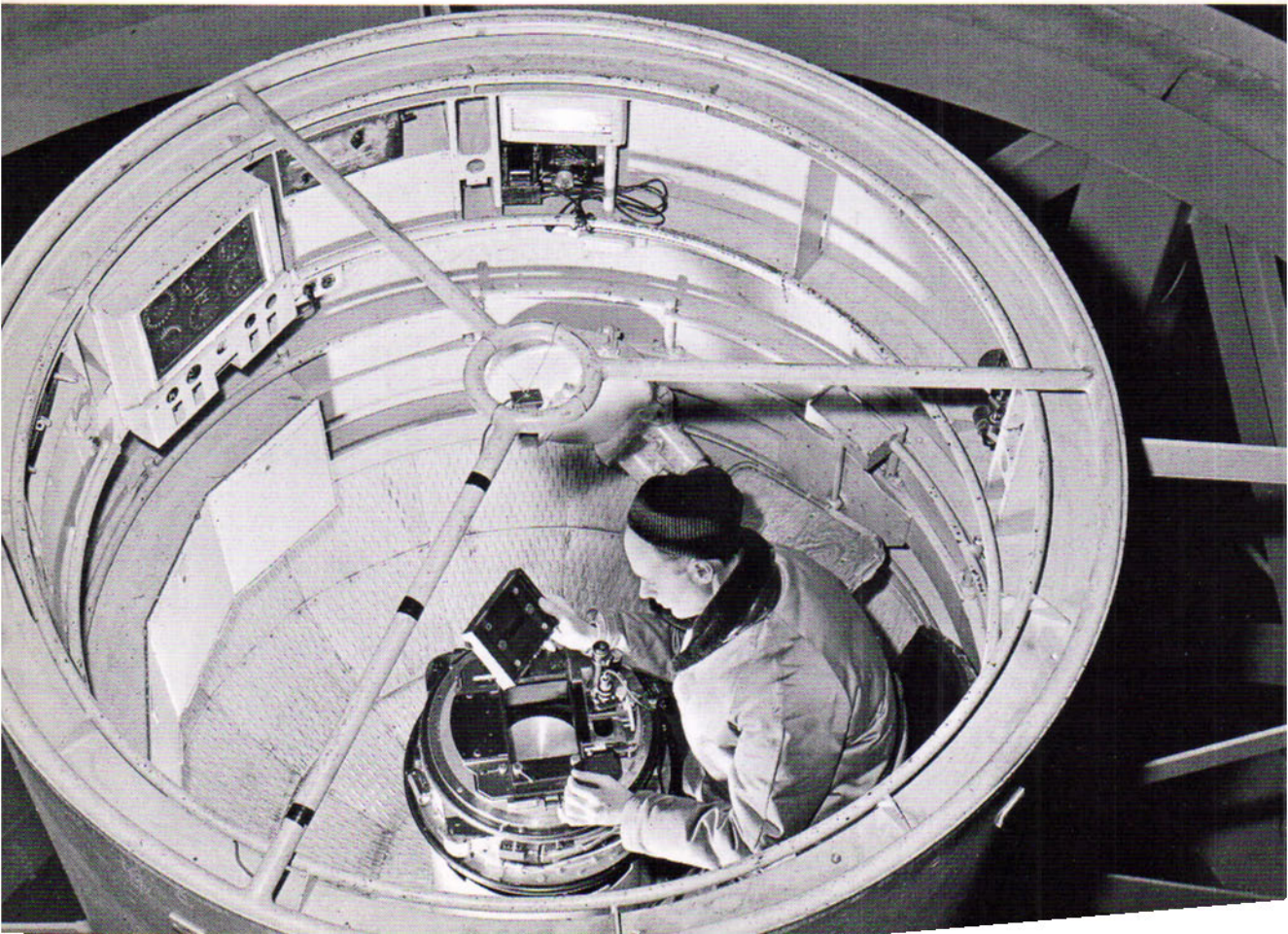
the many advantages of the reflectors. Of these, five major ones are presently under construction; all with apertures larger than 3 meters: a six-meter telescope near Zelenchuskaya in the Caucasus, USSR, and four 3.8-meter telescopes, to be located at Kitt Peak, Arizona, Cerro Tololo and La Silla, in Chile and in Australia.

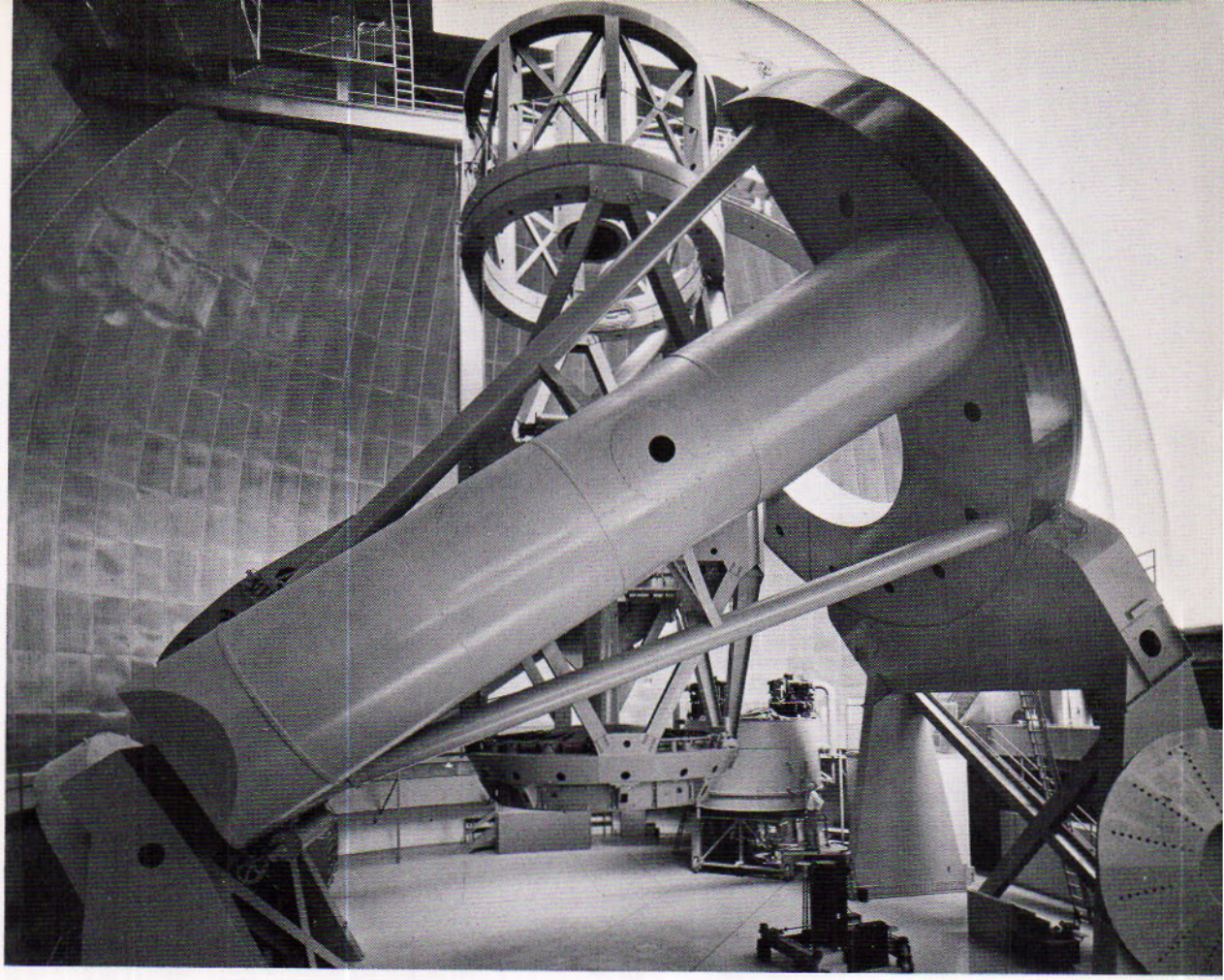
4.24 The 200-inch (508 cm) Telescope

This giant telescope on Palomar Mountain is operated by the Hale (formerly Mount Wilson and Palomar) Observatories. It is named the Hale telescope as a memorial to George E. Hale, who was successively director of the Yerkes and the Mount Wilson Observatory and who had a prominent part in the planning of the Palomar telescope.

The large mirror is a circular disk of Pyrex glass nearly 5.2 meters in diameter and 68.6 cm thick, having its concave upper surface coated with

Figure 4.24A
The observer's cage at the prime focus observing position of the 200-inch Hale telescope. Note plateholder and guiding eyepiece. *(Photograph from the Hale Observatories)*





aluminum. Its focal length is 16.9 meters. The back of the disk is cast in a geometrical rib pattern so that no part of the glass is more than 5.1 cm from the outside air. The telescope tube, 6.1 meters in diameter, moves in declination within the yoke that forms the polar axis. The skeleton tube, together with the mirror at the lower end and the observer's cage near the upper end, weighs 127 metric tons. In this cage, 1.52 meters in diameter, the observer is carried by the telescope as he works at the prime focus—a pioneer feature in telescope construction; the cage obstructs less than 10 per cent of the incoming light. The telescope is housed in a dome 41.8 meters in diameter.

With the aid of a corrector lens the photographs at the prime focus show in good definition an area of the sky 6.1 meters in diameter, or about 10.2 cm in diameter on the plate. When the 104 cm convex mirror is set in the converging beam, it reflects the light back through the 102 cm opening in the large mirror to focus below it. The effective focal length is 81.4 meters at the Cassegrain focus and 155.4 meters at the coudé focus in the laboratory.

Figure 4.24B
The 200-inch Hale telescope pointing to zenith. Notice size of human figure at bottom of picture. *(Photograph from the Hale Observatories).*

4.25
The 120-inch
(305 cm) Telescope

This large reflecting telescope at the Lick Observatory resembles the 200-inch telescope in many features of its design. The Pyrex disk of its mirror also has a ribbed back; it was cast as a trial run for the larger disk and was originally intended for use in testing that mirror. The 120-inch (305 cm) mirror has a focal length of (15.24 meters); thus the ratio of aperture to focal length is 1 to 5 instead of the ratio of 1 to 3.3 of the Palomar telescope. The scale of the photographs at the prime focus is nearly the same as for the larger telescope, and the light-gathering power is a third as great. There is an observer's cage at the prime focus. The dome is 29.6 meters in diameter. In both of these domes there is a gallery where visitors may view the telescope through a window without interfering with its operation.

Figure 4.25
The 120-inch (305 cm) mirror just being removed from its aluminizing tank. *(Lick Observatory Photograph.)*

4.26 Possible Improvements in Observing

As long as astronomical observing was almost entirely visual, the long-focus refracting telescope was ideal. Added power was effected by making larger telescopes, culminating in the 40-inch (102 cm) Yerkes refractor. At the beginning of the present century, photography was rapidly replacing visual procedures for many purposes. Photographic refractors were being used with focal ratios around 1 to 5 for added speed and width of field. The new large telescopes were reflectors having about this ratio and operating as well in the infrared as in the ordinary photographic range. The Schmidt telescope with a possible ratio as much as 1 to 1 greatly increased the speed and field.

To further increase the effectiveness of optical observing under our troublesome atmosphere, astronomers are thinking not so much of larger and still more expensive optical telescopes as of ways to improve the reception with present ones. Further improvements may be expected in photographic processes. Photoelectric techniques are developing rapidly. With a successful electronic image converter at its focus, as W. A. Baum has remarked, the 200-inch (508 cm) telescope might reach as far into space as would conventional direct photography with a 5080 cm telescope.

Observations from space will obviously improve our observational capability. We should also expect to see many more special purpose telescopes-telescopes for observing specific objects such as the Magellanic Clouds and for specific tasks such as astrometry and photometry, a trend radio astronomy has already started.

THE RADIO TELESCOPE

Radiations from the Galaxy at radio wavelengths were first detected, in 1932, by K. G. Jansky of the Bell Telephone Laboratories, who was investigating radio noise. In 1936, G. Reber, an electronic engineer at Wheaton, Illinois, built a paraboloidal antenna 9.1 meters in diameter for his pioneer radio map of the Milky Way. In 1946, radio reception from the sun was announced, the first discrete radio source was discovered, and radar began to be employed extensively in the recording of meteor trails. The hydrogen emission line in the radio spectrum was first observed in 1951. By means of this useful line, spiral arms of our galaxy were traced by astronomers in 1954. These and other results already achieved in this important new branch of astronomy will be noted in later chapters.

4.27 Radio Astronomy

This branch of Astronomy is the study of the heavens by use of radio waves. These waves that can come through to the ground range in length from a few millimeters, where they begin to be absorbed by atmospheric molecules, to 20 or 30 meters, where they cannot ordinarily penetrate the ionosphere. Because of their much greater lengths compared with light waves, the radio waves can pass through the clouds of our atmosphere and also through the interstellar dust that conceals all but 5 per cent of the universe from the optical view. Radio reception from the celestial bodies is as effective by day as by night.

Whereas optical spectra of the celestial bodies contain lines, the radio spectrum is almost entirely continuous. The remarkable exception is the line at the wavelength of 21 cm, formed by the neutral hydrogen atom at the ground state. By its Doppler shifts this line provides radio astronomy with means of measuring the velocities of sources in the line of sight.

The observational data in radio astronomy are procured in two ways: (1) from celestial radiations, and (2) by signals transmitted from the earth and reflected back from celestial bodies. An advantage of the second method is that the signals are under the observer's control; a disadvantage is that its application is limited to the nearer bodies of the solar system. The intensity of the reception falls off as the 4th power of the distance of the reflecting body, whereas in the first method the intensity inversely as the 2nd power of the distance from the source. In addition to echoes of radio signals returned from the moon, reflected beams from Mercury, Venus, Mars, and the sun's corona have been detected, and echoes from meteor trails are frequently recorded.

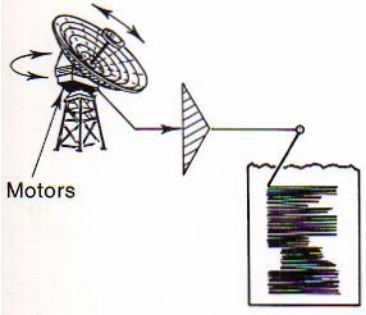


Figure 4.27A
Schematic diagram of the main elements of a radio-telescope.

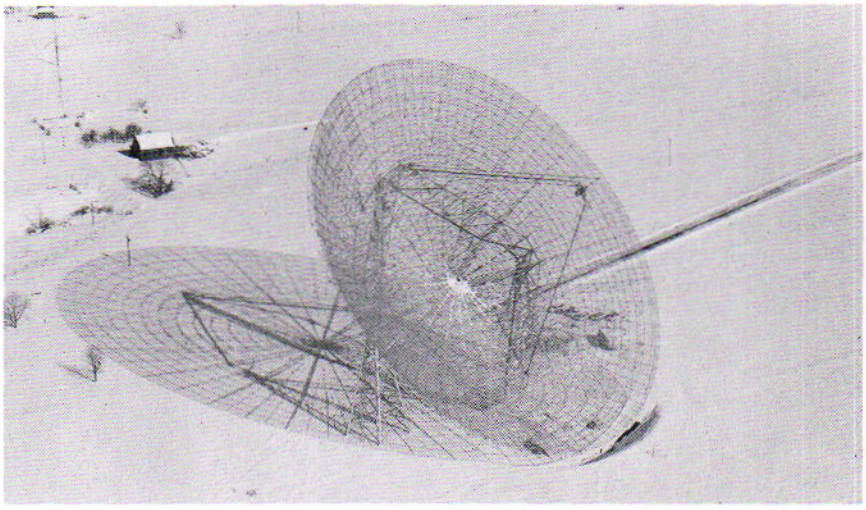


Figure 4.27B
The 300-foot paraboloidal transit telescope of the National Radio Astronomy Observatory. Note the projection effect in the shadow. *(National Radio Astronomy Observatory)*

Figure 4.27C
Owens Valley interferometer. *(California Institute of Technology)*

Cosmic radiations at radio wavelengths arise from two different causes. The first type is *thermal radiation;* its spectrum is that of black-body radiation defined like its optical counterpart by Planck's formula (10.5) for a particular temperature. This kind of radiation is strongest in the centimeter lengths. The second type is *nonthermal radiation,* which may be mainly the sort of radiation emitted by fast-moving particles revolving in the magnetic field of the synchrotron in the radiation laboratory. It is much stronger in the meter wavelengths than in shorter ones, and is separated thereby from the thermal radiation.

4.28
The Radio Telescope

This electronic device is analogous to the optical telescope. Its antenna, like the optical objective, serves to collect the radiations and to concentrate them on the receiver. Instead of an image of the source of radiation, we have here only the strength of the signal collected by the

antenna from a particular direction of the sky and recorded by a registering meter. The receiver output may be registered by a moving-chart pen recorder, or else by printed numbers or punches on tape.

An important problem of radio astronomy comes from the low power of the celestial radio source. The total power falling on the entire surface of the earth from the brightest radio source other than the sun is 100 watts, according to F. D. Drake. Of this amount a large radio telescope can collect only about 10^{-14} watts. The problem is to separate the weak radio signal from the noise within the receiver itself, which may be thousands of times greater. An effective solution is by use of a *maser*, for example, a highly refrigerated and magnetized synthetic ruby, to amplify the radio signal relative to the receiver noise.

The antennas of radio telescopes may be either revolving single paraboloids (Fig. 4.27B) or multiple element collectors built in a variety of forms. (Fig. 4.28)

Two paraboloid antennas can be coupled as an interferometer (Figure 4.27C) as we have already explained. The resolution of a single paraboloid is the same as given in section 4.20. In the case of the interferometer the denominator is modified by adding the length of the baseline (S_λ) measured, as before, in wavelengths as our unit. If we use a 25

Figure 4.28
The CSIRO interferometer. The large Parkes dish is the one in the background. *(CSIRO photograph, courtesy of J. Bolton).*

meter telescope at a wavelength of 10 cm, its resolution is very nearly 0.004 radians, or about 14′ of arc. Now if we use two such paraboloids as an interferometer and place them 4 km. apart we can ignore the factor a_λ and the resolution is increased to about 0.0000025 radians or about 0.″5 of arc.

With the development of portable atomic clocks (3.11) it is not even required that the two telescopes be physically connected. Here the trick is to synchronize two clocks with great precision. One then records on tape the signal being received by each telescope and the clock pulses as well. The two tapes are then played together making use of the timing pulses and the fringes are "observed". Great care must be exercised with the clocks and tape recorders and since the baselines are not usually on an east-west line the results require careful interpretation. This technique has been used successfully, first by N. W. Broten and his colleagues with a baseline of 3000 km. giving a resolution of 0.2 sec and present systems span oceans. We can even look forward to the day when one of the telescopes will be located on the moon giving a baseline of 38.4×10^8 wavelengths at 10 cm, hence a resolution of about 0.00007 seconds of arc.

Such resolution is required for very distant extra galactic objects and may be unachievable because of limitations set by interstellar scintillation. The spectacular results already achieved have led radio astronomers to suggest that coherent detection techniques using lasers may yield similar results in the visible spectrum.

4.29 The Paraboloidal Type

This type of radio telescope has as its antenna a "dish" of sheet metal or wire mesh, resembling in form the mirror of an optical reflecting telescope. The antenna collects the radio waves from a source and focuses them on a dipole adjusted to the desired wavelength, from which the energy is conveyed to the receiver and meter. The paraboloidal telescope is generally steerable and is usually equatorially mounted. Thus it can be directed to the right ascension and declination of the point of the heavens to be investigated, and it can then be made to follow that point in its diurnal motion.

The University of Manchester in England has a 76.2 meter telescope at the Jodrell Bank Experimental Station, and the Radio-physics Laboratory in Australia has a 64 meter telescope at a site about 320 kilometers west of Sydney. Stanford University has a 45.7 meter telescope with altazimuth mounting. The National Radio Observatory at Green Bank, West Virginia, has a 45.2 meter steerable telescope and a 91.4 meter transit-type telescope. The California Institute of Technology operates twin 27.4 meter paraboloids in Owens Valley about 402 kilometers north of Pasadena. They are equatorially mounted on flatcars that move

on north-south and east-west tracks 488 meters long from their intersection; these telescopes may be used either separately or together as an interferometer giving high resolution. An increasing number of single steerable paraboloids in various parts of the world have diameters between 18 and 27 meters. A 305 meter fixed bowl designed by Cornell University scientists is located in a natural depression at Arecibo, Puerto Rico.

4.30 Resolving Power

The resolving power of a radio telescope relates to the fineness of detail that can be distinguished. Calculated by the same formula (4.20) as for the optical telescope, it is the angular distance between two radio point sources that can be just separated. This critical distance is directly proportional to the wavelength of the radiation and inversely to the diameter of the antenna in the same units. For this purpose the main beam of the antenna, where its absorption is greatest, is analogous to the diffraction disk of the optical telescope.

Because it works with much longer wavelengths, the paraboloidal radio telescope is far less effective in separating detail than is an optical objective of the same diameter. Thus the critical separation for a 15.2 meter paraboloid at the wavelength of 21 cm is about 48′, compared with the theoretical separation of 0″.023 for visual light with the 200-inch (508 cm) Hale telescope. The radio resolution can be improved by employing larger telescopes and shorter wavelengths, or by interferometer methods promoted in a number of different ways.

4.31 Multiple Element Radio Telescopes

Interferometer devices enhance the collecting and resolving powers of radio telescopes at relatively small expense compared with that of increasing the diameters of single paraboloidal antennas. An example of this type is the original radio telescope at the Radio Observatory of Ohio State University, designed and operated by J. D. Kraus and associates (Fig. 4.31A). The antenna consists of 96 helices mounted on a steel frame 49 meters long. The frame is pivoted on a horizontal east-west axis so that it may be rotated to face any part of the meridian from the south horizon to the north celestial pole.

A pioneer example of the multiple element radio telescope was designed by W. N. Christiansen in Australia for scanning the sun's disk in the centimeter wavelengths. It employs an array of 32 paraboloidal reflectors each 1.8 meters in diameter, which is spread over a line 217 meters long; the resolution is 1 minute of arc. The classic Mills Cross, designed by B. Y. Mills at Sydney, is a crossed array of dipoles 460 meters long. The new interferometric telescope of Cambridge University, completed in 1958, has a fixed east-west array of unit parabolas 442 meters long and a second, moveable aerial at a considerable distance. Multiple element radio telescopes are increasing remarkably in complexity and effectiveness.

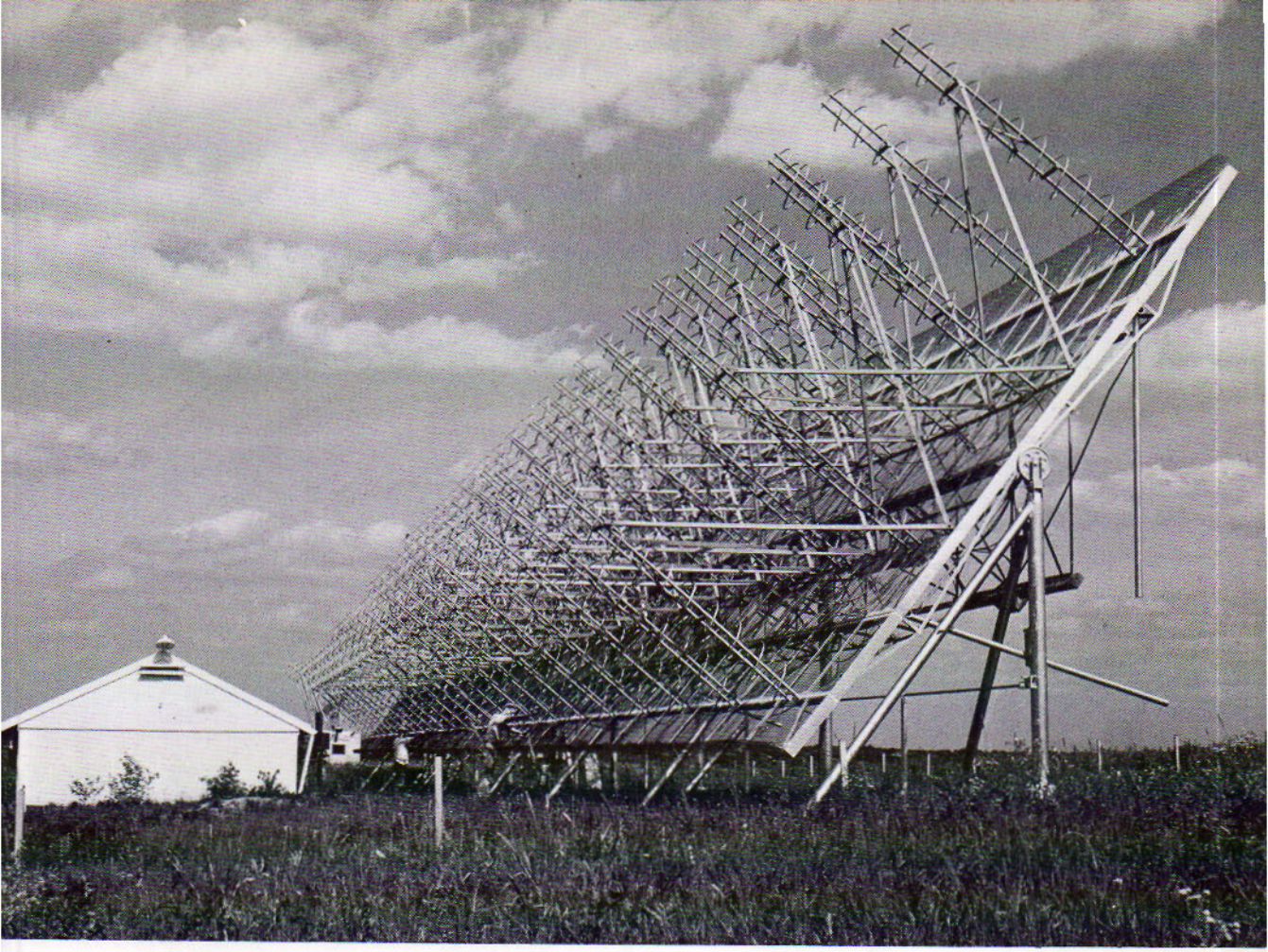


Figure 4.31A
The original Ohio State University radiotelescope, with an array of 96 helixes. *(Ohio State University).*

Perhaps the most exciting development in telescopes is the development of *synthetic-aperture telescopes* for radio astronomy. Large arrays, such as the Mills Cross discussed above, become extremely expensive if they are to achieve the resolution required by astronomy. It is beyond the scope of this book to develop the analysis used in aperture synthesis, but basically we can say that if the radiation from the source is correlated, then the radiation pattern on a series of antennas has a definite pattern related to a mathematical construct called the Fourier components of the *Fourier transform.* If we can observe the components we can obtain the Fourier transform and hence reconstruct the radiation pattern of the source.

The first application of this complex development was made by J. H. Blythe and it has been used very successfully by several observatories. With the development of *electronic phase switching* it has been possible to remove the need for moving telescopes in the system and hence reduce

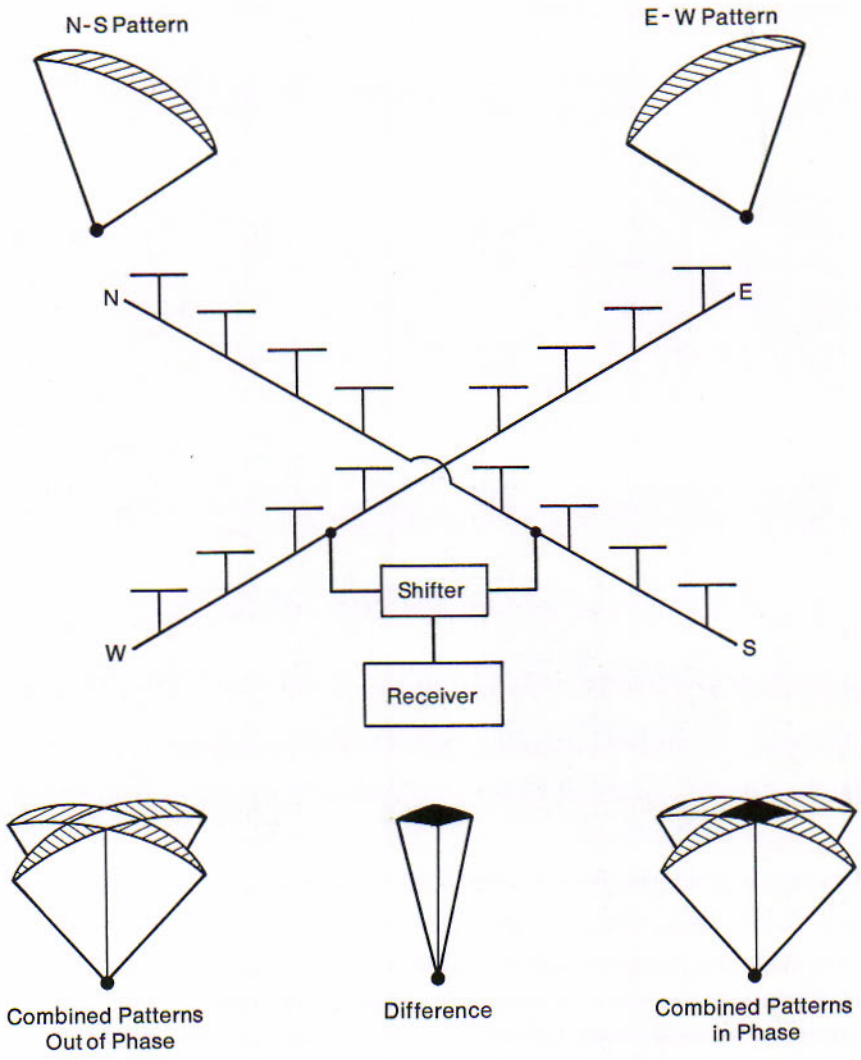


Figure 4.31B
Mills Cross diagram. Subtracting the out-of-phase signal from the in-phase signal results in a very narrow beam giving high resolution.

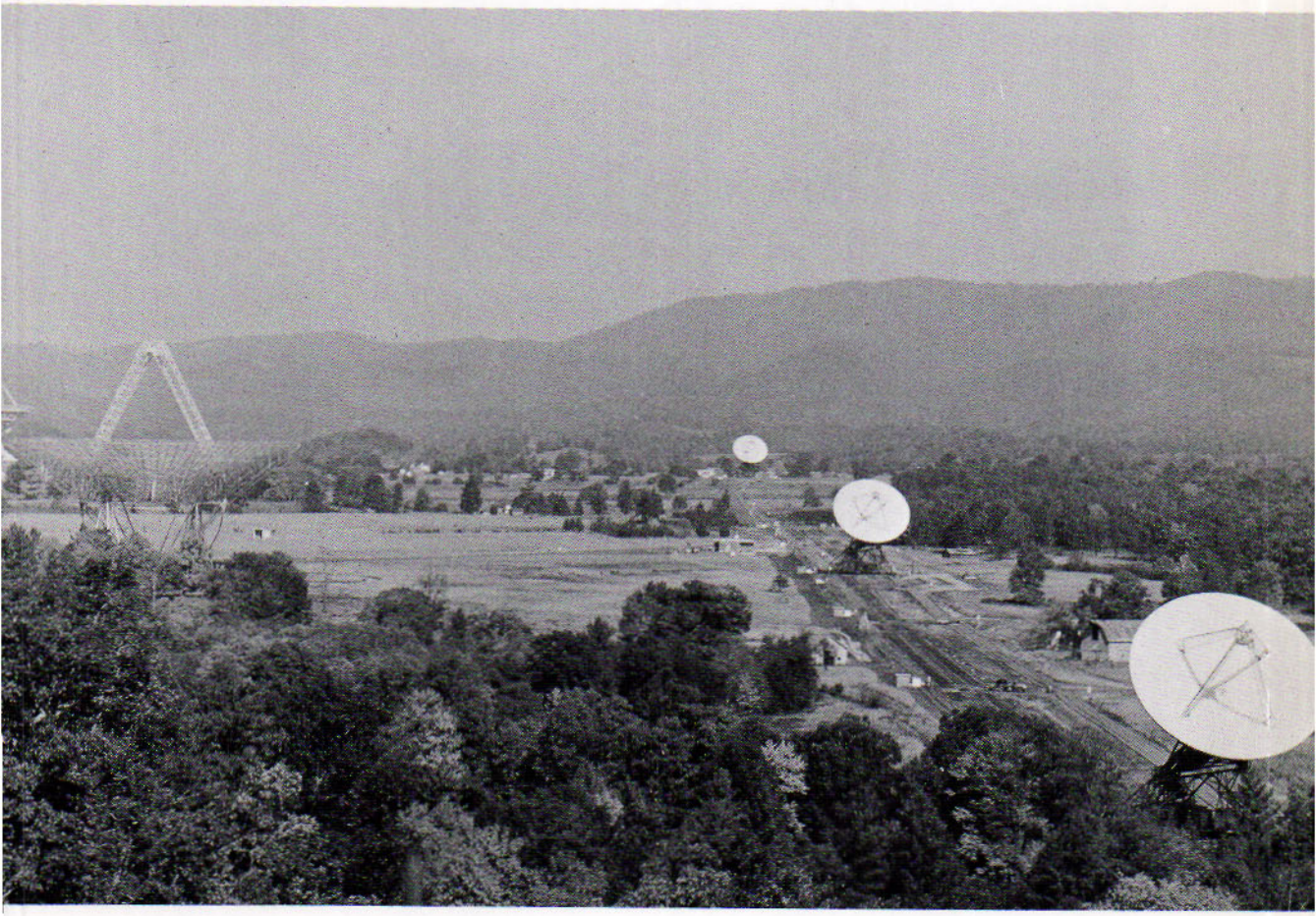
the observing time required. Thus a Mills Cross with the cross staff offset to form an L, can be made, by phase switching into an aperture synthesis instrument.

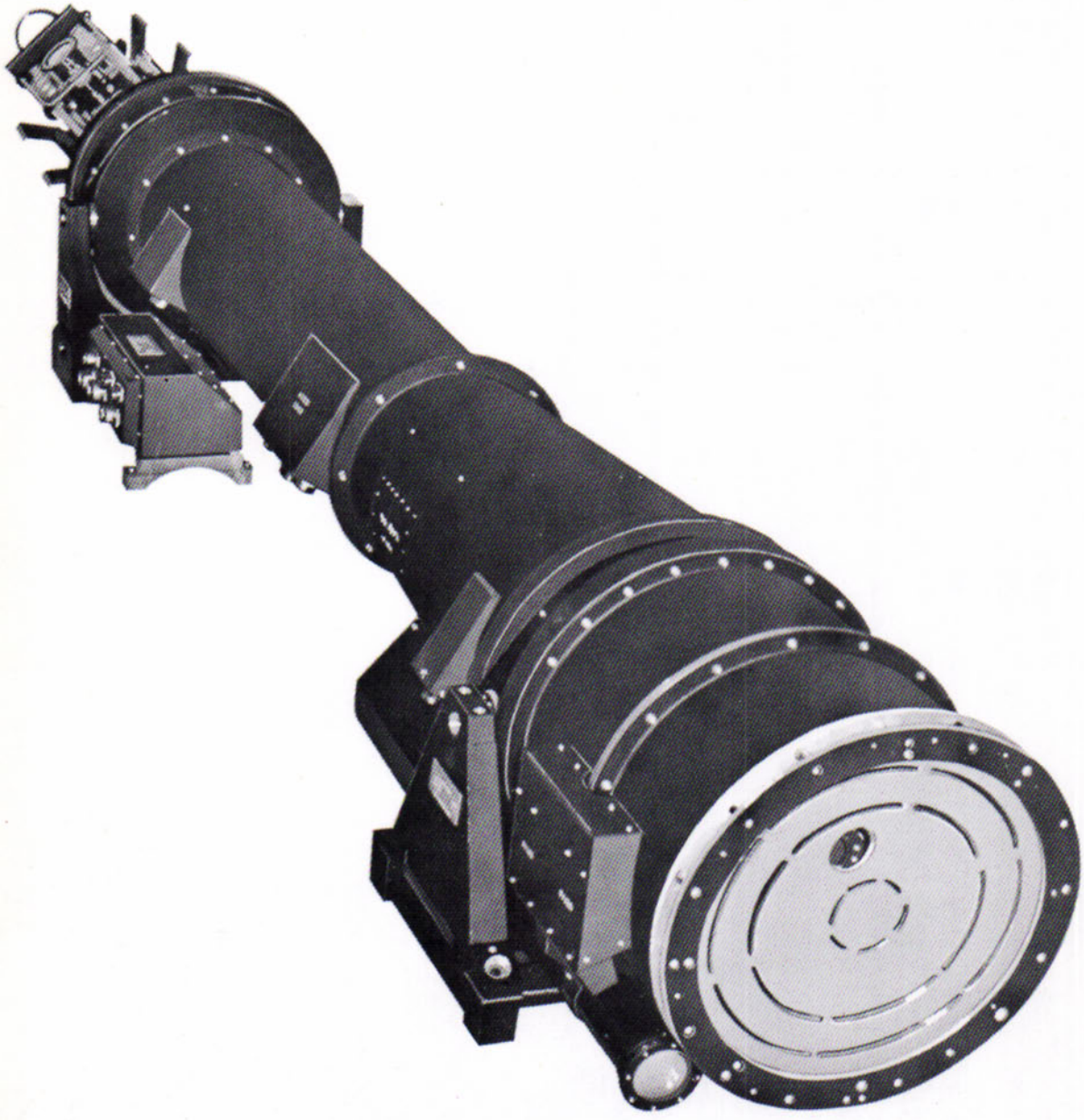
The "final" step has been to supersynthesize the array by using several telescopes at various positions and using the rotation of the earth to sample the Fourier components. One rotation of the earth with the appropriate delays for baseline projections at a given spacing gives a

set of Fourier components. One or two elements in the array are then set at different positions and the earth's rotation builds up another set of components until all of the components are obtained to synthesize a map of the brightness distribution.

The first supersynthesis array was developed at Cambridge, England. NRAO has used their interferometer as a supersynthesis instrument (Fig. 4.31C) and the Dutch opened a very extensive supersynthesis telescope at Westerborg in 1970.

Figure 4.31C
The National Radio Astronomy Observatory's three-element interferometer, also used as an aperture synthesis instrument. *(NRAO Photograph)*





A grazing incidence X-ray telescope designed for the Apollo manned skylab. Overall length: 284 cm. *(American Science and Engineering, Inc., courtesy of R. Giaconi).*

X-RAY ASTRONOMY

At this writing the youngest branch of observational astronomy is in the X-ray region of the electromagnetic spectrum. The wavelengths in this region range from 0.1Å to 100Å.

These telescopes are really "stacked" detectors contrived to signal the penetration of radiation of a given energy range and having an electronic anti-coincidence circuit to be sure that it does not count penetrations from the wrong direction or crossing only one detector. The simplest such device is a series of Geiger Counter tubes shielded from penetration from the side (Figure 4.32). The anti-coincidence circuit is arranged so that tube A must detect an event before tube C or it is not counted. Such telescopes are flown in balloons, rockets and satellites and have discovered many discrete sources of X rays.

A more advanced form of X-ray telescope resembles the more conventional telescope. X rays are reflected at angles near grazing incidence so it is possible to use paraboloidal surfaces to focus X rays on a detector yielding an image of the source. So far efforts to make such telescopes have been less than satisfactory but theoretically they should be the ideal X-ray telescopes.

4.32
X-ray Telescopes

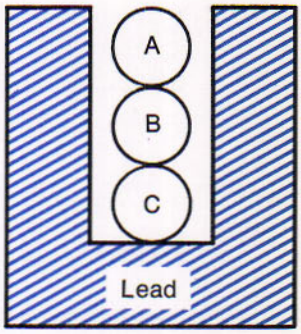


Figure 4.32

REVIEW QUESTIONS

1. Now that we have noted the effect of atmospheric refraction in increasing the altitudes of the stars, it is in order to re-examine some conclusions of Chapter 1. Modify the following statements:
   (a) The duration of sunshine is 12 hours on days when the sun is at an equinox.
   (b) The polar caps in which the stars never set or never rise have radii equal to the observer's latitude.
   (c) The midnight sun is visible only within the arctic and antarctic circles. (State a reason in addition to refraction why this statement is not quite correct.)
   (d) As seen from the north pole the celestial equator coincides with the horizon.
2. Describe what occurs when a narrow beam of white light is passed through a glass prism. Suppose that it is monochromatic red light instead.
3. Why has the grating replaced the prism in modern spectrographs?
4. Describe the appearance of: (a) The continuous spectrum; (b) the bright-line

spectrum; (c) the dark-line spectrum. What is the nature of the source producing each kind of spectrum?

5. State and explain the effect in the spectrum when the source is approaching or receding from the observer.
6. When two double convex lenses are appropriately separated to form a simple refracting telescope, what does each lens contribute?
7. Why does a single lens as an objective produce a blurred image? How is the definition improved by an achromatic objective?
8. Distinguish between the Schmidt telescope and the ordinary type of reflecting telescope.
9. Why are all very large optical telescopes reflecting telescopes?
10. The following questions are among those asked by visitors in the dome of an observatory. How would you answer them?

    (a) How much does this telescope magnify?
    (b) Does it magnify stars as well as planets?
    (c) How far can you see with the telescope?
    (d) How can you point the telescope to a star that is invisible to the naked eye?
    (e) Why is this telescope better than a smaller one?

11. Explain that the resolving power of a paraboloidal radio telescope is more limited than that of an optical telescope of equal aperture.
12. Mention some advantages of the radio telescope in studies of the heavens.

PROBLEMS

1. The objective of a 1 meter telescope has a focal length of 15 meters. How much brighter does a star appear through this telescope than to the unaided eye if the aperture of the lens of the eye is 5 millimeters?
   Answer: 40,000 times.
2. What is the magnifying power of this telescope with an eyepiece of 3 cm focal length?
   Answer: 500.
3. What is the theoretical resolving power of this telescope?
   Answer: 0.″14.
4. What is the diameter of the moon's image as photographed with this telescope? The moon's angular diameter is 0°.5.
   Answer: The diameter is slightly more than 13 cm.
5. A line at the wavelength of 4000 angstroms (10.21) in the spectrum of a star is displaced 1 angstrom to the red. Calculate the star's velocity in the line of sight (11.9).
   Answer: 74.9 km a second, recession.

REFERENCES

Ingalls, Albert G., ed., *Amateur Telescope Making, Book 1,* New York: Scientific American, 1959.

Amateur Telescope Making, Advanced, New York: Scientific American, 1959.

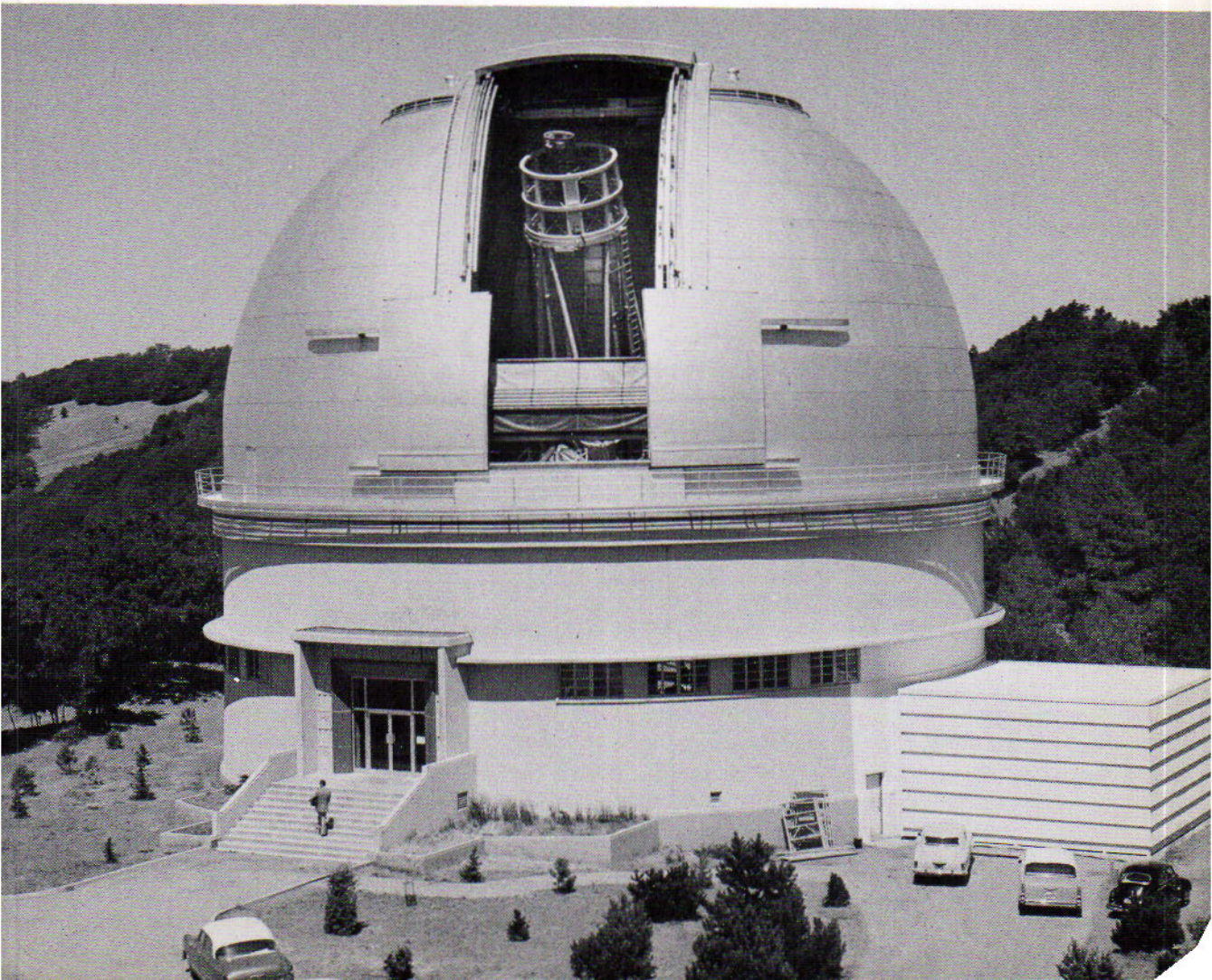
Amateur Telescope Making, Book 3, New York: Scientific American, 1956.

King, Henry C., *The History of the Telescope,* Cambridge, Mass.: Sky Publishing Co., 1955.
Kopal, Zdnek, *Telescopes in Space,* London: Faber & Faber, 1968.
Miczaika, G. R. and William M. Sinton, *Tools of the Astronomer,* Cambridge, Mass.: Harvard University Press, 1961.

FOR FURTHER STUDY

Kuiper, Gerard P. and Barbara M. Middlehurst, eds. *Telescopes,* Chicago: University of Chicago Press, 1960.
Strong, John, *Concepts in Classical Optics,* San Francisco: Freeman, Cooper & Co., 1958.

Dome of the 120-inch (305 cm) telescope of the Lick Observatory at Mt. Hamilton, California. *(Lick Observatory Photograph).*



5

THE MOON

MOTIONS OF THE MOON — THE MOON'S SURFACE FEATURES — THE TIDES

The earth is accompanied in its revolution around the sun by its single satellite, the moon, which is 3476 Kilometers in diameter, or a little more than one fourth the earth's diameter. Although it ranks only sixth in size among the satellites of the solar system, the moon is larger and more massive in comparison with the earth than is any other satellite with respect to its primary. The earth-moon system has more nearly the characteristic of a double planet. Our knowledge of the moon is being expanded at a rapid pace due to the use of manned and unmanned spacecraft.

The Moon from lunar orbit. *(NASA photograph)*

MOTIONS OF THE MOON

5.1 Revolution of the Earth and Moon Around the Sun

The earth's revolution around the sun has been hitherto described without reference to the influence of the moon. Because the earth and moon mutually revolve around their common center of mass once in nearly a month while they are making the annual journey around the sun, the orbit of each one relative to the sun is slightly wavy. What we have called the "earth's orbit" is strictly the orbit of the center of mass of the earth-moon system. Imagine the earth and moon joined by a stout rod between their centers; the *center of mass* is the point of support of the rod for which the two would balance.

$m_1a_1 = m_2a_2$

If the earth and moon were equally massive, the center of mass would be halfway between their centers. Very slight shifts in the directions of the nearest planets during the month have revealed that the center of mass is in fact only 4667 km from the earth's center toward the moon and is, therefore, within the earth. Evidently the moon is much the lighter of the two. A simple calculation shows that its mass is a little less than $\frac{1}{80}$ of the earth's mass.

Astronomical diagrams are often unable to keep the same scale of distances throughout, or of distances and sizes of the celestial bodies. If the distance between the earth and sun in Fig. 5.1 were made equal to the length of the printed page, the distance between the earth and moon on that scale would scarcely exceed the diameter of a period on the page. A drawing exactly to scale would show that the annual orbits of both the earth and moon are always concave to the sun.

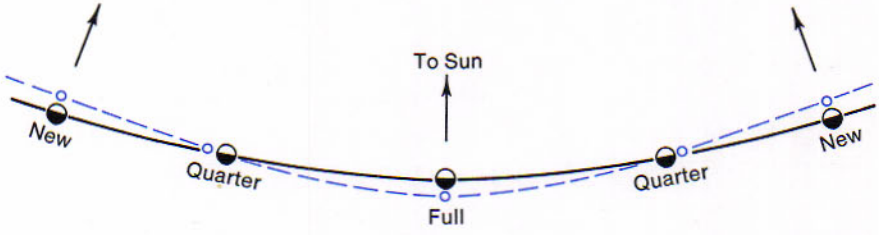


Figure 5.1
Orbits of the earth and moon around the sun.

5.2 The Moon's Orbit Relative to the Earth

The moon's revolution may be considered in three ways: (1) Its annual motion around the sun, which is disturbed by the attraction of the earth; (2) its monthly motion around the center of mass of the earth and moon, in which the sun is the chief disturbing factor; or what amounts to nearly

the same effect in this case: (3) its monthly motion with respect to the earth's center. It is this relative motion with which we are especially concerned. The *orbit of the moon,* for most purposes, is its path around the earth; it is an ellipse of small eccentricity e = 0.055, having the earth's center at one focus.

The moon's speed in this orbit averages slightly more than one kilometer a second. By the law of equal areas (3.6), which applies to all revolving celestial bodies, the speed is greatest at *perigee,* where the moon is nearest the earth, and is least at *apogee,* where it is farthest from the earth. The major axis, or *line of apsides,* revolves eastward once in about 9 years. This is one of the many variations in the moon's orbit that arise mainly from the influence of the sun. The size of the orbit has been determined by measuring the moon's parallax.

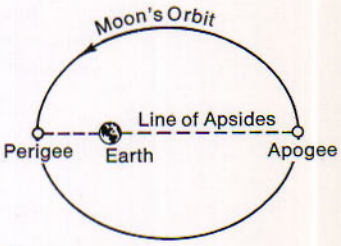


Figure 5.2
The moon's orbit relative to the earth. The orbit is an ellipse of small eccentricity (much exaggerated in the diagram) with the earth at one focus.

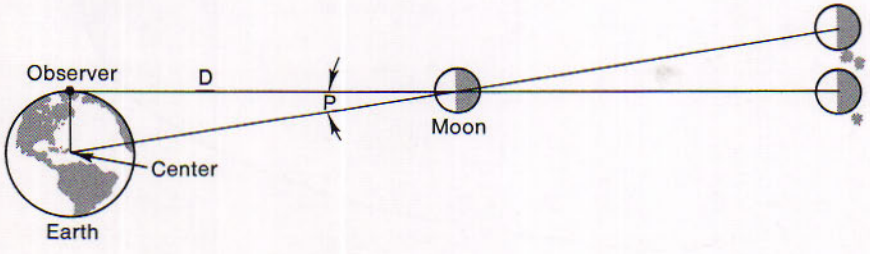
5.3 Parallax; Relation to Distance

Parallax is the difference in direction of an object as viewed from two places, or from the two ends of a base line. As an example of the parallax effect we may note the shifting of a nearby object against a distant background when the eyes are covered alternately. For the same base line (B) the parallax (p) becomes smaller as the distance (D) of the object is increased. When the parallax of an object is measured, its distance may be calculated, supposing that the direction of the object is perpendicular to the base line as seen from one end of that line.

$$D = \frac{B}{\sin p}$$

Here we have a means of measuring the distances of inaccessible objects such as the celestial bodies. The parallax of the relatively nearby moon can be determined by simultaneous observations of its positions among the stars from two places on the earth a known distance apart. Whatever stations are used, the observed parallax is standardized by calculating from it the parallax that would have resulted if the base line had been the earth's equatorial radius and the moon had been on the horizon. This equatorial horizontal parallax is regarded as the *parallax* of the moon.

Figure 5.3
The moon's directions differ by nearly its breadth as viewed from New York and San Francisco, or twice its diameter as viewed from the radius of the earth.



5.4 The Moon's Distance

The parallax of the moon at its mean distance from the earth is 57′2.″62. By the preceding formula the mean distance between the centers of the earth and moon is 384,404 km. This value is about 60¼ times the earth's equatorial radius. The distance at any particular time may differ considerably from the average because of the eccentricity of the moon's orbit and of variations in it. The distance at perigee may be as small as 356,400 km and at apogee may be as great as 466,700 km.

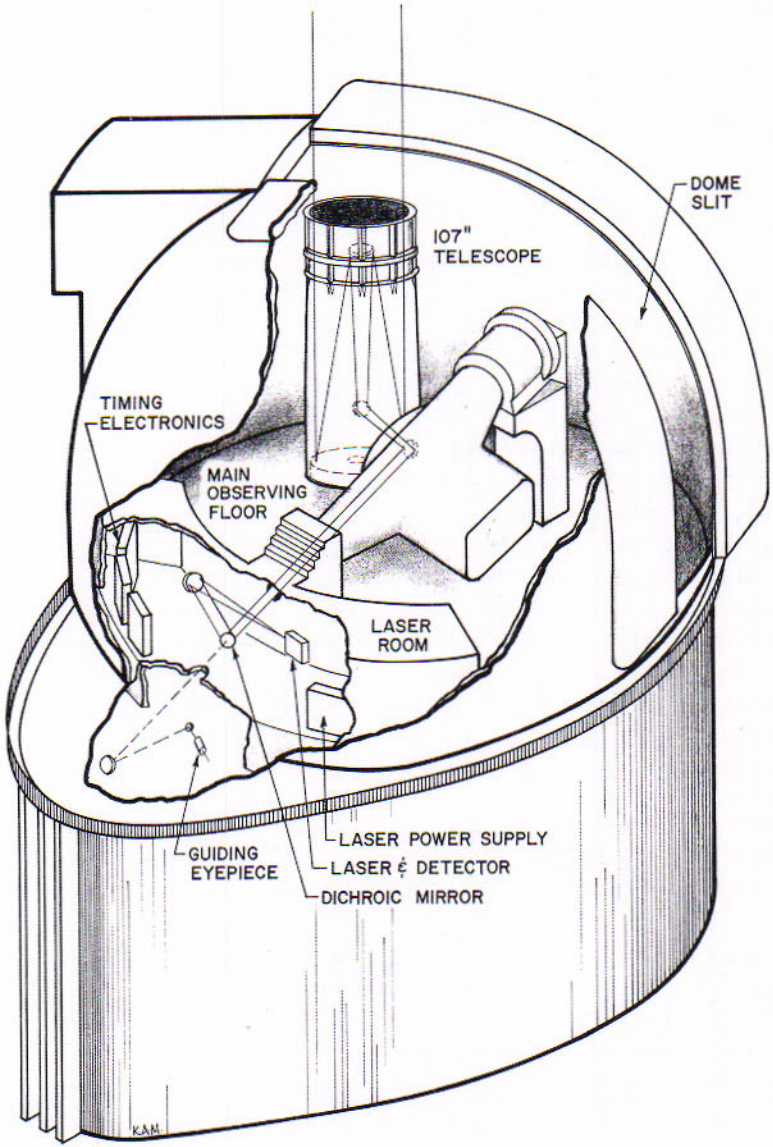


Figure 5.4
Experimental setup for laser ranging tests at the McDonald 107-inch reflector. *(Diagram by the McDonald Observatory).*

The moon's distance from us as light travels is 1.28 light seconds; this is the distance in miles divided by 300,000 km a second, the speed of light. In 1946, the U.S. Army Signal Corps beamed radar pulses toward the moon and received the echo of each pulse 2.56 seconds after it was sent out. This pioneer experiment has been repeated frequently. Naval Research Laboratory scientists in 1957 measured the two-way travel times of 60,000 pulses of 10-cm radio waves reflected from the moon. They found for the moon's mean center-to-center distance from the earth 384,404 km with an uncertainty of 1 km. This value is in close agreement with the distance found by triangulation.

Lick and McDonald Observatory astronomers using a laser retroreflector installed by Apollo 11 astronauts on 20 July 1969 have determined the distance to the moon at the moment of observation to be 384,404.377 km with an uncertainty of 0.001 km. As timing techniques improve the uncertainty should decrease even further. If additional retroreflectors are installed, a good value for the shape of the moon may be obtained and the distance to the geometrical center derived.

5.5 Aspects of the Moon

In its monthly revolution around us the moon moves continuously eastward relative to the sun's place in the sky. The moon's *elongation* at a particular time is its angular distance from the sun. Special positions receive distinctive names and are known as the *aspects* of the moon.

When the moon overtakes the sun, generally passing it a little to the north or south, the elongation is not far from 0°. The moon is in *conjunction* with the sun when the two bodies have the same celestial longitude. It is in *quadrature* when its elongation is 90° either east or west. The moon is in *opposition* when its celestial longitude differs by 180° from that of the sun, so that its elongation is not far from 180°.

Aspects of the planets relative to the sun are similarly reckoned. For the conjunctions of the planets with the moon and with one another, however, the predictions in the Diary of the *American Ephemeris and Nautical Almanac* are the times to the nearest hour when the two bodies have the same right ascension.

5.6 The Moon's Phases

The changing figures of the waxing and waning moon are among the most conspicuous of celestial phenomena and were among the first to be understood. The moon is a dark globe shining only by reflected light. As it revolves around the earth, its sunlit hemisphere is presented to us in successively increasing or diminishing amounts. These are the *phases* of the moon.

It is the *new moon* that passes between the sun and the earth, and its dark hemisphere is toward us. The moon is invisible at this phase except when it happens to cross directly in front of the sun's disk, causing an

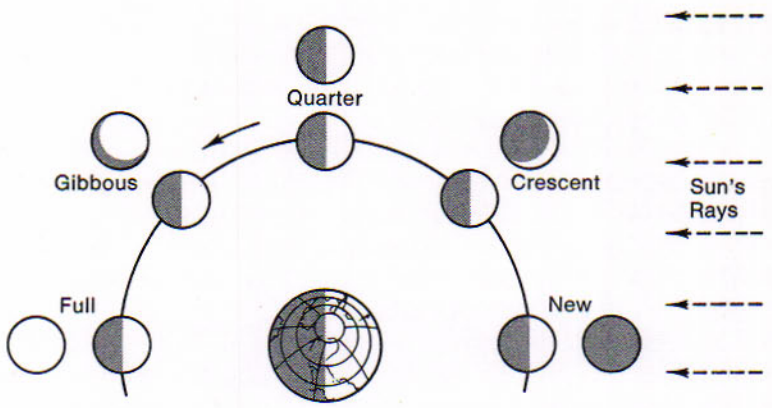


Figure 5.6
The phases of the moon. The outer images show the phases as seen from the earth.

eclipse of the sun. On the second evening after the new phase the thin *crescent* moon is likely to be seen in the west after sundown; this was the signal for the beginning of the new month in the early lunar calendars. The crescent becomes thicker night after night, until at the *first quarter* the sunrise line runs straight across the disk. Then comes the *gibbous* phase as the bulging sunrise line gives the moon a lopsided appearance. Finally, a round *full moon* is seen rising in the east at about nightfall.

The phases are repeated thereafter in reverse order as the sunset line moves across the disk; these are gibbous, *last quarter,* and new again. The moon's *age* is the interval at any time since the preceding new moon.

The *limb* of the moon is the edge of the moon. The *terminator* is the line between the bright and dark hemispheres of the moon; it is the line of the sunrise before the time of the full moon and of the sunset thereafter. Aside from irregularities in its course, which are caused by the mountainous character of the lunar surface and are often noticed without the telescope, the terminator generally appears elliptical, because it is a circle seen in projection. The full circle coincides with the edge of the moon at the full phase, whereas at the quarter phases it is turned so that it runs straight across the disk. Quarter phase occurs slightly before quadrature (one quarter of the orbit) due to the finite distance of the sun. Aristarchus used this fact to estimate the distance to the sun around 250 B.C.

The horns, or *cusps,* of the crescent moon point away from the sun and show nearly the course of the ecliptic in the moon's vicinity. Thus the positions of the thin crescent near moonset and moonrise depend on the angle then between the ecliptic and horizon (1.19). It is left to the reader

to explain why the horns are more nearly vertical after sunset in the spring than in the autumn, as viewed in middle latitudes.

When the moon is in the crescent phase, the rest of its disk is made visible by sunlight reflected by the earth. The brighter crescent seems to have a greater diameter than the earthlit part of the disk and so to be wrapped around it. This illusion of the difference in scale between the two parts becomes more striking as the quarter phase is approached, although by this time the earthlight has faded almost to invisibility.

The earth exhibits the whole cycle of phases in the lunar sky, and these are supplementary to the moon's phases in our skies. "Full earth" occurs there at the time of the new moon. Full earthlight on the moon is many times as bright as is the light of the full moon on the earth. The earth is not only a larger mirror to reflect the sunshine, but it is also a more efficient one because of its atmosphere and clouds. Full earthlight at the distance of the moon is about as bright as twilight on earth.

Earthlight on the moon is bluer than the direct sunlight, for much of the earthlight is selectively reflected by our atmosphere, and in this light the blue of the sky appears.

5.7 Earthlight on the Moon

Figure 5.7
Earthlight on the moon's surface, at the crescent phase. *(Yerkes Observatory Photograph).*

5.8 The Sidereal and the Synodic Month

Astronomically, the month is the period of the moon's revolution around the earth. As in the cases of the day and year the different kinds of month depend on the different points in the sky to which the motion is referred. The *sidereal month* is the true period of the moon's revolution; it is the interval between two successive conjunctions of the moon's center with the same star, as seen from the center of the earth. Its length

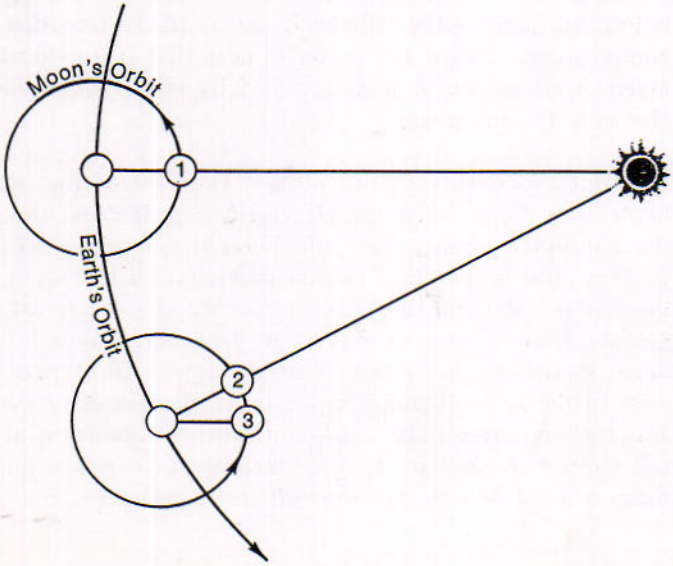


Figure 5.8
The synodic month is longer than the sidereal month. Between positions 1 and 2 the moon has made one revolution, completing the sidereal month. The synodic month does not end until the moon has reached position 3.

averages $27^d\ 7^h\ 43^m\ 11^s.5$, or nearly 27⅓ days, and varies as much as 7 hours because of perturbations of the moon's motion.

The *synodic month* is the interval between successive conjunctions of the moon and sun, from new moon to new moon again. This month of the phases is longer than the sidereal month by more than 2 days, the additional time the moon requires to overtake again the slower-moving sun. The length of the synodic month averages $29^d\ 12^h\ 44^m\ 2^s.9$, or a little more than 29½ days, and varies more than half a day.

The moon's eastward progress among the constellations in 1 day averages 360° divided by the length of the sidereal month, which equals 13°.2. The moon's motion in 1 hour is accordingly a little more than half a degree, or slightly more than its own diameter.

5.9 The Moon Rises Later from Day to Day

When the eastward motion is considered, as we have just now done, the moon overtakes the sun at intervals of the synodic month. With respect to the diurnal motion of the heavens, however, the moon keeps falling behind the sun, so that it returns to upper transit 28.5 times in 29.5 solar days. The interval between upper transits is 29.5/28.5 times 24 hours, or about $24^h\ 50^m$ of mean solar time. Thus the moon crosses the celestial meridian about 50 minutes later from day to day in the average.

The daily retardation of moonrise also averages about 50 minutes, but the actual retardation may differ greatly from this value, especially in high latitudes. In the latitude of New York the greatest possible delay may exceed the least by more than 1 hour. The variation depends mainly on the angle between the moon's path, which is not far from the ecliptic, and the horizon; the smaller the angle at moonrise, the less is the delay in rising from day to day. As we have already noted (1.19), the ecliptic is least inclined to the horizon in our northern latitudes when the vernal equinox rises. When the moon is near that point, its rising is least delayed, a circumstance that is especially conspicuous when the moon is also near the full phase.

5.10 The Harvest Moon

The full moon that occurs nearest the *time* of the autumnal equinox, September 23, is called the Harvest Moon. Because the sun is then near the autumnal equinox, the full moon is near the position of the vernal equinox and is therefore in that part of its path that is least inclined to the horizon at moonrise. The peculiarity of the harvest moon, as distinguished from other occasions when the moon is near full, is its minimum delay in rising for a few successive nights. Thus there is bright moonlight in the early evening for an unusual number of evenings in middle and higher northern latitudes conveniently occurring at the time of the fall harvest. A similar effect is observed in corresponding southern latitudes around March 21. The full moon following the Harvest Moon is

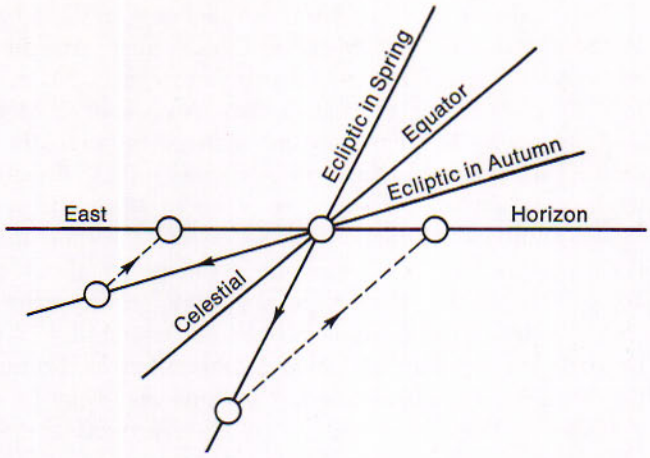


Figure 5.10
Explanation of the harvest moon. Because of its eastward motion along its path, which nearly coincides with the ecliptic, the moon rises later from night to night. For the nearly full moon, the delay is least in our northern latitudes in autumn, when the ecliptic is least inclined to the horizon at moonrise.

known as the Hunter's Moon for much the same reason and is convenient for the fall hunting.

5.11 The Moon's Apparent Path; Regression of the Nodes

The moon's path on the celestial sphere during a month is nearly a great circle that is inclined about 5° to the ecliptic. The path among the constellations for a particular month may be traced by plotting on a celestial globe the right ascensions and declinations of the moon's center from tables in the *American Ephemeris and Nautical Almanac.* When the plotting is continued through successive months, it is seen that the path shifts westward rather rapidly, keeping about the same angle with the ecliptic.

The *nodes* of the moon's path are the two opposite points where it intersects the ecliptic. The *ascending node* is the point where the moon's center crosses the ecliptic from south to north; the *descending node* is the point where it crosses from north to south. *Regression of the nodes* is their westward displacement along the ecliptic, just as the equinoxes slide westward in their precessional motion, but at a much faster rate; a complete revolution of the nodes of the moon's orbit is accomplished in 18.6 years. From this and other changes in the moon's orbit, for which the sun's attraction is mainly responsible, the moon's course among the constellations is considerably different from month to month, although it always remains within the confines of the zodiac.

5.12 Ephemeris Time by the Moon

The moon's monthly motion around the heavens is rapid and independent of the earth's variable rotation. It provides an important means of determining time on a uniform basis. Positions of the moon have been observed for a long time on occasions when it transits the meridian or when it passes in front of stars (6.13). Because the opportunities for such observations are restricted, astronomers have sought more convenient and accurate means of determining time by the moon-clock.

The dual-rate moon position camera was designed for the purpose by W. Markowitz at the U.S. Naval Observatory. Attached at the focus of a telescope of moderate size, its plate carriage is driven by a motor to follow the stars in the diurnal motion. A dark filter of glass placed before an opening at the center of the carriage is gradually tilted by a motor, shifting the moon's image optically so as to hold it stationary with respect to the stars during the exposure of about 20 seconds.

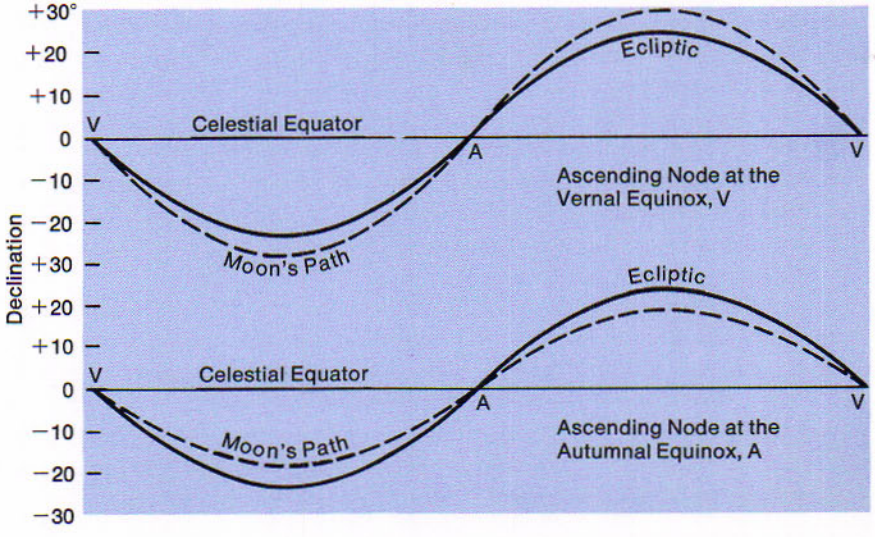
The photograph, showing the moon and the stars around it, is observed with a special device where the positions of about 10 stars and 30 to 40 points on the moon's bright edge are measured. Corrections for irregularities of the edge have been determined by C. B. Watts. The final result is the right ascension and declination of the moon's center at the universal time of observation. This time may then be compared with the ephemeris time when the moon is scheduled to reach this observed position.

The operation of the dual-rate camera, which began in 1952 with the 12-inch refractor of the Naval Observatory, is being extended to observatories in other parts of the world. It can supply uniform ephemeris time and its relation to the variable universal time, and it will also improve the understanding of the complex motions of the moon.

5.13 The Moon's Range in Declination

Because the moon's path on the celestial sphere departs only a little from the ecliptic, the moon moves north and south during the month about as much as the sun does in the course of the year. Near the position of the summer solstice the moon rises in the northeast, sets in

Figure 5.13
Effect of regression of the nodes of its path on the moon's range in declination.



the northwest, and is high in the sky in our northern latitudes at upper transit. Near position of the winter solstice about 2 weeks later, the moon rises in the southeast, sets in the southwest, and crosses the meridian at a lower altitude. An example of the many compensations in nature is furnished by the full moon which, being opposite the sun, rides highest in the long winter nights and lowest in the summer.

When the inclination of the moon's path to the ecliptic is taken into account, we note that the range in declination varies perceptibly as the nodes regress. When the ascending node coincides with the vernal equinox, the moon's path is inclined to the celestial equator 23½° plus 5°, or 28½°; this occurred in 1968. When the ascending node coincides with the autumnal equinox, which occurred in 1959, the inclination to the equator is 23½° minus 5°, or 18½°. Thus the moon's highest and lowest altitudes at upper transits in latitude 40° north average in the first case 78½° and 21½°, respectively, and in the second case 68½° and 31½°—a decrease in range of about 20°.

The variation of 10° in the moon's maximum declinations north and south in the 18.6-year cycle is chiefly responsible for nutation, the nodding of the earth's axis which accompanies its precessional motion.

5.14 The Moon's Rotation and Librations

The moon rotates on its axis in the same period in which it revolves around the earth, namely, the sidereal month of 27⅓ days. In consequence of the equality of the two periods the moon presents about the same hemisphere toward the earth at all times. It is always the face of the "man in the moon" that we see near the full phase and never the back of his head. An examination of the moon's surface throughout the month, however, shows that features near the edge of the disk are turned sometimes into view and at other times out of sight. The moon seems to rock slightly; and these apparent oscillations, or *librations*, arise mainly from three causes:

1. *The libration in latitude* results from the inclination of about 6½° between the moon's equator and the plane of its orbit. At intervals of 2 weeks the lunar poles are tipped alternately toward and away from us; at times we can see 6½° beyond the north pole, at other times the same distance beyond the south pole. The explanation of this libration is analogous to that of the seasons.

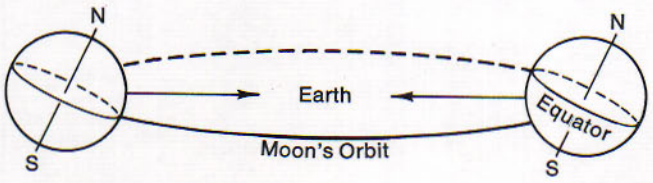
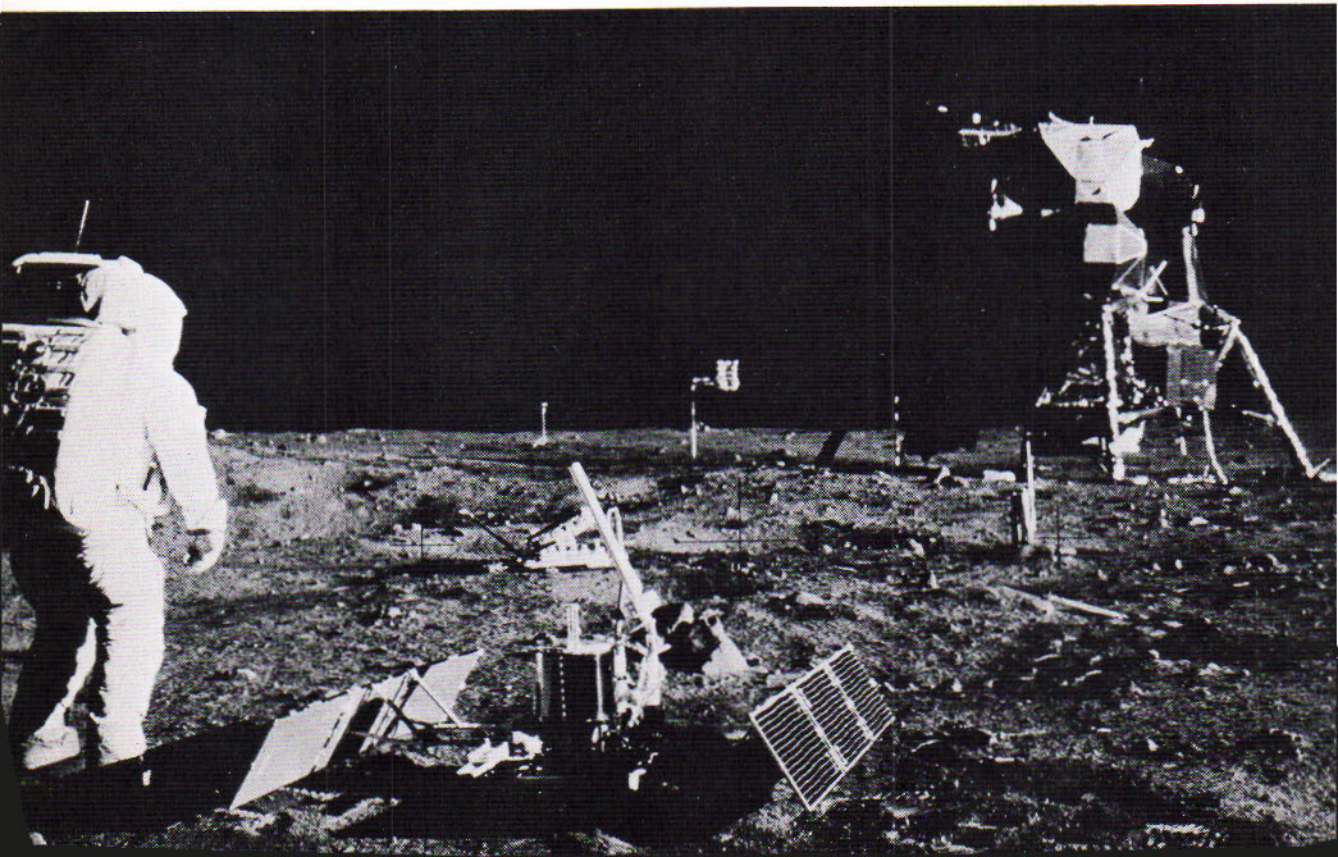


Figure 5.14A
The moon's libration in latitude. The inclination of the moon's equator to the plane of its orbit causes the moon's poles to be presented alternately to the earth.

2. *The libration in longitude* is caused by the failure of the moon's rotation and revolution to keep exactly in step throughout the month, although they come out together at the end. The rotation is nearly uniform, whereas the revolution in the elliptical orbit is not uniform (5.2). Thus the moon seems to rock in an east-west direction, allowing us to see as much as 7¾° farther around in longitude at each edge than we could otherwise.
3. *The diurnal libration* is a consequence of the earth's rotation. Even if the other librations were absent, so that the same hemisphere were turned always toward the center of the earth, we on the surface view the moon from slightly different directions during the day, and therefore see slightly different hemispheres. From the elevated position nearly 4000 miles above the center of the earth the observer can see about 1° farther over the western edge at moonrise and the same amount over the eastern edge at moonset.

Figure 5.14B
Lunar experiments. E. Aldrin finishing deployment of the seismometer carried by Apollo 11. Behind this equipment is the laser retroreflector; direction to the earth is indicated by the tilt of the reflector's face and the direction of the seismometer's antenna. Spacecraft Eagle is in the background. *(NASA photograph).*



In addition to the principal librations, there is a slight physical libration, because the moon's rate of rotation is not quite uniform. Fully 59 per cent of the moon's surface has been visible when the sidereal month is completed. The remaining 41 per cent is never seen from the earth; and throughout this region, of course, the earth would always be invisible to lunar observers.

Measuring the librations has always presented a challenge. Most of our information is derived from measuring well defined points on the moon with respect to the moon's limb and bright stars that appear in the nearby sky. A more accurate technique is to use well identified radar reflection points or even better using laser ranging techniques and *retro-reflectors* placed at various points on the moon. A start in this direction was made by the astronauts of Apollo 11 when they placed a retro-reflector on the moon. Unfortunately, Apollo 12 did not transport similar equipment to its landing site.

THE MOON'S SURFACE FEATURES

The unaided eye can discern only the dark areas of the moon's surface, which are known as the lunar seas, and occasional irregularities of the terminator, which suggest that the moon is mountainous. The telescope shows the mountains themselves and other details of the surface. The mountains are clearest near the terminator, either the sunrise or sunset line, where shadows are long and the contrast between mountain and plain is therefore more pronounced. Spacecraft, manned and unmanned, have observed and photographed the Moon in a detail unachievable by telescope.

5.15 Absence of Atmosphere on the Moon

There is no evidence that the moon has any permanent atmosphere. There is no twilight; the sunrise and sunset lines are abrupt divisions between day and night. The effect of twilight in prolonging the cusps of the crescent moon beyond the diameter could be observed if a lunar atmosphere had a density 10^{-4} that of the earth's atmosphere. No perceptible haze dims our view of the moon, even near the edge where an atmosphere would be most effective in this respect. When a star is occulted by the moon, it does not first become fainter and redder, as it

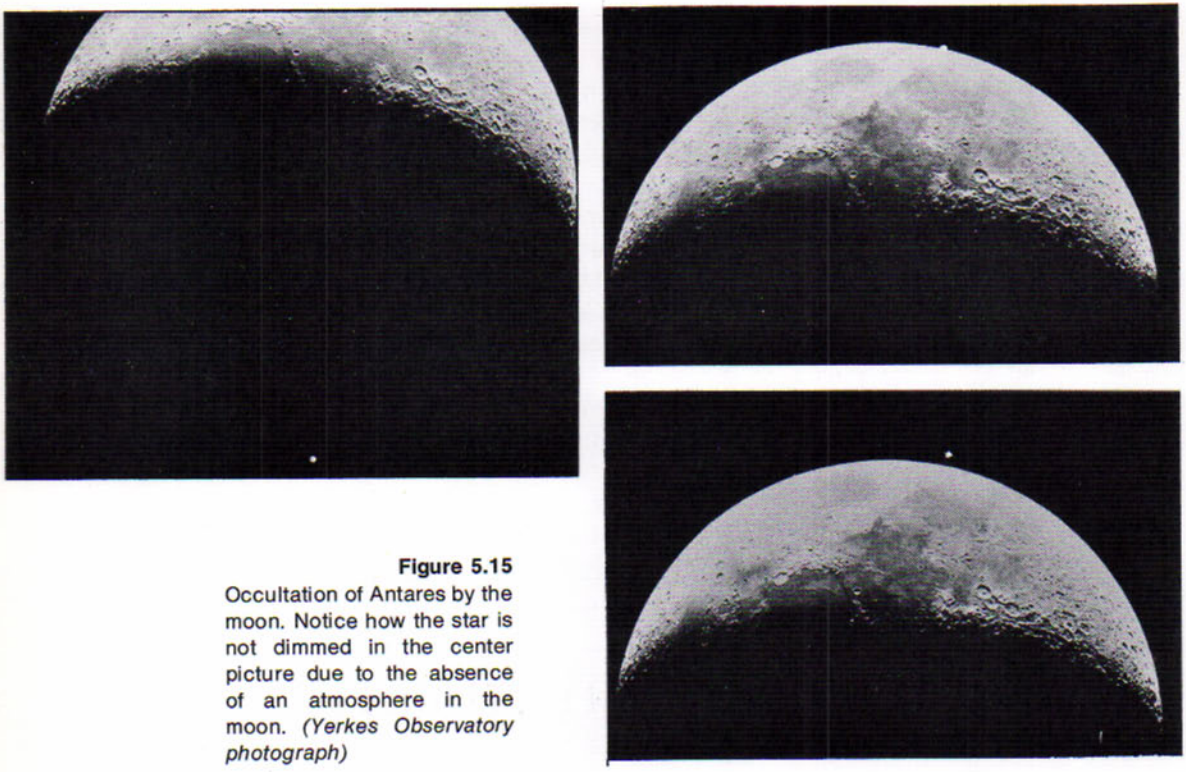


Figure 5.15
Occultation of Antares by the moon. Notice how the star is not dimmed in the center picture due to the absence of an atmosphere in the moon. *(Yerkes Observatory photograph)*

would behind a considerable amount of atmosphere. Instead, the star retains its normal brightness and color until it disappears almost instantly at the edge of the moon. A detailed analysis of polarimetric data yields an upper limit of 10^{-10} times that of the earth's atmosphere. A more delicate analysis for lunar aurora yields an upper limit of 10^{-15} and a very detailed study at radio wavelengths of an occultation of the Crab nebula by the moon gives an upper limit of something like 6×10^{-13} times the earth's atmosphere. Thus, except for possible temporary gaseous emissions we can conclude that the moon has no effective atmosphere.

These are indications that the moon has only an extremely rare atmosphere, at the most. The reason is found in the escape of gases from the weak control of the moon's attraction.

5.16 Escape of an Atmosphere

The molecules of a gas are darting swiftly in all directions and are incessantly colliding, so that some are brought momentarily almost to rest while others are propelled to speeds far exceeding the average. The speeds increase as the temperature of the gas is raised, and at the same temperature are greater for lighter gases than for heavier ones.

The *kinetic theory of gases* states that the average squared velocity of the molecules varies directly as the absolute temperature (T) of the gas and inversely as its mean molecular weight m. The average speed in km/sec at 0°C is 1.9 for hydrogen, 0.6 for water vapor, and 0.5 for nitrogen and oxygen. These speeds become 17 per cent greater at 100°C.

The ability of a celestial body to retain an atmosphere around it depends on the *velocity of escape* at its surface. This is the initial speed a molecule or any other object must have in order to overcome the pull of gravity and to get away. Calculation suggests that a celestial body will lose half its atmosphere in only a few weeks if the velocity of escape does not exceed 3 times the mean speed of the molecules in that atmosphere. The required time is increased to a few thousand years if the factor is 4, and to a hundred million years if the factor is 5 times the mean speed of the molecules.

The velocity of escape is about 11.2 km a second near the earth's surface, without allowance for air resistance, but is only 2.4 km a second near the surface of the moon. We conclude that the earth can retain the chief constituents of its atmosphere for an indefinite period, whereas the moon has been unable to keep any of these gases around it.

$V^2 = \frac{3kT}{m}$

m = mass of molecule

k = Baltzman constant

$V^2_{esc} = G(M + m)\frac{2}{r}$

M = mass of moon

r = radius of moon

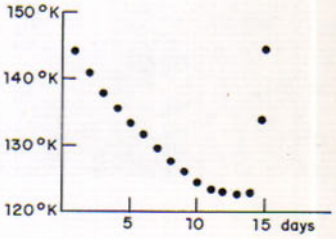
°C = °K–273

5.17 The Character of the Moon's Surface

The surface temperature of the moon varies from more than 100°C when the sun is overhead to −50°C at sunset and is reduced to −150°C at midnight. These values were determined by Pettit and Nicholson at Mount Wilson Observatory, who also observed a drop of 150°C in the temperature in 1 hour during a lunar eclipse. Similar values were obtained by the spacecraft Surveyor from the surface of the moon.

Such rapid cooling of the surface when the sunlight is withdrawn is caused partly by the absence of an atmospheric blanket. It is also promoted by the low heat conductivity of the surface material, so that the heat does not penetrate very far into the interior. In this respect the material has been likened to pumice or volcanic ash. Records of radiations from the moon in the radio wavelengths suggest that at only a few meters below the surface the temperature remains constant at about −40°C.

The *albedo*, or reflecting power, of the moon is only 7 per cent, in contrast with the value of 40 per cent for the earth; this refers to the ratio of the light reflected by the whole illuminated hemisphere of the moon to the light it receives from the sun. There are marked local variations from the average, from the darkest parts of the seas to the very bright floor of the crater Aristarchus. The reflectivity varies dramatically with the angle between the earth-moon-sun system. This is known as the moon's *photometric function* and rises to a very high peak at full moon (Fig. 5.17C). With allowance for shadows, the moon's reflecting power is comparable with that of rather dark brown rock—brown because the moonlight is redder than the direct sunlight.



Lunar cooling curve. Adapted from Surveyor III data.

Figure 5.17A
The full moon, printed in the position it appears when seen with the naked eye, following the astronautical convention. *(Yerkes Observatory photograph).*

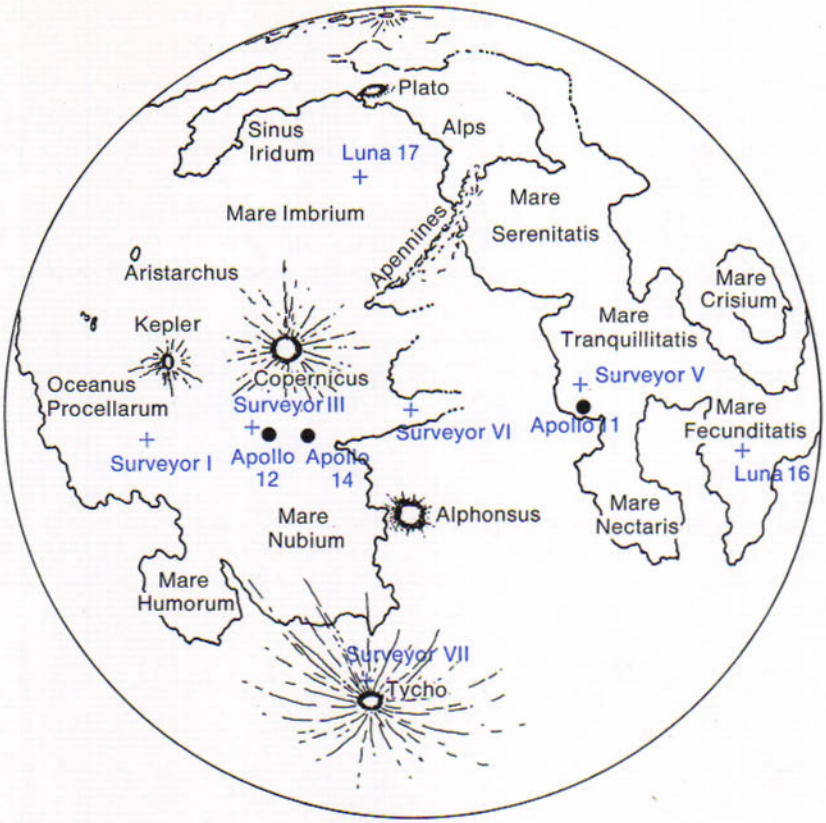


Figure 5.17B
An abbreviated identification chart of features in Fig. 5.17A.

It is not entirely bare, unbroken rock that we see. Although the rocks on the moon are not exposed to ordinary weathering by wind and water, their surfaces have been exfoliated by repeated expansion and contraction with the great range in the temperatures. The accumulation of meteorites and fragments of porous rocks shattered by the fall of meteorites add to the rubble. If we assume that the moon is uniform throughout, this rock has a density of 3.3 which is comparable to the basaltic lavas. Indeed, the various soil samplers carried on the Surveyor series spacecraft yielded a composition of the mare surface in excellent agreement with a basaltic composition. The Apollo 11 and 12 samples yield far more detail but are in essential agreement with this conclusion.

The first Apollo sample analysis is too new at this revision to be presented in detail. A few facts seem to be firm and will probably stand

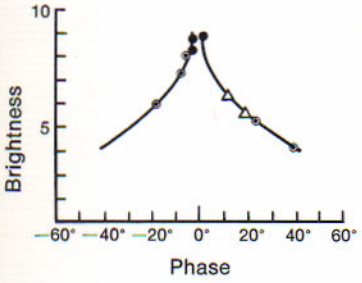
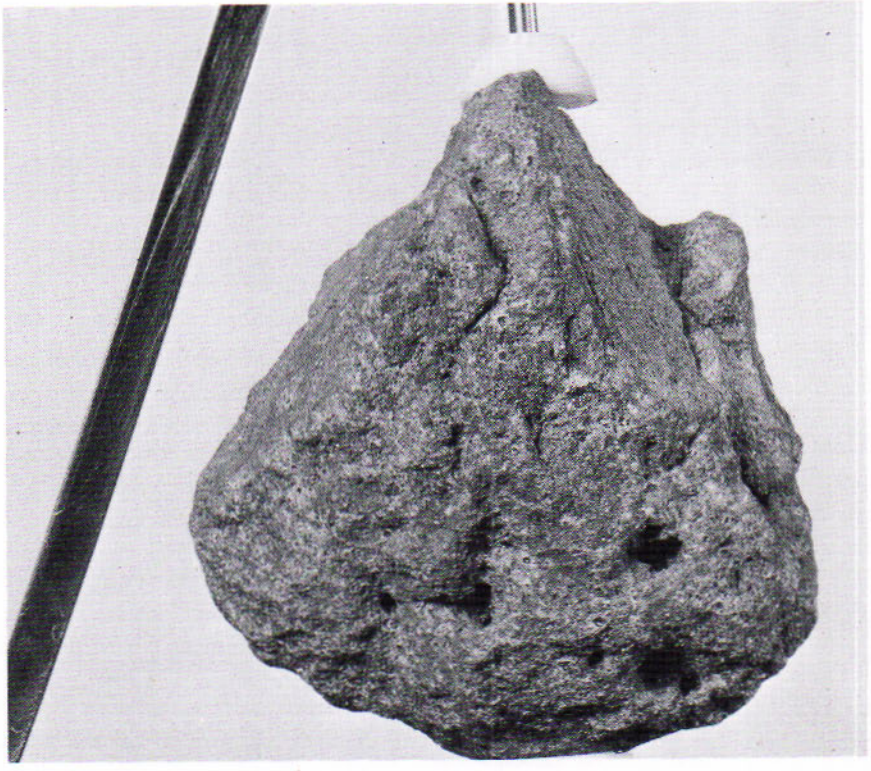


Figure 5.17C
Lunar photometric function in Tycho.

up. The ages of the rocks and grits collected run from 2.5 to 3.5 billion years. A very fine grit of somewhat different composition has an age of more than 4.5 billion years. This fine grit may originate from impacts in the lunar highlands or may be ray material from deep craters. It contains a high degree of aluminum compounds giving it a high reflectivity such as we see in the highlands.

The rocks examined have been irradiated by space for periods of one to five hundred million years without being disturbed. Some of the rocks are essentially round and have been on the surface and turned over often enough that radiation effects and micro-impact pits are distributed isotropically on them. The periods between reorientations are on the order of one hundred million years. There are enough differences between the first two samples to conclude that the moon is heterogeneous and has had a complex geological history.

Figure 5.17D
Lunar stone collected by Astronaut Conrad during the Apollo 12 mission. It is designated Sample 12006,1 and is of igneous origin. The stone measures 7.2 centimeters between the jaws of the display pedestal. *(NASA photograph).*



The seismic experiments so far seem to indicate that there is very little activity within the range and sensitivity of the seismometers. Surely there are seismic events such as slides, for example, for there is visual evidence of these in pictures, but they must be rather rare and are probably induced by the solar heating and cooling cycle. What seismic activity there is correlates well with tidal effects of the earth on the moon at perigee. The near surface of the moon appears to be extremely rigid since the impact of the lunar rocket ascent stage after Apollo 12 left the surface caused the moon to "ring" for about fifty minutes. This interpretation seems reasonable since we must remember the fact that such a rigid surface is able to support the highlands and mascons (5.18) despite its relatively low density.

5.18 Selenography

The mapping of the details of the lunar surface dates from 1610, when Galileo made the first map of the moon as observed with the telescope. He had recognized the lunar mountains and had called the large dark areas "seas," a misnomer that persists in the present nomenclature.

J. Hevelius, at Danzig in 1647, published a lunar map showing 250 named formations; the names were after terrestrial features they may have seemed to resemble. All that survive from this plan are the names of 10 mountain ranges, including the Apennines and Alps. J. B. Riccioli, at Bologna in 1651, chose the more enduring plan of naming the lunar craters and some other features in his map after distinguished former scholars; he selected names such as Copernicus, Tycho, and Plato.

Many lunar maps have since been made, often including more detail than was contained in previous ones. Especially noteworthy in the 19th century were the maps by Beer and Mädler, at Berlin in 1837, and by J. Schmidt, at Athens in 1878. The most detailed recent map, showing at least 90,000 formations, was constructed by H. P. Wilkins, former director of the Lunar Section of the British Astronomical Association. Some of the later map makers have extended Riccioli's system to smaller features and have assigned them names of contemporary observers.

Selenography reached a high point of activity during the 1960's as the lunar exploration effort was intensified. The U. S. Geological Survey began its effort on a part-time basis using the McCormick refractor in Charlottesville, Virginia. It later centered its effort in Flagstaff, Arizona where the Aeronautical Chart and Information Center was doing its mapping at the Lowell Observatory. This latter organization produced excellent maps which the Geological Survey and other groups all over the world used for their interpretive work.

Extensive Orbiter series spacecraft photographs were used to complete the mapping efforts. Along with the photographs, the moon's magnetic and gravitational fields were mapped through the use of orbit-tracking

data. The magnetic field of the moon was known to be essentially zero, so nothing surprising was turned up, but a quite different story developed in the gravitational mapping.

The physical shape of the moon has long been known to resemble that of an egg, with the elongated end pointing generally towards the earth. The gravity field for such a body can be predicted and its effect upon an orbiting test probe (the spacecraft) can be given in the form of complex power series. Incremental differences from the predictions leads to the conclusion that there are many local mass concentrations, called *mascons,* relatively near the surface of the moon in several of the maria. An earth analogy would be the Mesabi Range in Minnesota. The mascons are apparently most numerous on the "earth" side of the moon.

5.19 Lunar Features

The moon, even to the naked eye, presents detail in the nature of dark and bright areas. The large dark areas were designated as *maria* (plural of Latin *mare,* or sea) before we knew their true nature and retain these designations for historical reasons. *Mare Tranquilitatis* (Sea of Tranquility) is an excellent example of a relatively smooth lunar *mare.* Similar small areas off a larger mare often received the designation of gulfs or bays of which Sinus Iridum (Bay of Rainbows) is an example. Ranger, Surveyor and Zond spacecraft have shown the *maria* to be boulder strewn plains with a coarse sand like structure. In fact the first team of astronauts to circumnavigate the moon described the surface by the term "dirty sand". The *maria* are pockmarked with small craterlets formed from secondary impacts and, presumably, slumps. All of the *maria* studied by the Surveyor spacecraft were essentially identical. The far side of the moon is remarkably free of large *maria* in sharp contrast with the near side.

Under oblique lighting the most striking features of the moon are the craters. These range in size from great craters such as Copernicus (60km across) down to as small as 30cm. Craters have a variety of general features, some craters have bright ray structures, some craters have raised rims, some have central peaks. Some, such as Aristarchus, have rough bright saucer shaped floors, while others have smooth dark floors resembling *maria,* as in the case of Plato. Some astronomers claim that the great walled plains such as Clavius (200km across) are actually craters. The far side of the moon is heavily cratered. Its most striking feature is *Mare Orientalis,* actually a series of ringed craters the outermost of which is fully 900km across. Enough of this feature was revealed at extreme librations to enable G. P. Kuiper and his co-workers to give its true form in 1959. Mare Nubium resembles *Mare Orientalis* but must be much older.

The rays radiating from certain craters were shown by the Ranger VII Spacecraft to be *ejecta blankets* and secondary craters caused by impacts of debris from the primary crater. The most striking ray pattern is that

Figure 5.19
The moon from lunar orbit. This excellent photograph by C. Gordon shows wrinkle ridges, chain craters and, just in front of Crater Enke (the pentagonal crater in the center), a very sharp rille. The large crater at the top on the terminator is Kepler. *(NASA photograph).*

radiating from Tycho. The most unusual craters are the rimless, almost circular ones that often occur in chains. These may be the result of slumping into a lava tube.

Rilles, or clefts which are sometimes 1km across and hundreds of kilometers long in some cases, are interesting features. They are of two classes, relatively straight almost linear and very twisted following a tortuous path. The origin or cause of the rilles is not clear. They may be related to graben or possibly slumps into lava tubes. Occasional faults such as the Straight Wall resulting from a 300m vertical shift in the plain floor are found. It would be hard not to associate these with earthquake like activity on the moon. Faults are of great geologic interest since they expose ancient interior rock to investigation and have little of the errosive effects suffered by most of the lunar surface.

A very common feature of the lunar surface is the existence of *domes.* These mound structures range from very small to several kilometers in diameter and height. They very much resemble pingos and have raised the tantalizing possibility that there may be sub-surface ice on the moon. Certain irregular lunar craters resemble open pingos such as are found on the Canadian shield.

Among the few formations that have any resemblance to terrestrial mountain ranges are the three that form the western border of Mare Imbrium; they are the Apennines, Caucasus, and Alps. Like others that border the seas, these mountains slope more abruptly on the seaward side and more gradually in the opposite direction. They are surmounted by many peaks, the highest ones rising nearly 6100m above the plain.

Still greater heights are measured in the Leibnitz and Doerfel mountains near the south pole and almost beyond the edge of the moon; some of those peaks have elevations of 8000m, almost as high as Mount Everest. Heights on the moon, however, are less easily compared because they are referred to the neighboring plains; and the plains themselves are at different levels.

The height of a lunar mountain is determined in one way by measuring the length of its shadow and by calculating the sun's altitude above the horizon as seen then from that point on the moon. The height may also be found by measuring the distance of the summit from the terminator as it catches the first rays of the rising sun or the last rays of the setting sun. At those instants the illuminated top of the peak looks like a little star out in the dark beyond the terminator. A sketch of either situation will show that enough data are then known to calculate the height of the mountain by solving a right triangle.

A very interesting feature is the *wrinkle ridge.* It is a raised irregular hummocky ridge with very gentle slopes. The heights are on the order of 300 meters and they may have a width up to 30 kilometers. An analogy with the eroded Appalachian ridges is commonly drawn, but their origin and history must be quite different.

5.20 Lunar Nomenclature

Names for features on the moon developed in a rather disorderly fashion. Most of the earliest names are retained in the current lunar nomenclature for sentimental reasons. In 1932 the International Astronomical Union adopted an extensive list of about 5000 designated lunar features. Since that time a series of rules have been developed that have been more or less adhered to.

The rules are rather general. All names should be latinized. Craters, walled plains and rings should be named after deceased astronomers and prominent scientists. Mountain-like chains should be named after terrestrial features. Large dark areas should receive names indicating psychic states of mind. Rifts and valleys should be named after the nearest designated crater. Less outstanding features should be designated by their coordinates.

These rules were brought into sharp focus with the first photography of the moon's far side. From the first composite picture of the far side 18 names were proposed and adopted by the I.A.U. Later some 230 names were proposed, but not acted upon because excellent photographs were

then available which contradicted some of the earlier named features. In fact, of the original 18 farside features 11 names had to be withdrawn including Montes Sovietici (the Mountains of the Soviets) which appeared very prominently in the Lunik III pictures.

In order to avoid similar embarrassment from over-interpretation in the future, the I.A.U. in 1967 decided that the names applied to features on the farside should follow the earthside naming scheme and have about the same density of named features. Features on the excellent Aeronautical Chart and Information Center charts were numbered and an international committee was assigned the task of assigning names.

5.21 Origin of the Surface Features

The two predominant theories called upon to explain the surface features are 1) impacts of meteors and 2) volcanism. Both theories have always been accepted, but one or the other has predominated at various times depending upon the current evidence and the astronomer.

During the recent "pre-space age era" the meteoritic theory generally held sway. With the intensification of lunar observing due to man's exploration of space, gaseous emissions referred to as "red spots" have been observed giving some support to the volcanic concept and a moon that is not completely "dead". Actually, gaseous emissions have been observed for centuries but have not generally been accepted. The detection of "hot spots" by infra-red scans across the surface of the moon and independent observation of the same red spots have revived the concept of an active moon. Students of the moon, such as G. P. Kuiper, are even studying active earth volcanoes for clues as to the shape, size and structure that one should expect of lunar craters.

Many craters are the result of an impacting body. An example of such a crater would be Tycho with its rays of debris stretching out thousands of kilometers. Even more spectacular is the magnificent concentric ringed system of Mare Orientalis. Perhaps the strongest support for the impact theory can be derived from the very heavy cratering on the far side of the moon. This can be explained analytically if we assume that most large meteors have orbits close to the plane of the ecliptic. The earth would then tend to shield the near side of the moon although the ratio of the relative exposures is only slightly different from unity. Nonetheless, this slight difference operating over eons becomes effective. During the past decade evidence of very large impact craters on the earth has increased substantially. These craters are very old and have suffered erosion from wind and rain to the point that they are almost obliterated. They seem not to be preferentially distributed which would be in agreement with theory since the rotation of the earth exposes all sides equally.

Orbiter photographs clearly show lava flows overlapping other features.

This is convincing evidence of volcanic activity on the moon. Some craters, for example Ritter and perhaps Aristarchus, appear to be dead *caldera* in appearance. Many rilles appear to be collapsed lava tubes while others are clearly faults and are thus at least seismic in origin. The lava flows, rilles and faults could result from activity triggered by an impact. Certainly the impact that caused Tycho would have seismic effects felt everywhere on the moon.

THE TIDES

The rise and fall of the level of the ocean twice at any place in a little more than a day has been associated with the moon from early times. Newton correctly ascribed the tides in the ocean to the attractions of both the moon and sun and accounted for their general behavior by means of his law of gravitation.

5.22 Lunar Tides

To simplify the explanation of the tides we may imagine, as Newton did, that the whole earth is covered by very deep water. Because the gravitational attraction between two bodies diminishes as their separation is greater, the moon's attraction is greatest for the part of the ocean directly under the moon and is least for the part on the opposite side of the earth. The ocean is accordingly drawn into an ellipsoid of revolution, which in the absence of other effects would have its major axis directed toward the moon. This axis rotates eastward, following the moon in its monthly course around the earth.

Meanwhile, the earth is rotating eastward under the tide figure. The earth makes a complete rotation relative to a particular point in that figure once in a lunar day, which averages about $24^h\ 50^m$ of solar time. *High tide* occurs at a place of observation at intervals of $12^h\ 25^m$, and *low tide* at times halfway between them. These are occasions when the ocean level at the place is the highest and lowest, respectively, for that particular cycle.

Thus a succession of tide crests move westward around the earth. They are displaced behind the moon in its diurnal motion by friction with the ocean floor, because the water has not the depth required by the simple

static theory, and their progress is interrupted by land masses. High tide and the transit of the moon are generally far from simultaneous. The difference in time between these occurrences varies from place to place and is best determined by observation.

5.23 Spring and Neap Tides

The sun also causes tides in the ocean. It can be shown that the tide-producing force of a body varies inversely as the cube of the distance of that body and, accordingly, that the sun, despite its far greater mass, is less than half as effective as the moon in raising tides on the earth.

The two sets of tides may be considered as operating independently, the relative positions of their crests varying with the moon's phases. The *spring tide* occurs when the moon is new or full. Because the moon and sun are then attracting from the same or opposite directions, lunar and solar tides reinforce each other. The *neap tide* occurs when the moon is at either quarter phase. Then the moon and sun are 90° apart in the sky, so that one set of tides is partly neutralized by the other. When the moon is new or full and also in perigee, the difference in level between low and high tides is especially great.

The earth itself, like the ocean, is deformed by lunar and solar tides, but to a much smaller extent. Consequently, the observed tides in the ocean represent the differences between ocean and earth tides and to a lesser extent other tide-raising forces. We have already seen that the rotating earth causes an equatorial bulge. The rotation axis of this bulge is not perpendicular to the line joining the two bodies which complicates the moon's tidal action. There is another not inconsequential force which arises from the earth's rotating around the center of mass of the earth-moon system. This force tends to reinforce the lunar tides because it acts along the line joining the two bodies with a period of 27⅓ days and has a greater effect on the side opposite the moon.

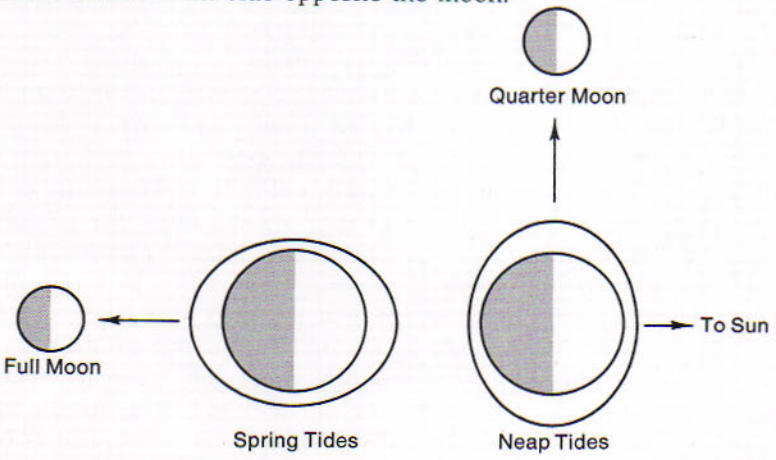


Figure 5.23
Spring tides and neap tides. Spring tides occur at new and full moon, when lunar and solar tides reinforce each other. Neap tides occur at the quarter phases when one set of tides is partly neutralized by the other.

5.24 Tidal Friction

The tides in the oceans and in the earth itself act as a brake on the earth's rotation; they tend to be held in position by the moon and sun and to impose some restraint on the daily rotation of the earth. The tides would accordingly be expected to reduce the speed of the rotation and to gradually lengthen the day (2.18). An increase in the period of the earth's rotation at the rate of $0^s.0016$ a century has been derived by comparing the observed times of early eclipses with the times when they would have occurred if the earth's rotation had remained perfectly uniform.

Although the rate of increase might seem negligibly small, the difference in the times has accumulated to a considerable amount within the period in which eclipses have been recorded. The error of a mean solar clock compared with a clock having constant rate is $\frac{1}{2}at^2$ seconds, where in this case $\frac{1}{2}a = 0^s.0008 \times 365\frac{1}{4} \times 100$, or about 29 seconds, and t is the number of centuries intervening. In 20 centuries the error in the earth-clock had amounted to 11,600 seconds, or about $3\frac{1}{4}$ hours. Tidal friction in the shallow seas has been assigned as sufficient reason for this effect. However, the actual rate of increase in the period of the earth's rotation and its cause as well may require further study.

The moon turns one hemisphere toward the earth, and some other satellites do the same with respect to their primaries. Such effects have been ascribed to tidal friction within the bodies themselves, which would have been unassisted by ocean tides. These bodies have no oceans and probably never had any.

5.25 Tidal Theory and the Origin of the Moon

There are four theories extant explaining the origin of the moon. The first is that the moon is a fragment of the earth, the second is that the moon was formed from the same nebula that formed the solar system, the third is that the moon is a captured body, and a fourth evokes a mechanism of accretion to build up the moon. For any one of these to be valid it must be compatible with the theory of tides.

Because of *tidal friction* the earth's period of rotation *increases.* Since angular momentum must be conserved, the moon moves away from the earth. At a distance of about two earth radii reinforcing tides occur that serve to break up any large massive body so we can only discuss the tidal effects beyond this point. At 2.5 earth radii the length of the month was then something like a quarter of its present value, and the day was a still smaller fraction of the present day. Under the action of the tides both month and day slowly increased in length, the month at first faster than the day. Eventually the day will lengthen at a faster rate than the month, until the two become equal to 47 of our present days.

At that remote time in the future, when the moon is much farther away than it is now, the earth-moon system will be internally stable; the earth will turn one hemisphere always toward the moon, just as the moon

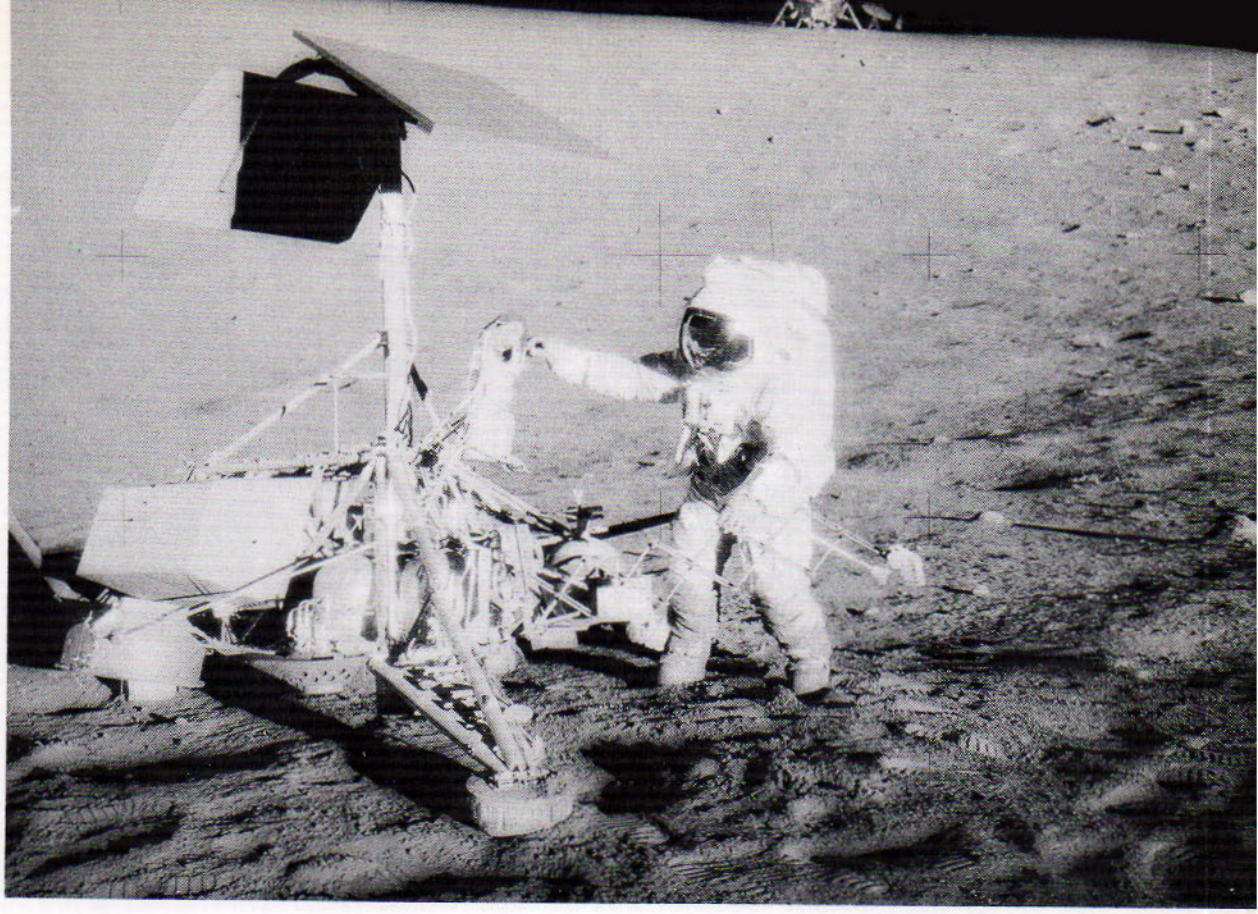


Figure 5.25
Surveyor III in Lunar Crater. C. Conrad is shown detaching parts of the television camera. Note the half covered rocks and completely uncovered rocks on the slopes of the crater. The Spacecraft Intrepid is in the background. *(NASA photograph)*

now presents one hemisphere to the earth. If it happens not to be our hemisphere that is turned moonward, the moon may become one of the sights to see on a trip abroad. At that stage the lunar tides cannot alter the system, but the solar tides will still operate on it and will force the earth and moon out of step again. The history of the system will then be repeated in reverse order, according to the theory, until the moon is brought back close to the earth, perhaps to be destroyed into a ring of debris.

Several observed features of the moon lead to the suspicion that the moon is composed of crystal rock materials such as basalt. This is deduced from the moon's photometric function, the dielectric constant at the surface, and the mean density. The theory is that the moon tore away from the earth from what is now the Pacific Ocean. Mechanically this theory suffers from a grievous difficulty because it has the moon moving through the tidally destructive zone.

The *nebular hypothesis,* discussed more fully in section 9.29, is more appealing in that the moon is formed simultaneously with the earth and from the same material thus avoiding the objections to the fission theory. The capture of a minor planet has certain appeal as well, but it requires such exact conditions to explain, first: the capture itself, second the circular orbit, and third, the fact that the capture process would deposit four times the energy required to melt the moon into the moon, that it has not been given too much serious thought in the last few decades. The *acretion theory* is beginning to receive serious support in that it overcomes most objections and accounts for some of the problems of conserving angular momentum. In this theory small particles "stick" together and become larger. The larger particle then sweeps up more particles which stick to it, etc., until all of the particles are swept up.

Exploration of the moon will certainly answer the questions concerning the moon's origin and thus shed some light upon the origin of the earth and the more general problem of the origin of the solar system. The moon, long considered a dead body, may not be so. Observations of transient phenomena have increased. These are usually in the form of pink or red spots resembling escaping gases. A simple seismometer on the surface of the moon may answer this question.

The exploration of the moon holds more than passive interest for astronomers. The moon may prove to be an excellent base for observing all the way from γ radiation to the infrared. The lack of an atmosphere has obvious observational advantages. The location of a radio observatory on the far side of the moon would give radio astronomy reception free of earth noise contamination and allow observing at very low frequencies.

REVIEW QUESTIONS

1. Give the four theories used to explain the origin of the moon. Why is one of these rejected? What can help select the correct one of the other three?
2. What are the points when the moon is nearest and farthest from the earth called?
3. What is meant by the aspects of the moon? List three aspects and indicate them on a simple diagram.
4. How could the ancient astronomer Aristarchus give the distance to the sun by observing the moon's phases? What units would he have used?
5. What is the difference between the sidereal and synodic months?
6. What are the principle lunar librations and their causes?
7. Why do we feel that the moon's surface is primarily composed of basaltic lava?
8. Explain the observational and theoretical reasons for our assertion that the moon has no atmosphere.

9. The velocity of escape from the surface of the moon is essectially a constant for a probe of small mass. If a gas molecule of mass m at temperature T can just escape the moon, what must the temperature be to allow a gas molecule of mass $2m$ escape?
10. Show that the velocity of escape for a spacecraft on the moon is approximately 0.05 that of the same spacecraft on the earth. Does this suggest a further use of the moon not discussed in the text?
11. Suppose that your lunar map did not contain the names of rilles. How would you find Hadley Rille?
12. Show with the aid of a diagram why the moon raises tides on opposite sides of the earth.

REFERENCES

Blagg, Mary A. and K. Muller, *Named Lunar Formations,* London: Lund, Humphries & Co. 1935

Baldwin, Ralph B., *The Measure of the Moon,* Chicago: Univ. of Chicago Press, 1963

Kuiper, Gerard P., *Photographic Lunar Atlas,* Chicago: Univ. of Chicago Press, 1960

Whipple, Fred L., *Earth, Moon and Planets,* Cambridge, Mass.: Harvard University Press, 1963

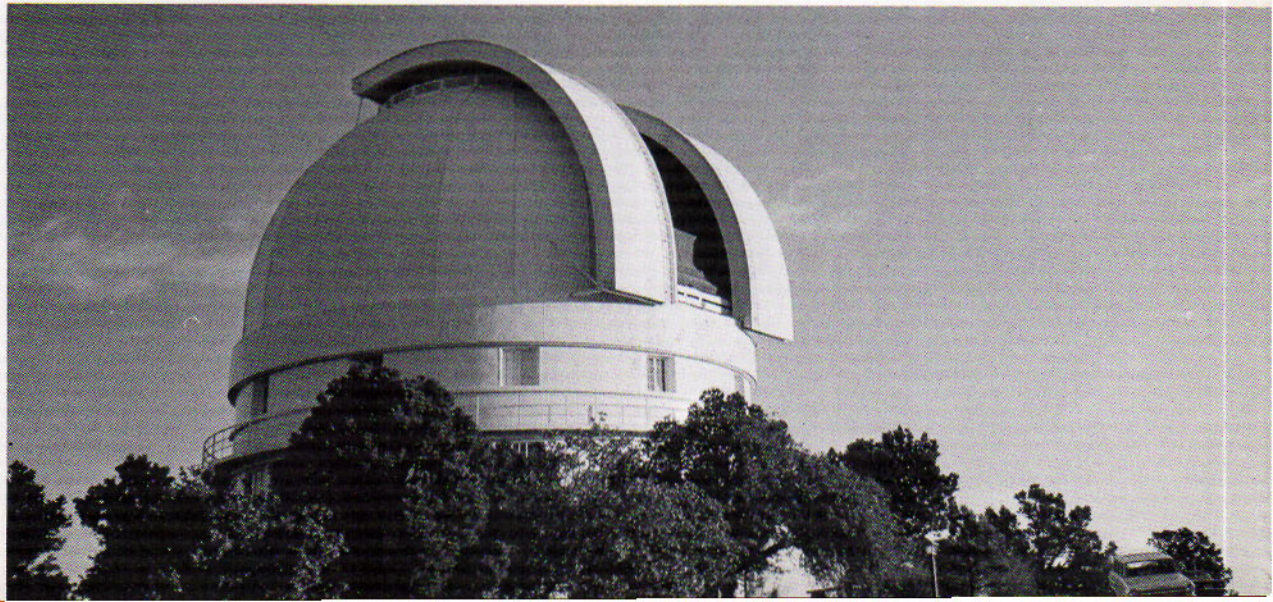
The U. S. Air Force LAC Charts, compiled by the Aeronautical Chart and Information Center are available from the U. S. Printing Office at a nominal cost per chart.

FOR FURTHER STUDY

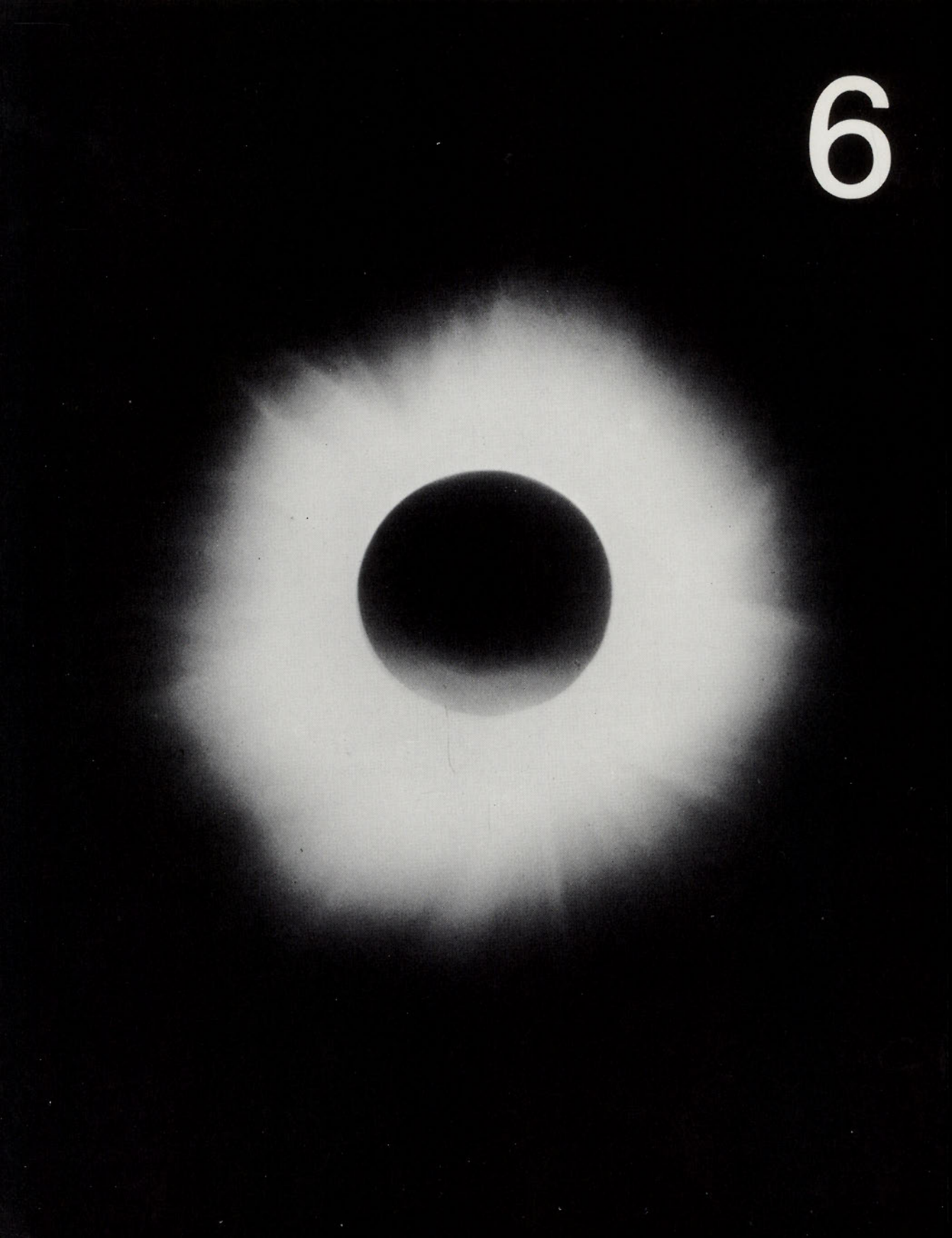
Hess, Wilmot N., Donald H. Menzel and John A. O'Keefe, eds., *The Nature of the Lunar Surface,* Baltimore: Johns Hopkins Press, 1966

Kopal, Zdenek, editor, *Physics and Astronomy of the Moon,* New York: Academic Press, 1962.

Kuiper, Gerard P., ed., *The Atmospheres of the Earth and Planets,* Chicago: University of Chicago Press, 1949



Dome of the McDonald Observatory, University of Texas. *(McDonald Observatory photograph)*


6

ECLIPSES OF THE MOON AND SUN

Eclipses of the moon occur when the moon, at the full phase; passes through the earth's shadow and is thereby darkened. Eclipses of the sun occur when the moon, then at the new phase, passes between the sun and the earth, so that its shadow falls on the earth; the observer within the shadow sees the sun wholly or partially hidden by the moon.

6.1 Shadows of the Earth and Moon

Because the earth and the moon are globes smaller than the sun, the shadow of each one is a cone having its apex directed away from the sun. This region, from which the sunlight is geometrically entirely excluded, is the *umbra* of the shadow and is sometimes called simply the *shadow*. It is surrounded by the larger inverted cone of the *penumbra*, from which the sunlight is partially excluded. There is no way of observing the shadows except as they encounter objects that shine by reflected sunlight.

The average length of the earth's shadow is 1,382,400 km as may be easily calculated from two similar triangles having as their bases the

Total solar eclipse, May 20, 1947, Bocainva, Brazil. *(Georgetown University photograph)*

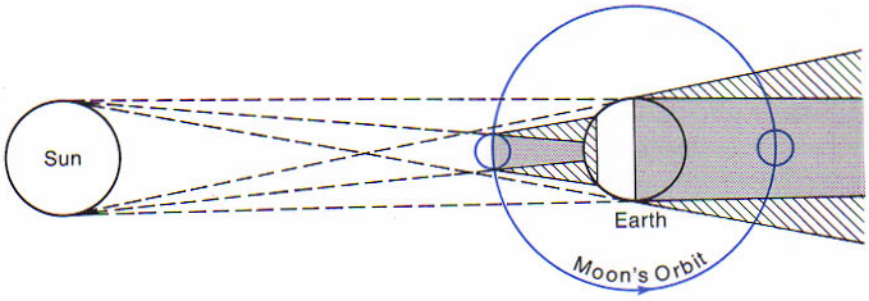


Figure 6.1
Cause of lunar and solar eclipses. When the moon is opposite the sun's position in the sky, it is eclipsed by the earth's shadow. When the moon is between the sun and the earth, its shadow falls on a portion of the earth; within this region the sun is eclipsed.

diameters of the earth and sun. By a similar procedure it is found that the length of the moon's shadow averages 373,370 km when the moon is between the sun and the earth. Because these shadows are more than 100 times as long as their greatest widths, they cannot conveniently be represented in diagrams in their proper proportions.

6.2 The Moon in the Earth's Shadow

If a screen could be placed opposite the sun's direction at the moon's distance from us, the umbra of the earth's shadow falling normally upon the screen would appear as a dark circle about 9170 km in diameter. Always opposite the sun, this shadow moves eastward around the ecliptic once in a year. At intervals of a synodic month the faster-moving moon overtakes the shadow and, whenever it then encounters the shadow, enters at the west side and moves through at a rate that is the difference between the speeds of the moon and the shadow; the hourly rate is about 30′, or very nearly the moon's apparent diameter.

Umbral eclipses of the moon are total and partial. The longest eclipses occur when the moon passes centrally through the shadow; the duration of the whole eclipse in the umbra is then about $3^h\ 40^m$, and of the total phase is $1^h\ 40^m$. Noncentral eclipses are shorter, depending on how nearly the moon's path approaches the center of the shadow. When the least distance from the center exceeds the difference between the radii of the shadow and moon, there is no total phase.

A lunar eclipse is visible, weather permitting, wherever the moon is then above the horizon, that is, from half of the earth and also the part that is rotated into view of the moon while the eclipse is in progress. The times when the moon enters and leaves the penumbra and umbra and of the beginning and end of totality are published in advance in various almanacs.

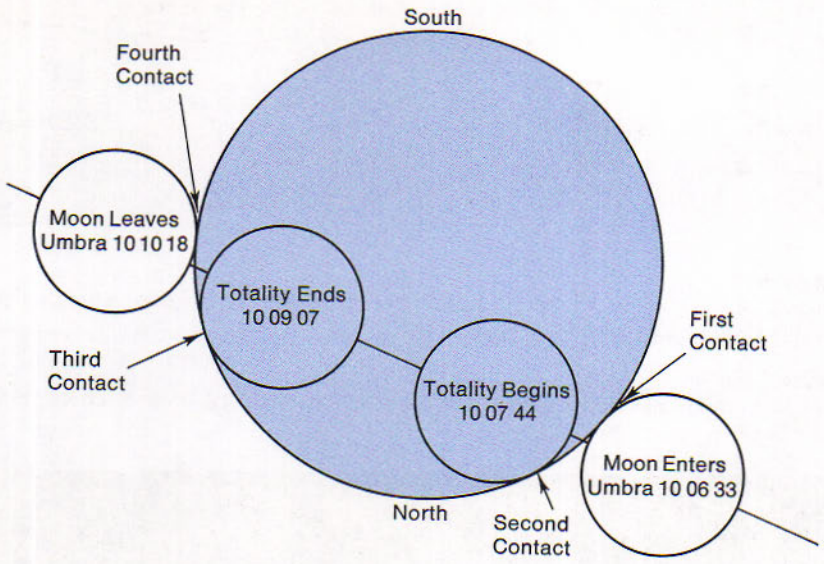


Figure 6.2
Path of the moon through the umbra of the earth's shadow during the eclipse of February 10, 1971. Universal times are given.

In the 6-year interval from 1972 to 1978 inclusive there are 11 umbral lunar eclipses scheduled, which are visible from a considerable area of the United States and Canada. Six of these are total.

6.3 Eclipses of the Moon

The moon is darkened so gradually in its passage through the penumbra of the shadow that this phase of the eclipse is less conspicuous. Soon after the moon enters the umbra, a dark notch appears at the eastern edge and slowly overspreads the disk. So dark in contrast is the shadow that the moon might be expected to vanish in total eclipse. As totality comes on, however, the entire moon is usually plainly visible.

Even when it is totally eclipsed, the moon is still illuminated by sunlight. The light filters through the earth's atmosphere around the base of the shadow and is refracted and diffused into the shadow and onto the moon. Red predominates in this light for the same reason that the setting sun is red. On rare occasions there is so much cloudiness around the base of the shadow that the eclipsed moon is very dim.

Penumbral eclipses occur when the moon passes through the penumbra of the earth's shadow without entering the umbra. The darkening of the part of the moon in the penumbra is visible to the eye when the least distance of the edge of the moon from the umbra does not exceed 0.35 of the moon's diameter. The darkening is detected in the photographs when the least distance does not exceed 0.65 of the moon's diameter, and by photometric means when the distance is still greater.

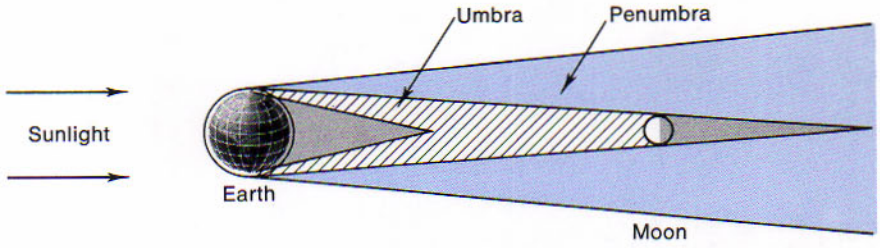


Figure 6.3A
Visibility of the moon in total eclipse. Sunlight is diffused by the earth's atmosphere into the shadow and onto the eclipsed moon.

Figure 6.3B
The moon in the penumbra of the earth's shadow, during eclipse of March 23, 1951. *(Griffith Observatory photograph).*

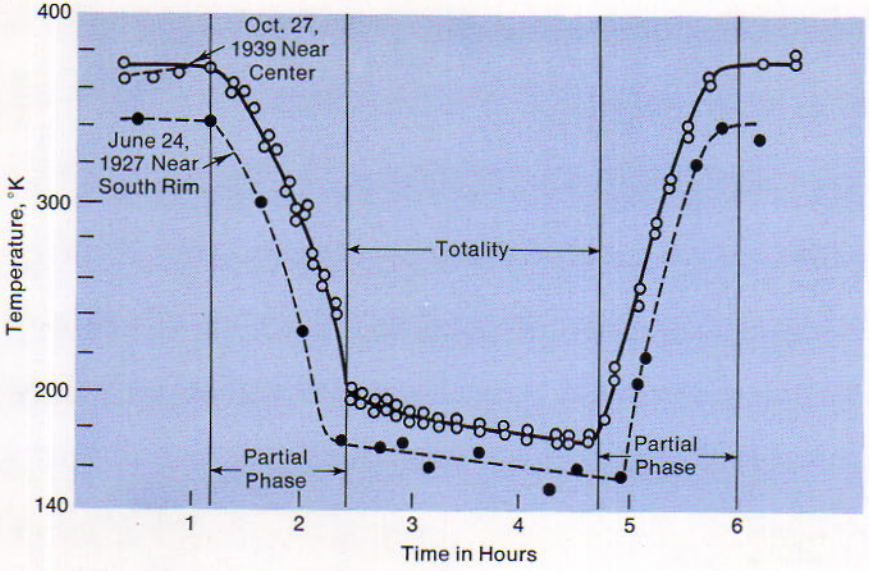


Figure 6.3C
Lunar eclipses' cooling curves.

Lunar eclipses, at times very spectacular, offer very little of a scientific nature. *Cooling curves* obtained at different wavelengths as the earth's shadow sweeps across the moon, offer clues to the nature of the surface material. Such observations led to the discovery of "hot spots" which are either areas of high heat capacity or areas where a warm interior material is fairly near to the surface such as would be the case in volcanic areas on the earth. Some of the "hot spots" coincide with reported areas of transient activity on the moon (5.21). Typical cooling curves are given in Figure 6.3C. Note that the cooling curve shown in Section 5.17 is essentially an extension of these eclipse cooling curves.

6.4 The Moon's Shadow on the Earth

In average conditions the umbra of the moon's shadow fails to reach the earth. The average length of this part of the shadow is 373,370 km, which is almost 4800 km less than the mean distance of the moon's center from the nearest point of the earth's surface. The fact that the umbra occasionally extends to the earth at solar eclipse results from the eccentricity of the earth's orbit around the sun and of the moon's orbit around the earth. At aphelion the length of the umbra is increased to 379,800 km and at perigee the moon's center may be as close as 350,000 km to the earth's surface. In these extreme conditions the umbra may fall on the earth 29,800 km inside the umbra's apex.

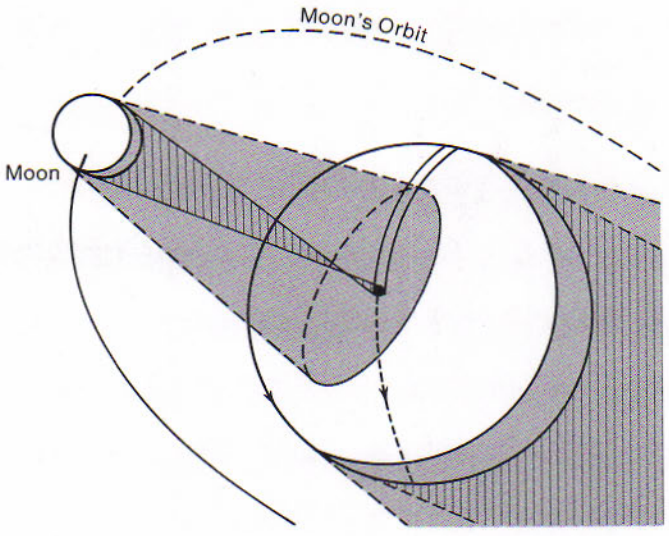
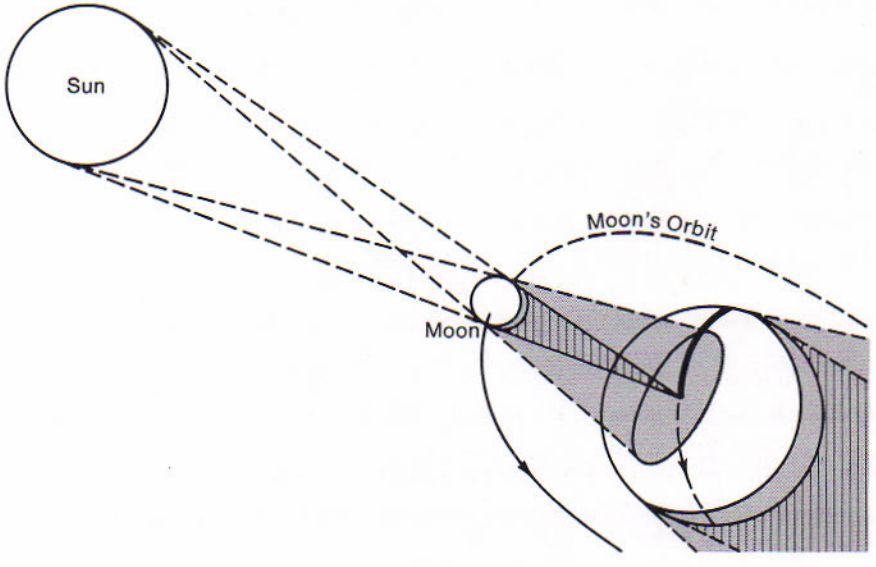
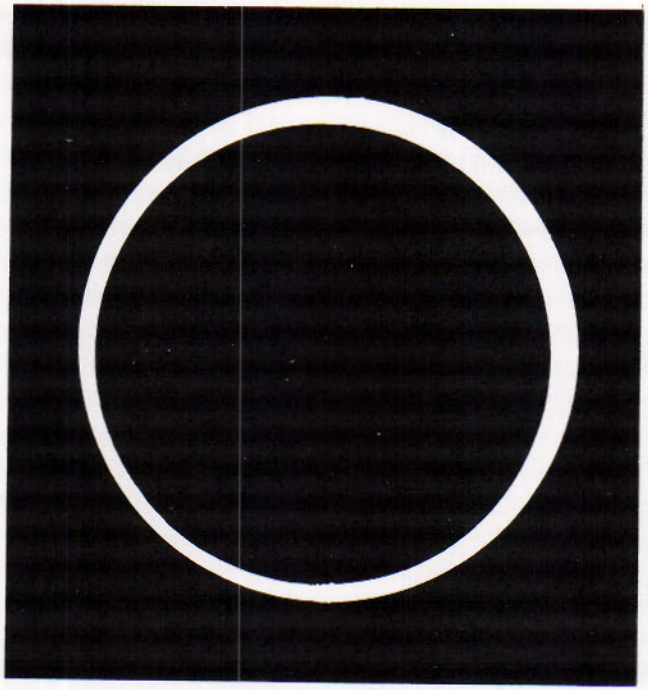


Figure 6.4A
Annular eclipse of the sun. The umbra of the moon's shadow does not reach the earth's surface. Within this shadow geometrically produced, a thin ring of the sun's disk remains visible around the moon.

Figure 6.4B
Path of total eclipse. The moon's revolution causes the shadow to move in an easterly direction over the earth's surface. From within the umbra the eclipse is total. Elsewhere within the larger circle of the penumbra the eclipse is partial.





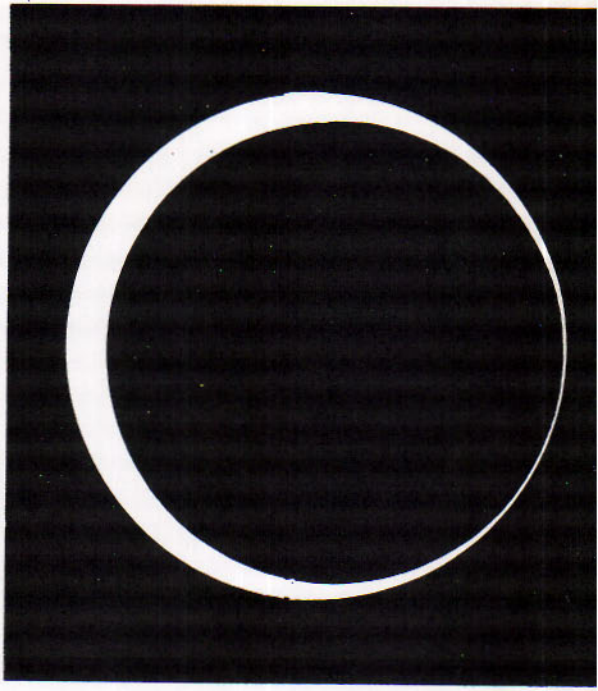


Figure 6.5A
The progression of the 1955 annular eclipse as seen from Mergui, Burma. *(Georgetown University photograph)*

6.5 Total and Annular Solar Eclipses

A *total eclipse* of the sun occurs when the umbra of the moon's shadow falls on the earth. If the observer is then within the umbra, he sees the dark circle of the moon completely hiding the sun's disk. The umbra can never exceed 427 km in diameter when the sun is overhead.

An *annular eclipse* occurs when the moon is directly between us and the sun, but the umbra of the shadow does not reach the earth. If the observer is within the circle of the umbra produced beyond its apex (Fig. 6.5A), he sees the moon's disk projected against the sun; the dark disk then appears slightly the smaller of the two, so that a bright ring, or annulus, of the sun remains uneclipsed. Annular eclipses are 20 per cent more frequent than total eclipses.

Around the small area of the earth in which the eclipse appears total or annular at a particular time, there is the larger partly shaded region of the penumbra, from 3200 to 4800 km in radius. Here the eclipse is *partial;* the moon hides only a fraction of the sun's disk, the fraction diminishing with increasing distance from the center of the shadow. When the axis of the shadow is directed slightly to one side of the earth, only the partial eclipse can be seen.

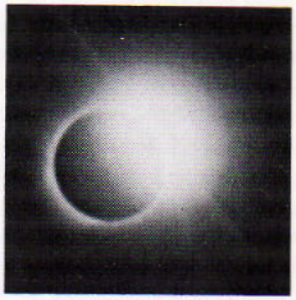


Figure 6.5B
The so-called "diamond-ring effect", 1963. *(Leander McCormick Observatory photograph)*

6.6 Path of the Moon's Shadow

As the moon revolves around us, its shadow moves generally eastward by the earth at the rate of about 3380 km an hour. Because the earth's rotation at the equator is at the rate of 1674 km an hour, also eastward, the effective speed of the shadow at the equator, when the sun is overhead, is 1706 km an hour. In other parts of the earth where the speed of rotation is less, the effective speed of the shadow is greater. The speed may reach 8000 km an hour when the sun is near the horizon. Considering the high speed of the shadow and its small size, it is evident that a total eclipse of the sun cannot last long in any one place; the maximum possible duration scarcely exceeds $7^m\ 30^s$ but can be extended by use of high flying aircraft. An annular eclipse may last a little longer. The partial phase accompanying either type of eclipse may have a duration of more than 4 hours from beginning to end.

The *path* of total eclipse, or of annular eclipse, is the narrow track of the central part of the shadow as it sweeps generally eastward over the earth's surface, from the time it first touches the earth at sunrise until it departs at sunset. Meanwhile, the penumbra moves over the larger surrounding region in which the eclipse is partial.

TABLE 6.I
Lunar Eclipses*
(Ephemeris Time)

DATE	TYPE	MOON ENTERS UMBRA	TOTAL ECLIPSE BEGINS	MIDDLE OF ECLIPSE	TOTAL ECLIPSE ENDS	MOON LEAVES UMBRA
1970 Feb 21	Partial	$08^h03^m.1$		$08^h30^m.7$		$08^h58^m.4$
Aug 17	Partial	02 17 .9		03 24 .1		04 30 .3
1971 Feb 10	Total	05 52 .6	$07^h03^m.8$	07 45 .3	$08^h26^m.8$	09 38 .0
Aug 6	Total	17 55 .8	18 53 .8	19 43 .9	20 33 .9	21 32 .0
1972 Jan 30	Total	09 12 .0	10 35 .8	10 54 .1	11 12 .2	12 36 .0
Jul 26	Partial	05 55 .9		07 16 .3		08 36 .9
1973 Jan 18	Penumbral			21 17 .9		
Jun 15	Penumbral			20 50 .7		
Jul 15	Penumbral			11 39 .2		
1973 Dec 09–10	Partial	01 09 .8		01 45 .1		02 20 .4
1974 Jun 4–5	Partial	20 39 .5		22 16 .7		23 53 .9
Nov 29	Total	13 29 .3	14 35 .8	15 14 .1	15 52 .4	16 58 .9
1975 May 25	Total	04 00 .8	05 04 .2	05 48 .8	06 33 .3	07 36 .7
Nov 18–19	Total	20 39 .4	22 03 .4	22 24 .2	22 44 .9	00 09 .0
1976 May 13	Partial	19 16 .5		19 55 .1		20 33 .7
Apr 14	Penumbral					
Nov 17	Penumbral					
1977 Apr 4	Partial	03 31 .0		04 19 .0		05 07 .2
Sep 27	Penumbral					
Oct 26	Penumbral					
1978 Mar 24	Total	14 33 .6	15 37 .6	16 23 .2	17 08 .9	18 12 .8
Sep 16	Total	17 21 .0	18 25 .2	19 05 .0	19 44 .7	20 48 .9
1979 Mar 13	Partial	19 29 .7		21 08 .8		22 48 .1
Sep 6	Total	09 18 .7	10 32 .1	10 55 .0	11 17 .9	12 31 .3
1980 Mar 1	Penumbral					
Jul 27	Penumbral					
Aug 25	Penumbral					

* Courtesy of R. Duncombe, U.S. Naval Observatory

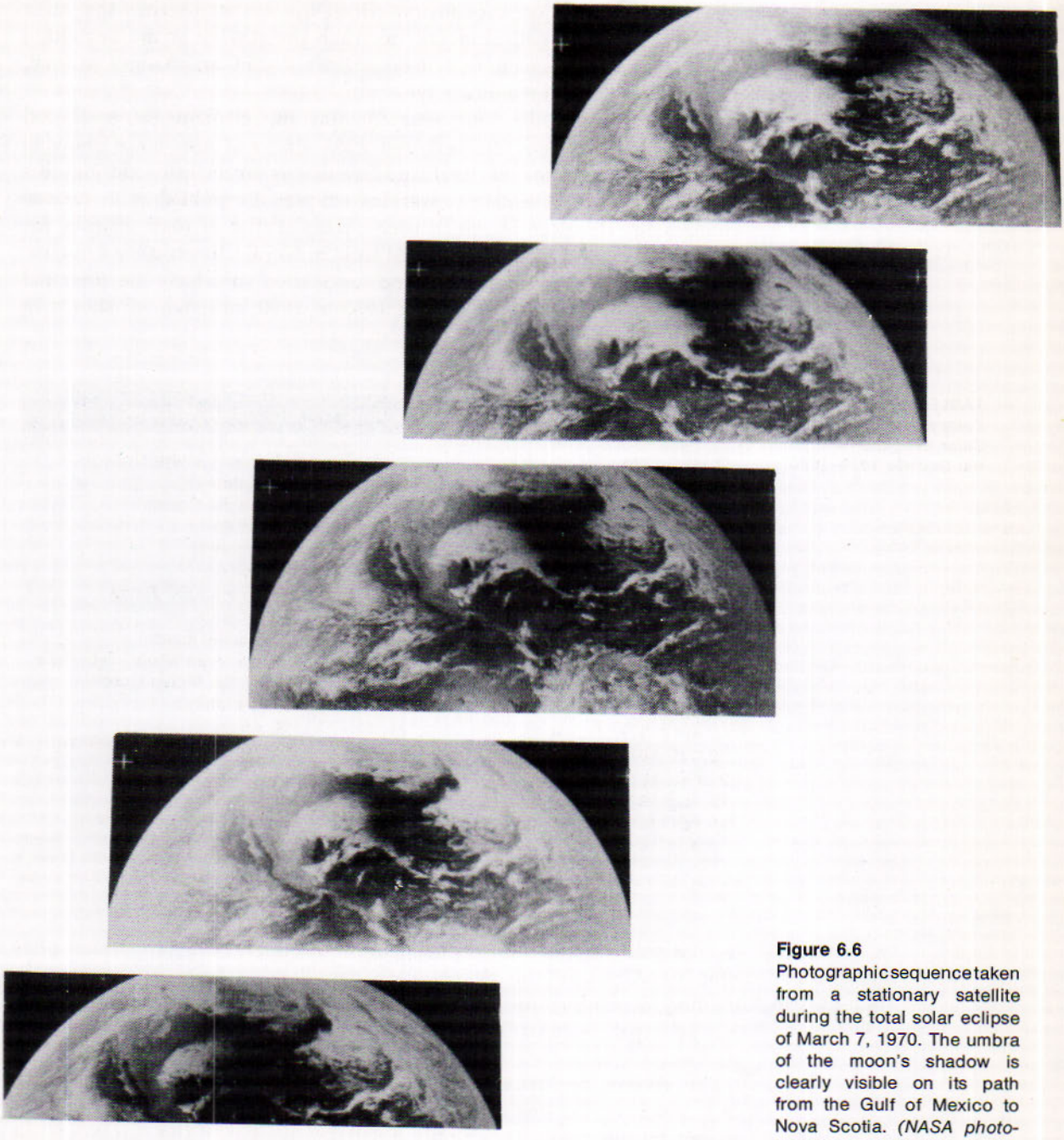


Figure 6.6
Photographic sequence taken from a stationary satellite during the total solar eclipse of March 7, 1970. The umbra of the moon's shadow is clearly visible on its path from the Gulf of Mexico to Nova Scotia. *(NASA photograph)*

An eclipse occurs occasionally in which the umbra touches the earth at the middle of its path, but fails to reach the surface at the beginning and end of its path. Such an eclipse begins as annular, changes to total, and later reverts to the annular type.

Oppolzer's *Canon der Finsternisse* contains the elements of solar and lunar eclipses between 1208 B.C. and 2163 A.D., and also maps showing the approximate paths of total and annular solar eclipses during this interval. More accurate data concerning eclipses are published in various almanacs for the year in which each occurs. Paths of total eclipses for many years in advance are published in U.S. Naval Observatory *Circulars*.

The dates, durations at noon, and land areas in which the principal total solar eclipses are visible from 1970 to 1980 inclusive are shown in table 6.II.

TABLE 6.II
Total and Annular Solar Eclipses in the Decade 1970–1980*

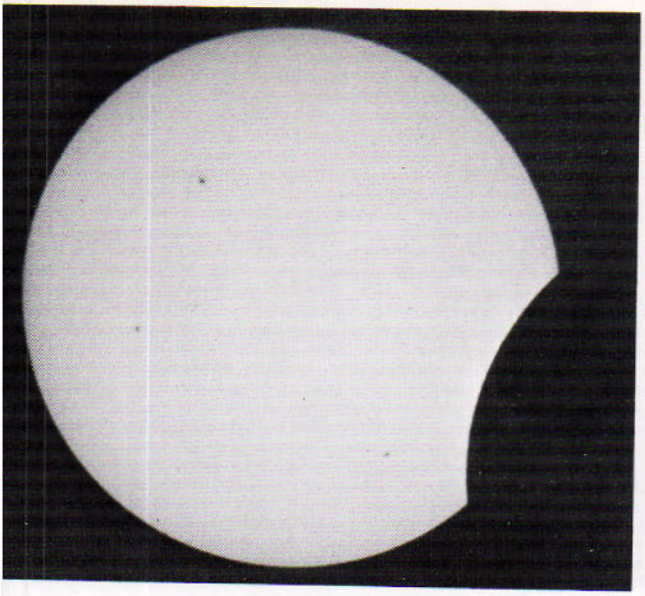
DATE	TYPE	DURATION (min)	GENERAL AREA
7 Mar 1970	T	3	Mexico, Eastern U.S.A., Canada
31 Aug 1970	A	6	S. Pacific Ocean
16 Jan 1972	A	brief	Antarctic
10 Jul 1972	T	3	N. America
4 Jan 1973	A	8	S. America
30 Jun 1973	T	7	S. America, N. Africa
24 Dec 1973	A	12	S. America
20 Jun 1974	T	brief	S. Indian Ocean
29 Apr 1976	A	7	Africa, Asia Minor, S.E. Asia
23 Oct 1976	T	5	S. Africa, Indian Ocean, Australia
18 Apr 1977	A	7	S. Africa
12 Oct 1977	T	3	Pacific Ocean
26 Feb 1979	T	3	N. America
22 Aug 1979	A	6	Antarctic
16 Feb 1980	T	4	Africa, India, S.E. Asia
10 Aug 1980	A	3	Pacific, S. America

A = Annular, T = Total *Courtesy of R. Duncombe, U. S. Naval Observatory

6.7 The Sun in Total Eclipse

A total solar eclipse ranks among the most impressive of celestial phenomena. Although the details vary considerably from one eclipse to another, depending on the diameter of the shadow and other factors, the principal features to be noted are much the same on all these occasions.

The eclipse begins at a particular place with the appearance of a dark notch at the sun's western edge. Thereafter the sun is gradually hidden by the moon. When only a narrow crescent of the sun is left, an unfamiliar pallor overspreads the sky and landscape. Immediately before



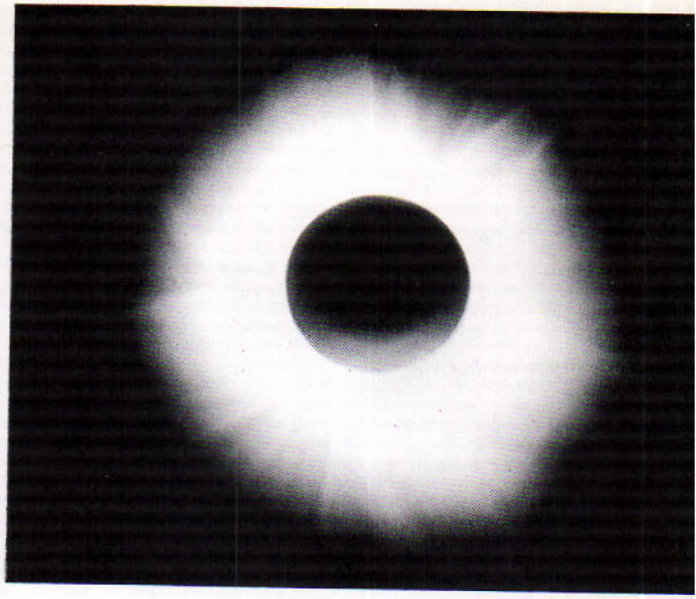


Figure 6.7A (Left)
The face of the sun at the start of the March 7, 1970 eclipse. Notice three spot groups. *(Photograph by Pocahontas)*

Figure 6.7B (Right)
The total solar eclipse of May 20, 1947 at Bocainva, Brazil. Note the round structure of the corona, typical of periods of maximum solar activity. Compare with Figure 10.29A. *(Georgetown University Observatory photograph)*

totality the sky darkens rapidly; shadow bands, like ripples, move across white surfaces; some animals become disturbed and some flowers begin to close. As the umbra of the shadow rushes in on the observer, the remaining sliver of the vanishing sun breaks into bright "Baily's beads" and quickly disappears. The so-called "Baily's beads" are caused by the rim of the sun when it shines through the irregularities on the limb of the moon. With the coming of totality the corona bursts into view; it is brightest close to the eclipsing moon and fades out in streamers. Flame-like prominences sometimes appear; their bases near the west edge of the sun are gradually uncovered, while those around the east edge are hidden as the moon moves across. Some bright stars and planets in the sun's vicinity may become visible to the unaided eye.

Totality ends as abruptly as it began. The corona vanishes, and the features of the partial eclipse recur in reverse order.

6.8 Value of Total Solar Eclipses

As spectacles to be long remembered, total eclipses of the sun have a popular appeal that may lead to further interest in astronomy. These brief occasions when the sun's disk is completely hidden by the moon also offer opportunities for observing features near the sun's edge that are ordinarily concealed by the glare of the sunlight. Available then for study are details of the corona, some that cannot be seen at all at other times and some that can be observed when there is not an eclipse only by use of special devices. The visibility of objects near the sun's place

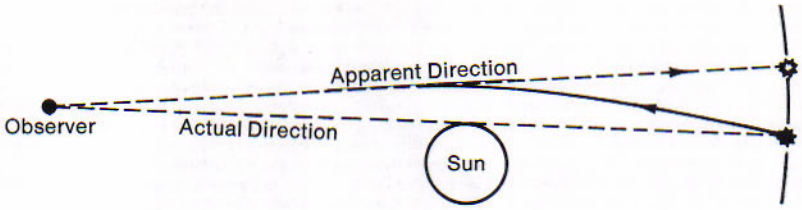


Figure 6.8
Apparent displacement of a star near the sun's edge. By the theory of relativity, starlight passing near the sun deflected so that the star is apparently displaced outward from the sun's position.

in the sky has permitted the search for a once suspected planet nearer the sun than Mercury's distance, the discovery of certain comets, and a successful test of the theory of relativity.

The theory of relativity requires that stars close to the sun's place be apparently displaced slightly away from the center of the sun; the predicted maximum displacement, for a star at the sun's edge, is 1.″75. According to the theory, space around a massive body is warped, thus a photon travelling along a straight line ("geodesic" is the correct term) appears to be deflected. Such displacements of stars were first observed by English astronomers at an eclipse in 1919 and have been verified at subsequent eclipses. The procedure has been to compare photographs of the region of the sky immediately around the eclipsed sun with other photographs of the same region taken at night at another time of the year. A comparison by G. Van Biesbroeck at the solar eclipse of 1952 February 25 showed an average displacement of 1.″70 at the sun's edge.

Solar eclipses are particularly useful in helping to clarify the dates of early events in history, because the dates and paths of the eclipses are precisely determined. This has proven very helpful also in locating ancient towns and cities whose exact geographical situation had been obliterated by the effects of time, but of which some historical or legendary mention existed. Lunar eclipses have been of some value in this respect also.

6.9 Eclipse Seasons

Eclipses of the sun and moon occur respectively when the moon is new and full. Although these phases recur every month, eclipses come less frequently. The reason is that the moon's apparent path on the celestial sphere is inclined 5° to the ecliptic. Each time when the moon overtakes the sun, or the earth's shadow opposite the sun, both sun and shadow have moved eastward on the ecliptic from the previous positions. Traveling in its path inclined to the ecliptic, the moon passes north or south of the sun and shadow, unless the two are near the intersections of the moon's path and the ecliptic. Only then can the moon eclipse the sun or be eclipsed in the shadow of the earth.

Eclipse seasons occur around the times when the sun passes one of the nodes of the moon's path. As the nodes regress rapidly westward, the eclipse seasons come more than half a month earlier from year to year. The interval between two successive arrivals of the sun at the same node is the *eclipse year;* its length is 346.620 days. The eclipse seasons in 1972 are in January and July.

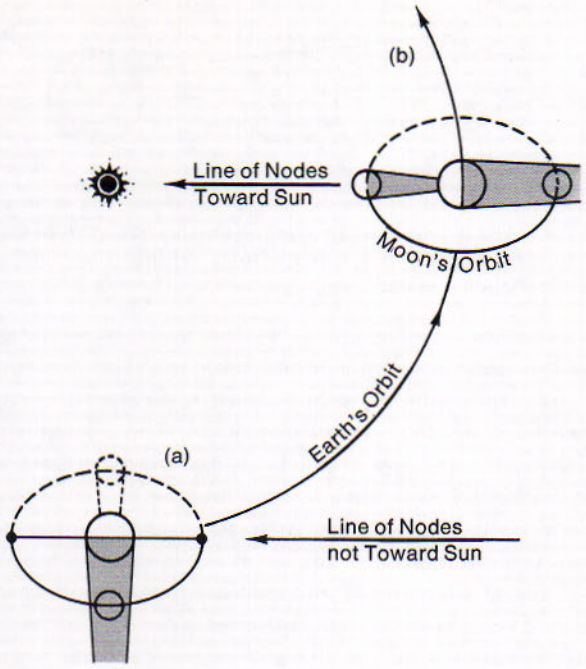


Figure 6.9
The inclination of the moon's orbital plane to that of the earth prevents the occurrence of eclipses except during two opposite seasons, such as *b* when the sun is near the line of nodes of the moon's path. At other times, as at *a*, the moon cannot pass between the earth and the sun or through the earth's shadow.

6.10 Solar and Lunar Ecliptic Limits

The *solar ecliptic* limit is the angular distance of the sun from the node at which it is grazed by the moon, as seen from some point on the earth. Within this distance the sun will be eclipsed; beyond it an eclipse cannot occur. The value of this limiting angular distance varies with the changing linear distances from us, and therefore the apparent sizes, of the sun and moon, and with fluctuations in the angle between the moon's path and the ecliptic. The extreme values, or major and minor limits, are respectively 18° 31′ and 15° 21′. When the sun's center is beyond the major limit, an eclipse is impossible; when it is within the minor limit, an eclipse is inevitable.

The *lunar ecliptic limit* (Fig. 6.10) is likewise the greatest distance of the sun from the node at which a lunar eclipse is possible. The major and minor limits for umbral eclipses are 12° 15′ and 9° 30′, and for penumbral eclipses are the same as for solar eclipses.

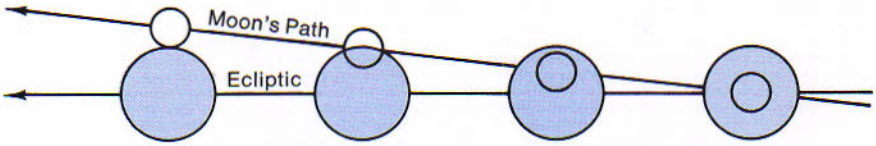


Figure 6.10
The lunar ecliptic limit. In order to eclipse the moon, the earth's shadow must be near one of the nodes of the moon's path. The greatest distance of the center of the shadow from the node, or of the sun's center from the opposite node, at which an eclipse is possible is the lunar ecliptic limit.

6.11 Frequency of Eclipses

The number of eclipses during each eclipse season is determined by comparing the double ecliptic limits with the distance the sun moves along the ecliptic in a synodic month with respect to the regressing node; this distance is 29.5/346.6 of 360°, or 30°.6. The question is whether the sun, and the earth's shadow opposite it, can possibly pass through the eclipse region without being encountered by the moon. They can do so if the double ecliptic limit is less than 30°.6, although usually they may not escape even then. If the double limit is greater than 30°.6, one eclipse must occur at each node, and two are possible. Because the eclipse year is 18.63 days shorter than the average calendar year, the first eclipse season may return before the end of the calendar year, and in this event one additional eclipse may result.

Two solar eclipses of some kind must occur each year; for twice the minor solar ecliptic limit is 30°.7. Five may occur, and as many as 3 of these may be total or annular. Two lunar eclipses of some kind must occur each year, and 5 are possible; of these no umbral lunar eclipse need occur, but 3 are possible.

The minimum number of eclipses in a year is therefore 4, 2 of the sun and 2 of the moon which may both be penumbral. The maximum number is 7. Thus solar and lunar eclipses are equally frequent for the earth as a whole, although many penumbral lunar eclipses among them cannot be detected. At any one place, lunar eclipses are seen the more often because of the greater area of the earth from which a lunar eclipse is visible.

6.12 Recurrence of Eclipses

The *saros* is the interval of $18^{y}\ 11\frac{1}{3}^{d}$ (or a day less or more, depending on the number of leap years included) after which eclipses of the same series are repeated. It is equal to 223 synodic months, which contain 6585.32 days, and is nearly the same in length as 19 eclipse years (6585.78 days). Not only have the sun and moon returned to nearly the same positions relative to each other and to the node, but their distances from us are nearly the same as before, so that the durations of succeeding eclipses in a series also differ very little. Knowledge of the saros, as it applies to cycles of lunar eclipses at least, goes back to very early times.

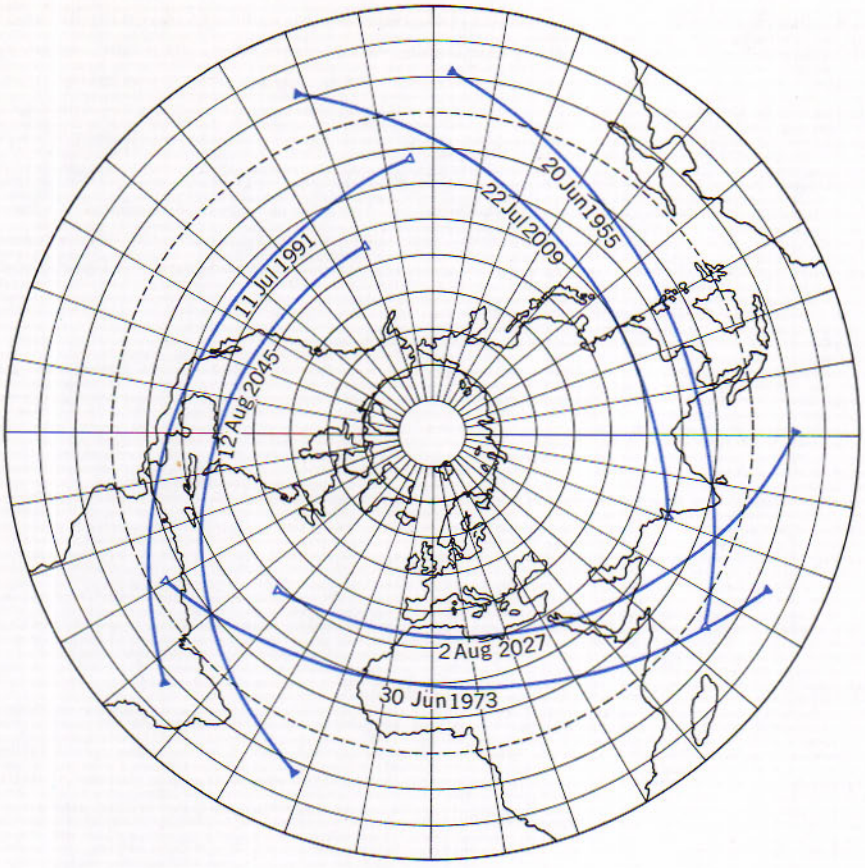


Figure 6.12
Paths of a family of solar eclipses through the year 2045.

The effect of the one-third of a day in the saros period is to shift the path of the following eclipse 120° west in longitude, and after 3 periods around to nearly the same part of the earth again. At the end of each period the sun is nearly a half-day's journey, or about 28′, west of its former position relative to the node. Thus a gradual change in the character of succeeding eclipses in a series is brought about, together with a progressive shift of their paths in latitude.

Eclipses at intervals of the saros accordingly fall into series, each series of solar eclipses containing about 70 eclipses in a period of about 1200 years. A new series is introduced by a small partial eclipse near one of

the poles. After a dozen partial eclipses of increasing magnitude and decreasing latitude of their paths, the series becomes total or annular for 45 eclipses, reverts to about a dozen diminishing partial eclipses, and finally disappears at the opposite pole. A family that begins at the descending node enters at the south pole and exits at the north pole, one beginning at the ascending node moves in the opposite direction. The family represented by the eclipses of 1955 and 1973 is remarkable because the descending node occurs near perigee and hence the durations of totality are not far from the greatest possible one. Figure 6.12 shows six eclipses from this family and graphically demonstrates the progressive shifting of the path of totality.

There are twelve notable families of total solar eclipses in progress at present. A family of lunar eclipses runs through about 50 saroses in a period of about 870 years.

6.13 Occultations by the Moon

In its eastward movement around the heavens the moon frequently passes over, or *occults*, a star. Because the moon moves a distance equal to its own diameter in about an hour, this is the longest duration of an occultation. The star disappears at the moon's eastern edge and emerges at the western edge. Observations of the times of such occurrences have been useful for determining the moon's positions at those times for comparison with its predicted positions. These data are now obtained more conveniently and precisely with the dual-rate moon position camera (5.12).

The predicted times of disappearance behind the moon and of subsequent reappearance of the brighter stars and planets, which were formerly published in the American and British Almanacs, have been omitted from these almanacs beginning with the year 1960. The predictions for 12 stations in the United States and Canada are now included annually in a special Occultation Supplement in *Sky and Telescope.*

Figure 6.13A
Jupiter and its satellites emerging from behind the moon. *(Griffith Observatory photograph)*

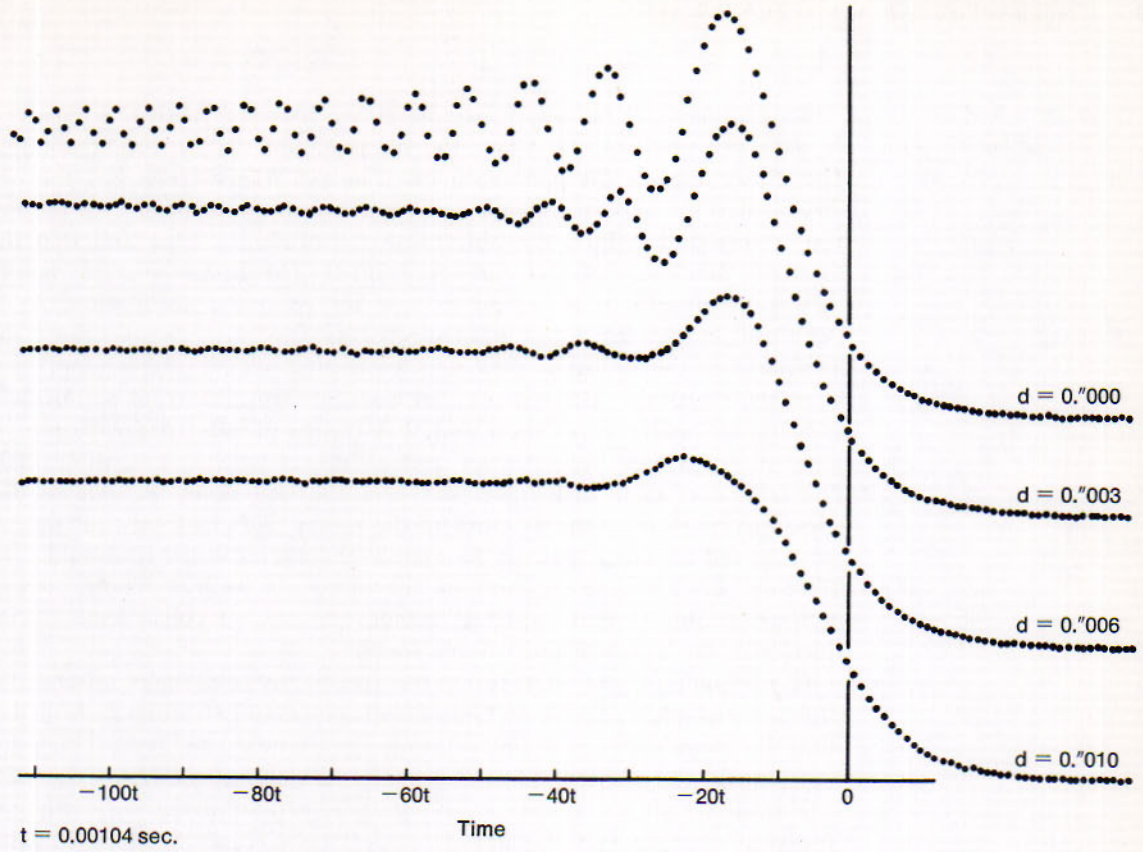
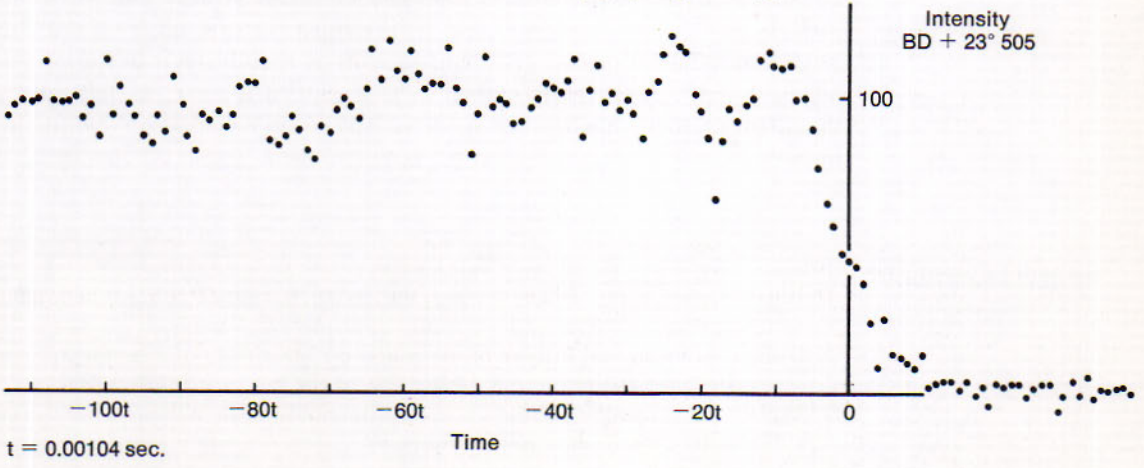


Figure 6.13B
Lunar occultation curves. The upper four are *theoretical* occultation curves for a star regarded as a point source 0″.000 and for stars subtending 0″.003, 0″.006 and 0″.010. Below is the *observed* occultation curve of the star BD +23° 505 indicating an effective point source. *(Leander-McCormick Observatory graph by R. Berg)*



$\alpha'' = 3.5 \times 10^{-5}\, d$
d = aperture in cm.

Occultations by the moon are interesting to watch with the telescope or with the unaided eye when the occulted objects are bright enough. The abruptness of disappearance or of reappearance convinces the observer that the moon has not sufficient atmosphere around it to dim appreciably and redden the objects near the moon's edge and that the apparent diameters of stars are very small. The disks of stars require only a hundredth of a second or less for complete occultation. A. E. Whitford, employing a fast photoelectric system, first demonstrated the difference in the apparent diameters of several stars by observing these intervals. Physical optics tells us that a knife edge diffraction pattern will appear just before occultation and that the amplitude of the pattern is different for a true point source and an object with a discernible disk. The least detectable disk (α) is simply a constant times the ratio of the telescope aperture to the distance to the moon. This method thus far has not enjoyed as much success as simple theory indicates it should. The differing conditions at the limb of the moon, as well as seeing, narrow declination limits, finite spectral bands and the response time of the electronics are a few of the reasons for this.

The occultations of planets by the moon are sometimes spectacular. The unusual photographs of Jupiter and its satellites emerging from behind the moon were taken on the morning of January 16, 1947. The moon was then in the crescent phase between last quarter and new. The dark side from which the planet is emerging was illuminated by earthlight. The oval dark spot in the upper part of the pictures is Mare Crisium. The times are Pacific standard.

Jupiter disappeared behind the bright edge of the moon at 4 o'clock that morning and reappeared about 50 minutes later at the western edge. The first photograph, taken at 4:50, shows the 4th and 1st satellites already in sight. In the second photograph, taken at 4:54, the planet has just cleared the moon's edge, and in the third picture, taken at 5:03, the 3rd satellite has appeared. The 2nd satellite is readily seen in the original negative but is less clearly shown in the print; it is near the left edge of the planet's disk, having emerged at 3:57 from occultation behind the planet.

REVIEW QUESTIONS

1. What do we mean when we say the moon has undergone a penumbral eclipse?
2. If there is an umbral eclipse of the moon on March 21, what is the approximate right ascension and declination of the moon at mid-eclipse?
3. Explain why lunar eclipses are more commonly seen although they occur less frequently than solar eclipses.

4. Why are some solar eclipses annular and others total?
5. The solar eclipse of 2089 Oct 4 begins at 97°E longitude, 29°N latitude. The next eclipse of this series begins at 23°W, 25°N on what date? What are the coordinates and date for the beginning of the second eclipse of this series after the 2089 eclipse?
6. Assuming that the eccentricity of the earth's orbit is .0102, show that the earth's shadow is longer at aphelion by a factor of 51/49 compared to the length at perihelion.
7. Define the ecliptic limit. How does it vary?
8. How many solar eclipses must occur each year?
9. Discuss the value of solar and lunar eclipses.
10. Of what value are lunar occultations?

REFERENCES

Dyson, Sir Frank Watson and R. van der R. Woolley, *Eclipses of the Sun and Moon*, Oxford: Clarendon Press, 1937.

Oppolzer, Theodore R., *Canon der Finsternisse*, Vienna: 1885.

Smart, William M., *Spherical Astronomy*, New York: Dover Publications, 1944.

U.S. Government Printing Office, *The American Ephemeris and Nautical Almanac*, Washington, D.C.: Annual.

FOR FURTHER STUDY

Billings, Donald E., *A Guide to the Solar Corona*, New York: Academic Press, 1966.

Leander McCormick Observatory, University of Virginia.


7

THE SOLAR SYSTEM

THE PLANETS IN GENERAL — THE PATHS OF THE PLANETS — THE LAW OF GRAVITATION

The *solar system* consists of the sun and the many smaller bodies that revolve around the sun. These include the planets with their satellites, and the comets and meteors. It is a large system; the outermost planet, Pluto, is about 40 times as far from the sun as the earth's distance, and the majority of the comets have the aphelion points of their orbits still more remote. In comparison with the distance of even the nearest star, however, the interplanetary spaces shrink to such insignificance that we look upon the solar system as our own community and the other planets as our neighbors.

61″ Astronomic reflector building. U.S. Naval Observatory, Flagstaff Station, Arizona. *(Official U.S. Navy photograph)*

THE PLANETS IN GENERAL

The planets are classified by the positions with respect to the earth and their composition. The visible planets were long known to be different from the stars by their motions and lack of twinkling. Study has revealed an orderly arrangement for the motions and spacings of the planets.

7.1 The Principal and Minor Planets

The word *planets* (wanderers) was originally used to distinguish from the multitude of "fixed stars" the celestial bodies, except the comets, that move about among the constellations. Seven of these were known to early observers: the sun, the moon, and the 5 bright planets, Mercury, Venus, Mars, Jupiter, and Saturn, the last of which was believed to be almost as remote as the sphere containing the stars themselves. Thus Omar Khayyám ascended in his meditation "from earth's center through the seventh gate" to the throne of Saturn—from the center of the universe, as he understood it, almost to its limits.

With the acceptance of the Copernican system, the earth was added to the number of planets revolving around the sun, and the moon was recognized as the earth's satellite. Uranus, which is barely visible to the unaided eye, was discovered in 1781. Neptune, which is too faint to be seen without the telescope, was found in 1846. The still fainter Pluto, discovered in 1930, completes the list of the 9 known *principal planets.* In the meantime, in 1801, Ceres, the largest of the *asteroids,* or *minor planets,* was detected.

7.2 The Planets Named and Classified

The planets in order of average distance from the sun are:

- Inferior planets
  - Mercury
  - Venus
  - Earth
  - Mars
  - (Inner planets)
- The Asteroids or Minor planets
- Superior planets
  - Jupiter
  - Saturn
  - Uranus
  - Neptune
  - Pluto
  - (Outer planets)

As shown above, the planets are classified either as *inferior* or *superior* depending on whether their orbits around the sun are respectively smaller or larger than the earth's; or as *inner* or *outer* according to their relative

position inside or outside the orbital region of the asteroids. The inner planets are sometimes referred to as *terrestrial*, while the outer planets are called *gaseous* and also *jovian*.

7.3 Planets Distinguished from Stars

Planets are relatively small globes, which revolve around the sun and shine by reflected sunlight. Five of the principal planets look like brilliant stars in our skies, while a sixth is faintly visible to the naked eye. Examined with the telescope, the larger and nearer planets appear as disks. The stars themselves are remote suns shining with their own light; they appear only as points of light even with the largest telescopes.

The bright planets can often be recognized by their steadier light when the stars around them are twinkling. All planets can be distinguished by their motions relative to the stars. The right ascensions and declinations of the principal planets and four asteroids are tabulated in the *American Ephemeris and Nautical Almanac* at convenient intervals during the year; their positions in the constellations can be marked in the star maps for any desired date.

The planet Mercury is occasionally visible to the naked eye in the twilight near the horizon, either in the west after sunset or in the east before sunrise. Venus, the most familiar evening and morning "star" is the brightest starlike object in the heavens. Mars is distinguished by its red color; at closest approaches to the earth it outshines Jupiter, which is generally second in brightness to Venus. Saturn rivals the brightest stars and is the most leisurely of the bright planets in its movement among the stars.

7.4 The Revolutions of the Principal Planets

The revolutions of the principal planets around the sun conform approximately to the following regularities, which seem to have survived from an orderly origin and evolution of the planetary system (9.32):

1. The orbits are nearly circular. They are ellipses of small eccentricity, but with more marked departures from the circular form in the cases of Pluto and Mercury.
2. The orbits are nearly in the same plane. With the prominent exception of Pluto, the inclinations of the orbits to the ecliptic do not exceed 8°; so that these planets are always near the ecliptic and are generally within the boundaries of the zodiac.
3. The revolutions are all *direct*, that is, in the same direction (west to east) as the earth's revolution. This is the favored direction of revolution and rotation of all planets and satellites, as compared with *retrograde* motion (east to west), except for the rotation of Venus and the revolution of a few minor satellites of Jupiter and Saturn.
4. The distances of the planets from the sun are given approximately by the relation referred to as the Titius-Bode Relation (7.5).

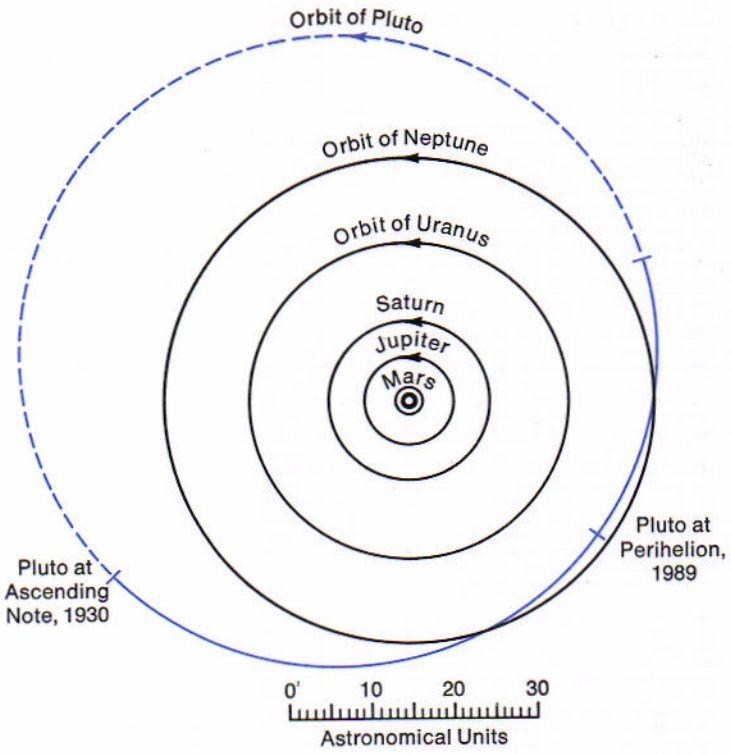


Figure 7.4
Approximate orbits of the principal planets. The orbits are in general nearly circles with the sun at the common center and are nearly in the same plane. The orbit of Pluto departs most conspicuously from these regularities.

The conspicuous departures from these regularities may be equally significant, as we note later, in showing alterations in the planetary system subsequent to its formation. As an example, it has been proposed that Pluto may have been a satellite of Neptune, which escaped from that planet. This theory, however, is not seriously considered at the present.

7.5 Bode's Relation

A convenient relation first proposed by Titius and championed by Bode is obtained by writing the numbers 0, 3, 6, 12, 24, . . . , doubling the number each time to obtain the next one, then by adding 4 to each number and dividing the sums by 10. The resulting series of numbers, 0.4, 0.7, 1.0, 1.6, 2.8, . . . , represents the mean distances of the planets from the sun, expressed in astronomical units (Table 7.I). The astronomical unit is the earth's mean distance from the sun.

TABLE 7.I
Distances and Periods of the Planets

NAME	MEAN DISTANCE FROM SUN			PERIOD OF REVOLUTION	
	Bode's Relation	*Astron. Units*	*Million Kilometers*	*Sidereal*	*Mean Synodic*
				days	days
Mercury	0.4	0.39	58	88	116
Venus	0.7	0.72	108	225	584
Earth	1.0	1.00	150	365¼	
Mars	1.6	1.52	228	687	780
				years	
Ceres	2.8	2.77	414	5	467
Jupiter	5.2	5.20	778	12	399
Saturn	10.0	9.54	1430	29	378
Uranus	19.6	19.18	2870	84	370
Neptune	38.8	30.06	4497	165	367
Pluto		39.52	5912	248	367

There was at the time when Bode called attention to the relation, however, one exception to it; the number 2.8 between the numbers for Mars and Jupiter corresponded to the mean distance of no known planet. Bode pointed out that the success of the relation in other respects justified a search for the missing planet. The discovery of Uranus, in 1781, at a distance in satisfactory agreement with the series extended one term further, so strengthened his position that a systematic search for a planet at 2.8 A.U. was undertaken by a group of European astronomers. As it turned out, the asteroid Ceres was discovered accidentally by Piazzi, in 1801, at the expected distance of 2.8 astronomical units. This is very nearly the average distance of the many hundreds of asteroids since discovered. The Titius-Bode Relation does not, however, predict Neptune. That is why it is not called a law.

$$\frac{A + B2^n}{10}$$

where
$A = 4$
$B = 3$
$n = -\infty, 0, 1, 2, \ldots$

7.6 Sidereal and Synodic Periods

The *sidereal period* of a planet is the interval between two successive returns of the planet to the same point in the heavens, as seen from the sun. It is the true period of the planet's revolution around the sun. This interval ranges from 88 days for Mercury to 248 years for Pluto.

The *synodic period* is the interval between two successive conjunctions of a planet with the sun, as seen from the earth; for an inferior planet the conjunctions must both be either inferior or superior (Fig. 7.7). It

is the interval after which the faster-moving inferior planet again overtakes the earth, or the earth again overtakes the slower superior planet. The relation between the two periods for any planet is:

$$\frac{1}{\text{synodic period}} = \pm\frac{1}{\text{sidereal period}} \mp \frac{1}{\text{earth's sidereal period}},$$

where the upper signs are for an inferior planet and the lower signs are for a superior planet. This is merely the statement of the fact that the rate at which the other planet gains on the earth, or the earth gains on the other planet, is the difference of the angular rates of their revolutions around the sun.

Mars and Venus have the longest synodic periods for the principal planets (Table 7.I) because they run the closest race with the earth. The synodic periods of the outer planets approach the length of the year as their distances from the sun, and therefore their sidereal periods, increase.

7.7 Aspects and Phases of the Inferior Planets

Because the inferior planets, Mercury and Venus, revolve faster than the earth does, they gain on the earth and therefore appear to us to oscillate to the east and west with respect to the sun's place in the sky. Their

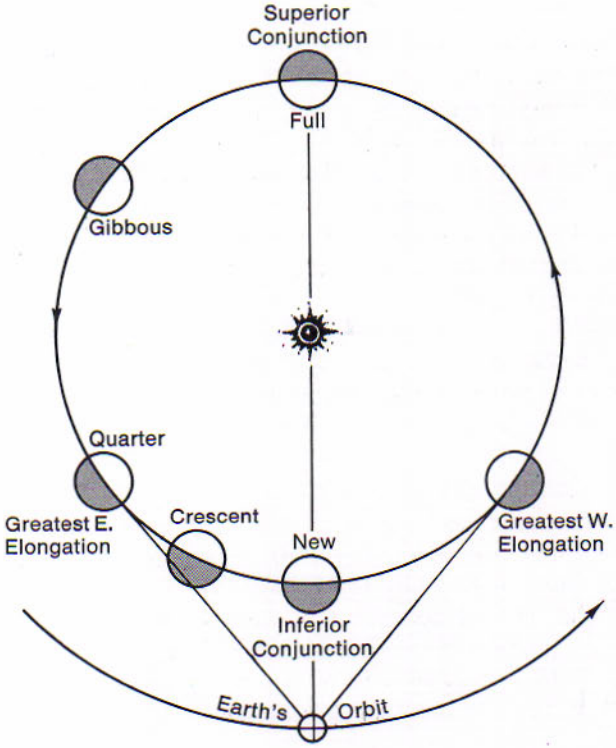


Figure 7.7
Aspects and phases of an inferior planet. The aspects differ from those of the moon. The phases are the same as the moon's.

aspects are unlike those of the moon (5.5), which has all values of elongation up to 180°.

After passing *superior conjunction* beyond the sun, the inferior planet emerges to the east of the sun as an evening star and slowly moves out to *greatest eastern elongation.* Here it turns west and, apparently moving more rapidly, passes between us and the sun at *inferior conjunction* into the morning sky. Turning east again at *greatest western elongation,* it returns to superior conjunction. Greatest elongation does not exceed 28° for Mercury and 48° for Venus.

The phases of the inferior planets resemble those of the moon (5.6). As these planets revolve within the earth's orbit, their sunlit hemispheres are presented to the earth in varying amounts; they show about the full phase at the time of superior conjunction, the quarter phase in the average near the elongations, and about the new phase at inferior conjunction.

Because the superior planets revolve more slowly than the earth does, they move eastward in the sky more slowly than the sun appears to do, so that they are overtaken and passed by it at intervals. With respect to the sun's position they seem to move westward (clockwise in Fig. 7.8), and to attain all values of elongation from 0° to 180°. The aspects of the superior planets are the same as those of the moon.

7.8
Aspects and Phases of the Superior Planets

Jupiter, as an example, emerges from *conjunction* to the west of the sun. It is then visible as a morning star, rising at dawn in the east. Jupiter appears to move westward with respect to the sun, because its eastward motion is not as fast as that of the sun. As Jupiter slips westward it comes successively to *western quadrature* when it is on the meridian at sunrise, to *opposition* when it is on the meridian at midnight, and to *eastern quadrature* when it is on the meridian at sunset. Setting earlier from night to night as it approaches the next conjunction with the sun, the planet is finally lost in the twilight in the west.

The superior planets do not exhibit the whole cycle of phases that the moon shows. At the conjunctions and oppositions their disks are fully illuminated, and in other positions they do not depart much from the full phase; for the hemisphere turned toward the sun is nearly the same as the one presented to the earth.

The *phase angle* is the angle at the planet between the directions of the earth and sun; divided by 180°, it gives the fraction of the hemisphere turned toward the earth that is in darkness. The phase angle is greatest when the planet is near quadrature; the maximum value is 47° for Mars, 12° for Jupiter, and is successively smaller for the more distant planets. Thus the superior planets show nearly the full phase at all times with the conspicuous exception of Mars, which near quadrature resembles the gibbous moon.

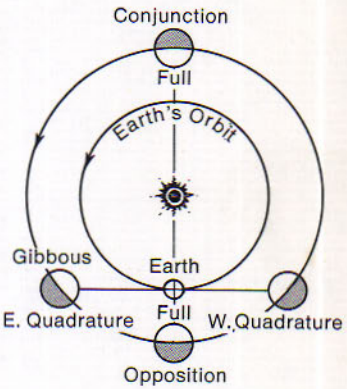


Figure 7.8
Aspect and phases of a superior planet. The aspects are similar to those of the moon. The only phases are full and gibbous.

THE PATHS OF THE PLANETS

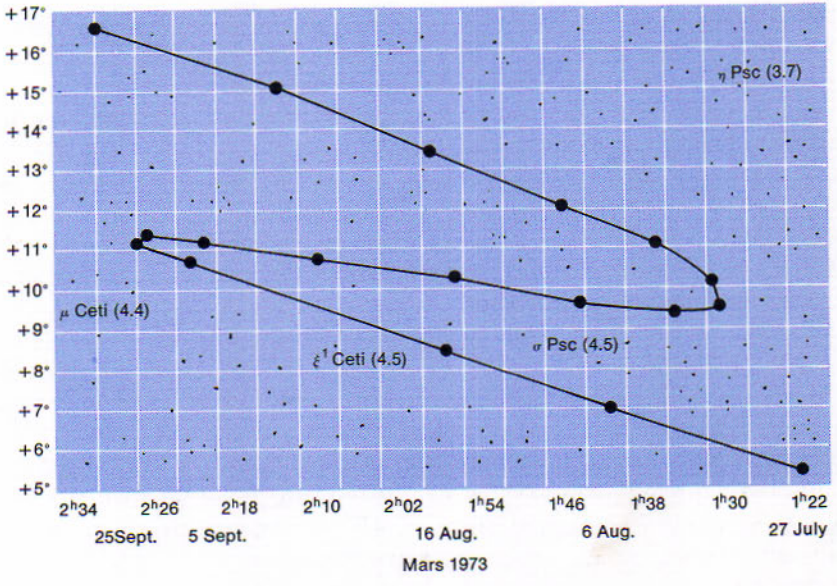
The first task of astronomy as a science was to explain the peculiar apparent motions of the planets with a simple, elegant theory. The accomplishment of this task led naturally to Kepler's celebrated laws and the scale of the solar system and has profoundly effected mankind's philosophical thought.

7.9 Apparent Motions Among the Stars

It is instructive to observe not merely that the planets move among the constellations, but also the complex paths they follow. Mars serves well as an example. Two or three months before the scheduled date of an opposition of Mars, note its position in the sky relative to nearby stars, and mark the place and date on a star map. Repeat the observation about once a week as long as the planet remains in view in the evening sky. A smooth curve through the plotted points, as in Fig. 7.9, represents the apparent path.

Against the background of the stars the motion is generally eastward, or *direct*, the same as the direction of the revolution around the sun. Once

Figure 7.9
Predicted path of Mars during the favorable opposition of 1973.



during each synodic period the planet turns and moves westward, or *retrogrades*, for a time before resuming the eastward motion. Thus the planets appear to move among the stars in a succession of loops, making progress toward the east and generally not departing far from the ecliptic.

7.10 Retrograde Motions Explained

The earth's eastward movement in its orbit around the sun tends to shift the planets backward, toward the west, among the stars. It is the same effect that one observes as he drives along the highway; objects pass by, and those nearer the road go by more rapidly than do those in the distance. This effect combines with the planets' real eastward motions to produce the looped paths that are observed.

A superior planet, such as Mars, retrogrades near the time of opposition; for the earth then overtakes the planet and leaves it behind. The direct motion becomes more rapid near conjunction, where the planet's orbital motion and its displacement caused by the earth's revolution are in the same direction.

An inferior planet retrogrades near inferior conjunction. This can be shown by extending the lines in Fig. 7.10 in the reverse direction, whereupon it is evident that the earth—an inferior planet relative to Mars, and then near inferior conjunction as viewed from that planet—is retrograding in the Martian sky. Mercury and Venus exhibit this effect to us. In general, a planet retrogrades for us when it is nearest the earth.

As long as the earth was believed to be stationary at the center of the system, the looped paths of the planets had necessarily to be ascribed entirely to the movements of the planets themselves. Complex motions such as these called for a complex explanation. The problem was finally simplified by the acceptance of the earth's revolution around the sun.

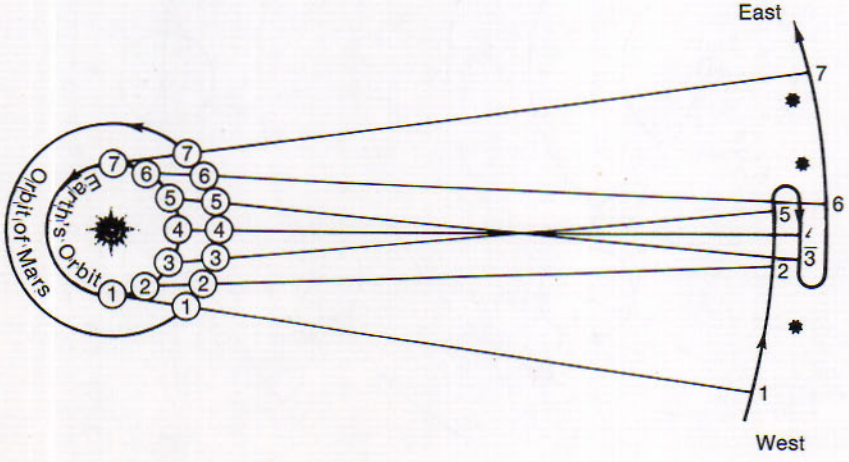


Figure 7.10
Retrograde motion of a superior planet. Positions of the earth at intervals of a month are numbered on the inner circle. Corresponding positions of Mars in its orbit are similarly numbered on the outer circle. As seen from the faster-moving earth, Mars retrogrades at positions 4 and 5, around the time of opposition.

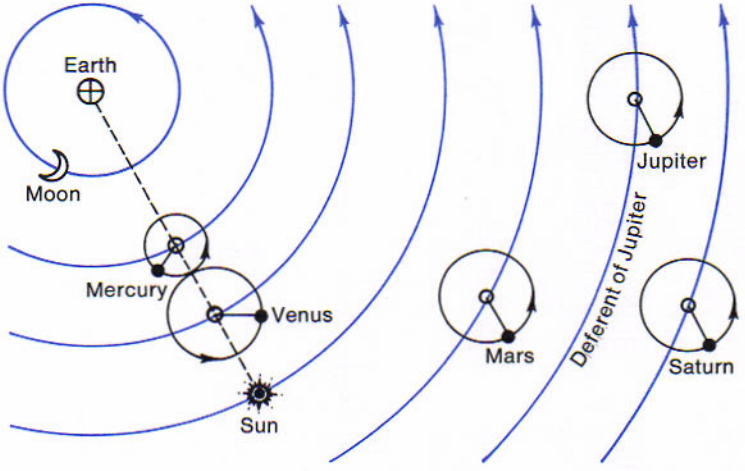
7.11 The Earlier Geocentric System

As early as the 6th century B.C., the earth was regarded by the Greek scholars as a stationary globe. Supported on an axis through the earth, the sky was a hollow concentric globe on which the stars were set; it rotated daily from east to west, causing the stars to rise and set. Within the sphere of the stars the sun, moon, and five bright planets shared in the daily rotation. They also revolved eastward around the earth, pausing periodically to retreat toward the west against the turning background of the constellations. The geocentric system remained almost unchallenged for more than 2000 years and was amplified meanwhile in attempts to account for the retrograde movements of the planets.

The problem of the planetary motions that the early scholars undertook to solve was simply kinematical. By what combination of uniform circular motions centered in the earth could the looped movements be represented? The most enduring solution of the early problem is known as the *Ptolemaic system,* because it is described in detail in Ptolemy's *Almagest.* In the simplest form of the system each planet moved uniformly in a circle, the epicycle, while the center of this circle was in uniform circular motion around the earth.

The working out of the solution consisted in obtaining by trial and error a better fit between the pattern and the apparent paths of the planets among the stars. The undertaking held the attention of astronomers, especially in the Arabian dominions, from Ptolemy's time to the revival of learning in Europe. Eventually the whole construction became cumbersome, but without satisfactory improvement in representing the observed planetary motions.

Figure 7.11
The Ptolemaic System of planetary motions.



7.12 The Heliocentric System

Nicolaus Copernicus (1473–1543) inaugurated a new era in astronomy by discarding the ancient theory of the central, motionless earth. In his book, *On the Revolutions of the Celestial Bodies*, published shortly before his death, he showed that all these motions could be interpreted more reasonably on the theory of the central sun. He assumed that the earth revolves around the sun once in a year and rotates daily on its axis.

In the *Copernican system* the sun was stationary at the center. Around it the planets revolved uniformly in circles, including the earth and its attendant, the moon. Epicycles were therefore still required, but their number was now smaller. With the additional assumption of the earth's rotation from west to east, the daily circling of all the celestial bodies from east to west became simply the scenery being passed by.

No longer required to rotate around the earth, the sphere of the stars could be imagined larger than before. This altered condition and the sun's new status as the dominant member of the system prepared the way for the thought that the stars are remote suns.

It is not surprising that the heliocentric theory met with disapproval on almost every hand; for it was a radical departure from the common-sense view of the world that had persisted from the very beginning of reflections about it. The new theory was supported at the outset by no convincing proof; its greater simplicity in representing celestial motions was the only argument Copernicus could offer in its defense. Moreover, it seemed to be discredited by the evidence of the celestial bodies themselves, as Tycho presently discovered.

7.13 Tycho's Observations

Tycho Brahe (1546–1601), native of the extreme south of Sweden, then a part of Denmark, spent the most fruitful years of his life at the fine observatory that the king of Denmark had financed for him on the island of Hven, about 32 km northeast of Copenhagen. During the last two years of his life, his observations were made at a castle near Prague.

The instruments of Tycho's observatory were mainly constructed of metal; they had larger and more accurately divided circles than any previously used. His improved methods of observing and his allowance for effects of atmospheric refraction, which observers before him had neglected, made it possible to determine the places of the celestial bodies in the sky with the average error of an observation scarcely exceeding a minute of arc. This was remarkable precision for observations made through the plain sights that preceded the telescope.

Tycho was unable to detect any annual variations in the relative directions of the stars, which he believed would be noticeable if the earth revolved around the sun. Either the nearer stars were so remote (at least 7000 times as far away as the sun) that their very small parallaxes could not be observed with his instruments, or else the earth did not revolve

around the sun. Because the first alternative required distances that seemed impossibly great to him at the time, Tycho rejected the Copernican assumption of the earth's revolution. However, in 1577 an event occurred that shook his belief. In that year a great comet appeared and Tycho's own observations showed that it was farther away than the moon. Since previously comets were believed to be atmospheric phenomena, here was an upsetting event that clearly required an unusual orbit in the Ptolemaic system to explain it.

As a substitute for the Copernican system, the *Tychonic system* again placed the earth stationary at the center. In that system the sun and moon circled around the earth, but the other planets revolved around the sun. Aside from slight effects that could not have been detected without a telescope, the Tychonic and Copernican systems were identical for calculations of the positions of the planets.

Tycho's most noteworthy contribution to the improvement of the planetary theory was his long-continued determinations of the places of the planets in the sky, their right ascensions and declinations at different times, especially of the planet Mars. These data provided the material for Kepler's studies, which resulted in his three laws of planetary motions.

7.14 Kepler's Studies

Johannes Kepler (1571–1630), a German, joined Tycho at Prague in 1600 and, as his successor, inherited the records of Tycho's many observations of the places of the planets. Beginning with the recorded places of Mars, Kepler at first undertook to represent them in the traditional way by combinations of epicycles and eccentrics, but was unable to fit all the places as closely as their high accuracy seemed to require. Experimenting further with ellipses having the sun at one focus, he was astonished to see the large discrepancies between observation and theory disappear.

Tycho's observations of the planets' right ascensions and declinations gave only their directions from the earth at the various dates. Kepler required for his studies their directions and distances from the sun. This he accomplished by the following device, as Fig. 7.14 is intended to show.

Consider the case of Mars, having a sidereal period of 687 days, and neglect for the present purpose the inclination of its orbit to the ecliptic. Compare two observations of the planet's position made 687 days apart. At the end of this interval Mars has returned to the same place in its orbit, while the earth is 43 days' journey west of its original place. Accordingly, the observed directions of Mars on the two occasions differ widely and, by their intersection, show where the planet is situated in space. By comparing such pairs of observations in different parts of the planet's orbit it is possible to determine the form of the orbit and its size relative to the earth's orbit.

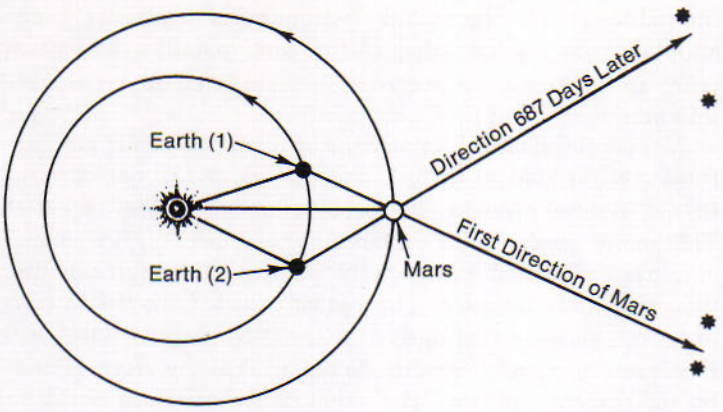


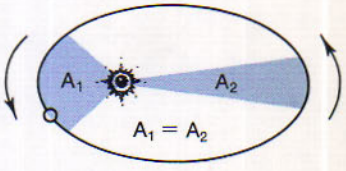
Figure 7.14
Kepler's method of determining the orbit of Mars. Pairs of apparent places of Mars separated by its sidereal period of 687 days gave the planet's direction and distance in astronomical units from the sun.

7.15 Kepler's Laws

Kepler's first two laws of the planetary motions were published in 1609; in his book entitled *Commentaries on the Motion of Mars.* The third, or *harmonic law,* the formulation of which gave him greater trouble, appeared in 1618, in his book *The Harmony of the World.* Kepler's laws are as follows:

1. The orbit of each planet is an ellipse with the sun at one of its foci.
2. Each planet revolves so that the line joining it to the sun sweeps over equal areas in equal intervals of time.
3. The squares of the periods (p) of any two planets are in the same proportion as the cubes of their mean distances (a) from the sun.

These laws assert that the planets revolve around the sun, but they do not necessarily include the earth as one of the planets. It was not yet possible to choose between the Copernican and Tychonic systems. The laws bring to an end the practice of representing planetary movements only by combinations of uniform circular motions. The third law determines the mean distances of all the planets from the sun in terms of the distance of one of them, when their sidereal periods of revolution are known. The usual yardstick is the earth's mean distance from the sun, which is accordingly called the *astronomical unit.*



7.16 The Scale of the Solar System

The problem of determining the scale of the solar system as accurately as possible is often called the solar parallax problem, because the sun's geocentric parallax may be regarded as the required constant. The corresponding value of the astronomical unit gives the linear scale. By

international agreement the astronomical almanacs since 1896 have adopted 8″.80 as the value of the solar parallax. The earth's mean distance from the sun is accordingly considered to be about 149.6 million kilometers.

Because the relative dimensions of planetary orbits are given by Kepler's third law restated to include the masses (7.22), one distance determined in the system provides the scale in miles as well as another distance. The many projects for obtaining the scale of the solar system have employed observed positions of planets that approach the earth nearer than the sun's distance. The optical result believed to have the highest order of precision was derived in 1950 by Eugene Rabe of the Cincinnati Observatory from his studies of the orbit of the asteroid Eros as perturbed by the principal planets. His value of the earth's mean distance from the sun is 149.6 million kilometers.

Radio methods of observing the scale of the system are also being employed effectively. The timing of many radar echoes from the planet Venus near its inferior conjunction in April, 1961, gave remarkably consistent results. These programs were reported by Millstone radar, Goldstone radar, and three other radio observatories and later extended through several conjunctions. The average value of the solar parallax was about 8″.7942 and the corresponding value of the earth's mean distance from the sun was about 149.6 million kilometers, or 92,950,000 miles.

7.17 Galileo; the Motions of Bodies

While Kepler was engaged in his studies of planetary orbits, Galileo Galilei (1564–1642), in Italy, was finding evidence with the telescope in favor of the Copernican system.

Galileo's discovery of four bright satellites revolving around the planet Jupiter dispelled the objection that the moon would be left behind if the earth really revolved around the sun. His discovery that Venus shows phases like those of the moon discredited the specification of the Ptolemaic system (Fig. 7.11) that kept the planet always on the earthward side of the sun, where it could never increase beyond the crescent phase. His explanation that the movements of spots across the sun's disk are caused by the sun's rotation provided an argument by analogy for the earth's rotation as well.

Galileo's chief contribution to the knowledge of the planetary movements was his pioneer work on the motions of bodies in general. His conclusion that an undisturbed body continues to move uniformly in a straight line, or to remain at rest, and his studies of the rate of change of motion of a body not left to itself prepared the way for a new viewpoint in astronomy. The interest was beginning to shift from the kinematics to the dynamics of the solar system—from the courses of the planets to the forces controlling them.

THE LAW OF GRAVITATION

After explaining the paths of the planets it then became necessary to find out what caused them to follow Kepler's laws. In a relatively brief period Isaac Newton expounded the required law and laid the groundwork for the detailed study of mechanics and its application to the heavens which is called celestial mechanics. The force, which Newton called gravity, *is classified as a weak force in modern physics and is still poorly understood.*

7.18 Force Equals Mass Times Acceleration

The concept of forces acting throughout the universe originates in our own experience with the things around us. If an object at rest, that is free to move, is pulled or pushed, it responds by moving in the direction of the pull or push. We say that force is applied to the object and, with allowance for disturbing factors such as air resistance and surface friction, we estimate the force by the mass of the object that is moved and the acceleration, or the rate at which its motion changes. In general, the acceleration of a body anywhere in any direction implies a force acting on it in that direction. The amount of the force is found by multiplying the mass of the body by its acceleration, or $f = m\alpha$.

Acceleration is defined as the rate of change of velocity. Since *velocity* is directed speed, acceleration may appear either as changing speed or changing direction, or both. A falling stone illustrates the first case; its behavior is represented by the relations:

$$v = v_0 + \alpha t; \qquad s = v_0 t + \tfrac{1}{2}\alpha t^2,$$

where v_0 is the speed when first observed, a is the acceleration of about 980.621 cm/sec^2 toward the earth, and v and s are the speed and the distance the stone has fallen after t seconds. If the stone starts at rest ($v_0 = 0$), it will fall about 490 cm in the first second, 1470 cm in the next second, and so on, increasing speed at the rate of 980 cm/sec^2.

A planet moving in a circular orbit illustrates acceleration in direction only; the speed is constant. If the planet describes an elliptic orbit, the speed also changes in accordance with Kepler's second law.

7.19 The Laws of Motion

The conclusions of Galileo and others concerning the relations between bodies and their motions were consolidated by Isaac Newton (1642–1727), in his *Principia* (1687), into three statements, which are substantially as follows:

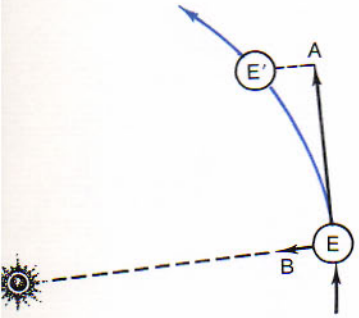


Figure 7.19
The earth's revolution explained by the laws of motion. At the position *E* the earth if undisturbed would continue on to *A*. It arrives at *E'* instead, having in the meantime fallen toward the sun the distance *EB*.

1. Every body persists in its state of rest or of uniform motion in a straight line, unless it is compelled to change that state by a force impressed upon it.
2. The acceleration is directly proportional to the force and inversely to the mass of the body, and it takes place in the direction of the straight line in which the force acts.
3. To every action there is always an equal and contrary reaction; or, the mutual actions of any two bodies are equal and oppositely directed.

The first law states that a body subject to no external forces moves uniformly in a straight line forever, unless it happens never to have acquired any motion. Up to the time of Galileo, the continued motion of a planet required explanation; since that time, uniform motion is accepted as no more surprising than the existence of matter itself. Changing motion demands an accounting.

The second law defines force in the usual way. Because nothing is said to the contrary, it implies that the effect of the force is the same whether the body is originally at rest or in motion, and whether or not it is acted on at the same time by other forces.

The third law states that the force between any two bodies is the same in the two directions. The earth attracts the sun just as much as the sun attracts the earth, so that $f_S = f_E$, or $m_S\alpha = m_E\alpha_E$. But the effects of the equal forces, that is, the accelerations, are not the same if the masses are unequal; the ratio of the accelerations is the inverse ratio of the masses affected.

7.20 The Law of Gravitation

By means of his laws of motion and by mathematical reasoning, Newton succeeded in reducing Kepler's geometrical description of the planetary system to a single comprehensive physical law. It will serve our present purpose simply to outline the sequence and chief results of Newton's inquiry.

By Kepler's first law, the path of a planet is an ellipse; it is continually curving. Consequently, the planet's motion is continually accelerated and, by the second law of motion, a force is always acting on the planet.

Since the planet moves, by Kepler's second law, so that the line joining it to the sun describes equal areas in equal times, it is easily proved that the force is directed toward the sun. Kepler had suspected that the sun had something to do with the planet's revolution, but he did not understand the connection. Later this second law becomes nothing more than a statement of the conservation of angular momentum.

Again from Kepler's first law, since the orbit is an ellipse with the sun at one focus, it can be proved that the force varies inversely as the square of the planet's distance from the sun. An elliptic orbit would also

result if the force varied directly as the distance, but in this event the sun would be at the center of the ellipse, not at one focus.

From Kepler's third law and the third law of motion, it can be shown that the attractive force between the sun and any planet varies directly as the product of their masses. In addition, Newton discovered that the moon's revolution is controlled by precisely the same force directed toward the earth. Although his experience did not extend beyond the solar system, he concluded that this force operates everywhere. These were the steps that led to the formulation of the *law of gravitation:*

Every particle of matter in the universe attracts every other particle with a force that varies directly as the product of their masses, and inversely as the square of the distance between them.

7.21 Examining the Law of Gravitation

The law of gravitation provides the key for the interpretation of celestial motions. It is therefore important to understand the meaning of the law. The statement is where f is the force, m_1 and m_2 are the masses of the two particles, and d is their distance apart.

$$f = \frac{Gm_1m_2}{d^2}$$

1. *The constant of gravitation, G,* is defined as the force of attraction between two unit masses at unit distance apart. If $m_1 = m_2 = 1$ gram, and d is 1 centimeter, then $G = f$. It is believed to be a universal constant, like the speed of light; but it is even more remarkable as a constant, for the speed of light is reduced by an interposing medium such as glass, whereas the force of gravitation is unaffected by anything placed between the attracting bodies.

   The value of this constant is best determined in the physical laboratory by the method first employed by the English scientist Henry Cavendish about 1798. It consists in measuring the attractions of metallic balls or cylinders. Heyl's determination at the National Bureau of Standards in 1942 gave the value $G = 6.673 \times 10^{-8}$ cm^3 sec^{-2} g^{-1}. Thus the attraction between gram masses 1 cm apart is only a 15-millionth of a dyne. Although it is very feeble between ordinary bodies, the gravitational force becomes important between the great masses of celestial bodies. The *Gaussian constant of gravitation* is much used in astronomical calculations; it is the acceleration produced by the sun's attraction at the earth's mean distance from the sun.
2. *The attraction of a sphere is toward its center,* as though the whole mass were concentrated there. Because of their rotations the celestial bodies are not spheres, but the flattening at their poles is often so small and the intervening spaces are so great that the distances between their centers may be used ordinarily in calculating their attractions. The attraction of a spheroid in the direction of its equator is greater

than that of a sphere of the same mass, and is smaller in the direction of its poles.

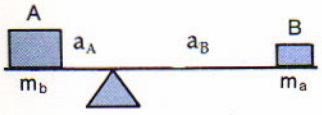
3. *The acceleration of the attracted body is independent of its mass.* If the force, f_1, on this body is replaced by the equivalent $m_1\alpha_1$ in the statement of the law of gravitation, the mass, m_1, cancels out, and the acceleration of the attracted body does not depend on its own mass. Galileo is said to have demonstrated this fact by dropping large and small weights from the leaning tower of Pisa. They fell together, thereby discrediting the traditional idea that heavy bodies fall faster than light ones.

$$\alpha_1 = \frac{Gm_2}{d^2}$$

The second, or attracting body, as we have chosen to consider it, is itself attracted and has the acceleration $\alpha_2 = Gm_1/d^2$ in the direction of the first. In Galileo's experiment this factor need not be taken into account; it becomes important when the two bodies have comparable masses. The acceleration of one body with respect to the other is the sum. Thus the relative acceleration of two bodies varies directly as the sum of their masses.

$$\alpha_1 + \alpha_2 = \frac{G(m_1 + m_2)}{d^2}.$$

4. *Two bodies, such as the earth and sun, mutually revolve around a common center between them.* Imagine the earth and sun joined by a stout rod. The point of support at which the two bodies would balance is the *center of mass* (or barycenter); it is the point around which they revolve in orbits of the same shape. If the masses of the two were equal, this point would be halfway between their centers. Because the sun's mass is 332,000 times as great as the earth's mass, the center of mass is not far from the sun's center. The relation is:



$$\frac{\text{Sun's center to center of mass}}{\text{Earth's center to center of mass}} = \frac{\text{earth's mass}}{\text{sun's mass}}.$$

The distance from the sun's center to the center of mass of the earth-sun system is therefore 1.496×10^8 km divided by 332,000 which equals 450 km, where we have taken the earth's values to actually be those of the earth-moon system and their barycenter.

7.22 Kepler's Third Law Restated

In its original form (7.15), Kepler's harmonic law gave a relation between the periods of revolution and the distances of the planets from the sun. As it is now derived from the law of gravitation, the relation involves the masses of the planets as well; it is as follows:

The squares of the periods of any two planets, each multiplied by the sum of the sun's mass and the planet's mass, are in the same proportion as the cubes of their mean distances from the sun.

Consider two planets, Mars and the earth. Let m represent the mass,

P the sidereal period of revolution of the planet, and a its mean distance from the sun. The revised harmonic law is in this case:

$$\frac{(m_S + m_M)P^2_{MS}}{(m_S + m_E)P^2_{ES}} = \frac{a^3_{MS}}{a^3_{ES}} .$$

The law in its original form was not far from correct, because the masses of all the planets are so small in comparison with the sun's mass that the ratio of the sums of the masses is nearly unity.

Let the units of mass, time, and distance in the above relation be respectively the sun's mass (neglecting the inconsiderable relative mass of the earth), the sidereal year, and the earth's mean distance from the sun. The denominators then disappear because their terms are all unity. Further, in the place of Mars and the sun take any mutually revolving bodies anywhere, denoting them by the subscripts 1 and 2. They may be the sun and a planet, a planet and its satellite, or a double star. In the more general form Kepler's third law becomes:

The sum of the masses of any two mutually revolving bodies, *in terms of the sun's mass,* equals the cube of their mean linear separation, *in astronomical units,* divided by the square of their period of revolution, *in years.*

$$m_1 + m_2 = \frac{a_{12}{}^3}{P_{12}{}^2} .$$

In this way the masses of the sun and of planets having satellites have been determined, the masses of the second bodies of the pairs being small enough in these cases to be neglected in comparison. The formula does not serve for the solitary planets, such as Mercury and Pluto, nor for the asteroids, the satellites themselves, the comets, and the meteor swarms. Their masses become known only in case they noticeably disturb the orbits of neighboring bodies.

7.23 The Relative Orbit of Two Bodies

We have noted (7.21) that two bodies, such as the earth and sun, mutually revolve around their center of mass, which is nearer the more massive body, so that the less massive component has the larger orbit. It can be shown (1) that the orbits are independent of any motion of the center of mass, that is, of the system as a whole; (2) that the individual orbits are the same in form, and that this is also the form of the *relative orbit* of one body with respect to the other. The relative orbit is often the only one that can be calculated; it is the one understood when one body is said to revolve around another.

Kepler's first law states that the orbits of the planets are ellipses. Newton proved that the orbit of a body revolving around another in accordance with the law of gravitation must be a conic, of which the ellipse is an example.

7.24 The Conics

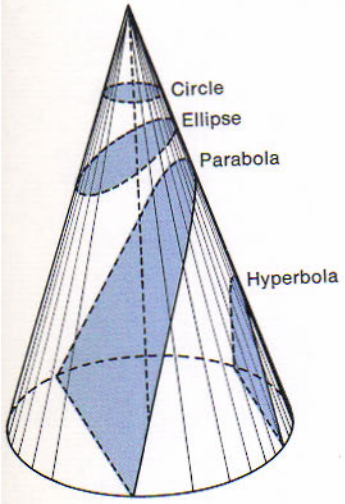


Figure 7.24
The conics.

The conics, or conic sections, are the ellipse, parabola, and hyperbola. They are sections cut from a circular cone, which for this purpose is a surface generated by one of two intersecting straight lines when it is turned around the other as an axis, the angle between them remaining the same.

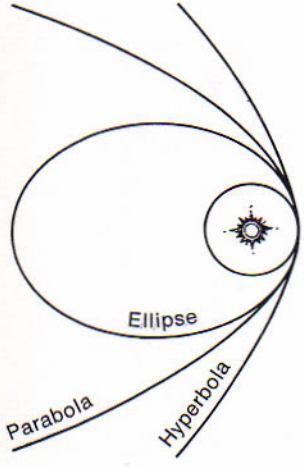
The ellipse (eccentricity 0 to nearly 1) is obtained when the cutting plane passes entirely through the cone, so that the section is closed. When the plane passes at right angles to the axis, the eccentricity of the ellipse is zero, and the section is a circle.

The parabola (eccentricity 1) results when the cutting plane is parallel to an element of the cone. This curve extends an indefinite distance, the two sides approaching parallelism. All parabolas, like all circles, have the same form but not the same size. The orbits of many comets are nearly parabolas.

The hyperbola (eccentricity greater than 1) is obtained when the cone is cut at a still smaller angle with the axis. It is an open curve like the parabola, but the directions of the two sides approach diverging straight lines. If a star passes another and is deflected by attraction from its original course, the orbit is hyperbolic.

7.25 Form of the Relative Orbit

$$V^2 = G(m_1 + m_2)\left(\frac{2}{r} - \frac{1}{a}\right),$$



The particular conic in which a celestial body revolves is determined by the central force and the velocity of the body in the orbit; for it is evident that the curvature of the orbit depends on the deflection of the body in the direction of its companion and the distance it has moved forward meanwhile in the orbit. This conclusion, among others, is obtained formally from the *equation of energy*, which is not derived here (see margin). where V is the velocity of revolution when the two bodies are at the distance r apart, and a is half the major axis of the resulting orbit.

It can be seen from this equation that the semimajor axis lengthens as the velocity is greater assuming a given periapsis distance. For a moderate speed, the orbit is an ellipse; for increasing speeds the length and eccentricity of the orbit grow greater, until a critical speed is reached at which the orbit becomes a parabola.

If the orbit is a circle, then $a = r$ in the above formula, so that V^2 is proportional to $1/r$. If the orbit is a parabola, a is infinite and V^2 is proportional to $2/r$. Therefore, if the speed of a body revolving in a circular orbit is multiplied by the square root of 2, or about 1.41, the orbit becomes a parabola. Because the earth's orbit is nearly circular, the *parabolic velocity* at our distance from the sun is the earth's velocity, 29.8 km a second, multiplied by 1.41, which equals 42 km a second. If its velocity should ever become as great as this value, the earth would depart from the sun's vicinity. Many comets and meteors, having their aphelion points far beyond the orbit of Neptune, cross the earth's orbit with speeds of this order.

7.26 The Elements of the Orbit

These are the specifications necessary to define it uniquely and to fix the place of the revolving body in the orbit at any time. The elements of the elliptical orbit of a planet, with their symbols, are the following:

1. *Inclination to ecliptic, i.* If the plane of the orbit is inclined to the ecliptic plane (i denotes the numerical value of the inclination), the line of their intersection is the *line of nodes,* which passes through the sun's position. The *ascending node* is the projection on the ecliptic, from the sun, of the point at which the planet crosses the ecliptic plane going from south to north.
2. *Longitude of the ascending node,* Ω. It is the celestial longitude of this node as seen from the sun, that is, the angle between the line of nodes and the direction of the vernal equinox. It fixes the orientation of the orbit plane, and, together with the inclination, defines this plane precisely.
3. *Angle from the ascending node to the perihelion point,* ω. It is measured from the ascending node along the orbit in the direction of the planet's motion, which must be specified; it gives the direction of the major axis of the orbit with respect to the line of nodes, and thus describes the orientation of the orbit in its plane.

   The foregoing are said to be the orientation elements.
4. *Semimajor axis, a.* This element, which is also known as the planet's *mean distance* from the sun, defines the size of the orbit and, very nearly, the period of revolution; for by Kepler's third law, P^2 is proportional to a^3 regardless of the shape of the ellipse.

   The semimajor axis is spoken of as the scale element.
5. *Eccentricity, e.* The eccentricity of the ellipse is the ratio c/a, where c is the sun's distance from the center of the ellipse (one half the distance between the foci). These five elements define the relative orbit uniquely.
6. *Time of passing perihelion, T.* This element and the value of the period of revolution permit the determination of the planet's position in the orbit at any time.

   If the orbit is circular, the longitude of perihelion drops out; if it is a parabola, the semimajor axis, which is then infinite, is replaced as an element by the *perihelion distance, q,* which defines the size of the parabola.

   The last two elements along with the period are referred to as the dynamical elements.

When the elements become known, the position of the planet or comet at any time can be computed; this, combined with the earth's position in its orbit at that time, gives finally the apparent place of the object as seen from the earth, its right ascension and declination. A tabulation

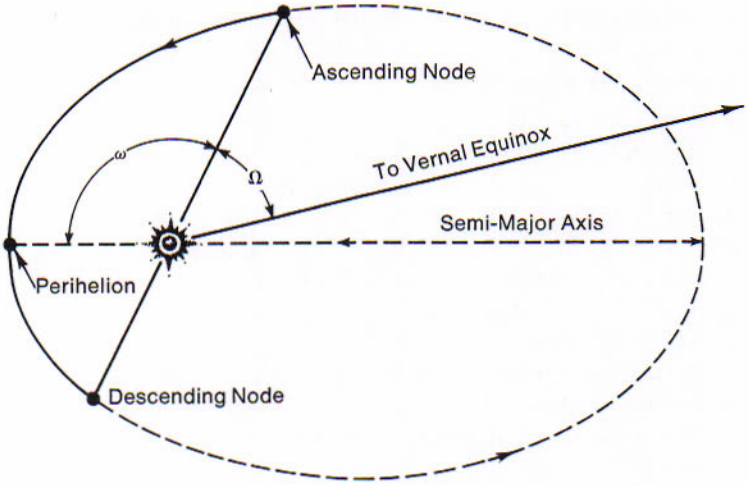


Figure 7.26
The orbit of a planet. The plane of the planet's orbit is inclined to the plane of the earth's orbit, represented by the plane of the paper.

of such places at regular intervals, often of a day, is an *ephemeris.* The astronomical almanacs give such tabulations for the sun, moon, principal planets, and certain asteroids for each year in advance.

7.27 Perturbations

Thus far we have dealt with the revolution of a body around the sun generally as though the body were acted on only by the sun's attraction. This is the *problem of two bodies,* which is solved directly and completely in terms of the law of gravitation. However, the body is subject to the attractions of other members of the solar system as well, so that it departs in a complex manner from simple elliptic motion. Thus we have in practice the *problem of three or more bodies,* the solution of which is more troublesome. It is fortunate for the orderly description of the planetary movements that the masses of the planets are small in comparison with the sun's mass and that their distances apart are very great. If it were not so, the mutual disturbances of the revolving bodies would introduce so much confusion that simple approximations, such as Kepler's laws, would have been impossible.

Because the sun's mass is dominant in the solar system, it is possible to derive at first the planet's orbit with reference to the sun alone and then to consider the departures from simple elliptic motion that are imposed by the attractions of other members of the system. *Perturbations* are the alterations so produced. As examples, the eccentricities and inclinations of planetary orbits fluctuate, perihelia advance, and nodes regress. All perturbations are oscillatory in the long run, so that they are not likely to alter permanently the general arrangement of the solar system.

7.28 Perturbations of Artificial Satellites

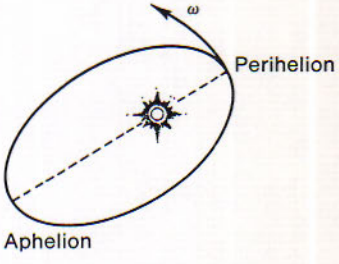
The orbits of artificial satellites revolving near the earth are strongly perturbed by the oblate earth, effects that are not considerably confused with perturbations by the sun. These orbits are also perturbed by resistance of the earth's atmosphere, giving information as to the density of the upper atmosphere at different elevations. As an example, consider the perturbations of satellite 1957-alpha (2.9) as reported by the Smithsonian Astrophysical Observatory from many observed positions of the rocket shell. The main perturbations of the orbit were:

1. *Regression of the nodes,* where the orbit crossed the plane of the earth's equator. The earth's equatorial bulge tended to pull the satellite's orbit into the equator plane, thus decreasing the original 65° inclination of the orbit. This tendency was resisted by the satellite's revolution. The result was a gyroscopic westward shifting of the orbit around the earth. The nodes regressed at the rate of 3°.1 a day, or three complete turns in about a year. For comparison (5.11) the nodes of the moon's path make a complete turn around the ecliptic in 18.6 years and the earth's precession takes almost 26,000 years.
2. *Advance of perigee.* The elliptical orbit of the satellite turned in its plane at the rate of 0°.4 a day, the perigee advancing in the direction of the revolution. The moon's perigee advances much more slowly.
3. *The effect of air resistance* was to make the satellite spiral toward the earth, revolving faster and in an orbit of decreasing eccentricity. L. G. Jacchia had drawn attention to semiregular fluctuations in the decrease of the revolution periods. These he attributes to corresponding variations in density of the earth's upper atmosphere produced by variability in intensity of short-wave solar radiations.

7.29 Relativity Effects

Events in the heavens and in the laboratory are usually represented practically as well by the formulas of either Newton or Einstein. A few exceptional cases are known where the predictions on the two theories differ significantly enough to be subject to the test of observation. These involve the presence of large masses or very high speeds. In such cases the observational evidence supports the theory of relativity.

The first test case was the rate of rotation of the major axis of Mercury's orbit. The perihelion point for this planet advances eastward at the rate of 573″ a century, whereas the rate predicted by the Newtonian theory on the basis of the attractions of other planets is 43″ less. Albert Einstein explained, in 1915, that the whole advance is predicted by the relativity theory; and smaller similar advances in the orbits of Venus and the earth are now known to be more accurately predicted by this theory.



A challenge to this simple explanation has been advanced by R. H.

$$\Delta\omega = \frac{3.84}{R^{5/2}(1-e)^2}$$

R(in A.U.)
$\Delta\omega$(per 100 years)

Dicke who has shown that if the core of the sun is oblate and rotating as a solid body it would explain the excess perihelion advance equally well. Special experiments on board spacecraft orbiting the sun can settle this problem one way or another.

$$\Delta\theta = \frac{1''.75}{n}$$

n in the units of the sun's radius

At this point it is probably useful to point out that the two other "classical" astronomical test cases for relativistic effects are under challenge as well. The apparent outward displacement of stars from the sun's place in the sky (6.8) rests upon the success of an exceedingly difficult experiment that has been, so far, far from convincing, although this impression has been somewhat changed by a study of the positions of two radio sources as they passed near the sun (the interferometric measurements are too new for inclusion here). The redward displacement of the lines in the spectra of a very dense star (11.30) due to the gravitational retardation of the outward moving photon is also being challenged because of the difficulty of measuring a real displacement in diffuse lines. At the surface of the sun the displacement amounts to 1 Å in the visible region of the spectrum.

$$\frac{d\lambda}{\lambda} \propto \frac{M}{R}$$

There are however, relativistic effects that are quite apparent, the existence of *synchrotron radiation* and its spectral changes as the velocities of the particles approach the speed of light, for one, and the *red shifts of galaxies* (18.25) that are greater than one, for another. Still another is the very precise *relativistic shift* in the *Mössbauer effect.* What we are trying to point out is not that one theory is better than another, but that the theory of relativity, which does a little better than the older Newtonian mechanics, may now be supplanted by an even more sophisticated theory.

$$\frac{d\lambda}{\lambda} = \frac{6M}{C^2R}$$

REVIEW QUESTIONS

1. Give three ways to distinguish the planets from the stars.
2. How do the planets most naturally divide: for kinematical studies? for physical studies?
3. Give the principle regularities of the solar system and the outstanding exceptions to these.
4. What is meant by the terms: synodic period, sidereal period?
5. How do the phases of the inferior planets disprove the Ptolemaic system?
6. Why was Tycho unable to prove the Copernican theory?
7. Weight is the local force of attraction on one body by another. The density of Uranus and Jupiter is about the same, but Jupiter has a diameter 3 times that of Uranus. What is the ratio of weights on Jupiter compared to Uranus?
8. While it is difficult to rank the achievements of men, would you consider Tycho's work more important than Kepler's?
9. What are the three classes of orbital elements?
10. In two general classes of effects the relativity theory gives results different from Newtonian Mechanics. What are these two general classes?

REFERENCES

Blanco, Victor M. and S. W. McCusky, *Basic Physics of the Solar System,* Reading, Mass.: Addison-Wesley Co., 1961

Glasstone, S., *Sourcebook on the Space Sciences,* Princeton: D. van Nostrand, Co. 1965.

Page, Thornton and L. W. Page, *Wanderers in the Sky,* New York, The Macmillan Co., 1965

Tucker, R. A. R., *The Paths of the Planets,* New York: American Elsevier, 1967

Smart, W. M., *Text-Book on Spherical Astronomy,* N.Y.: Cambridge Univ. Press, 5th edition, 1962

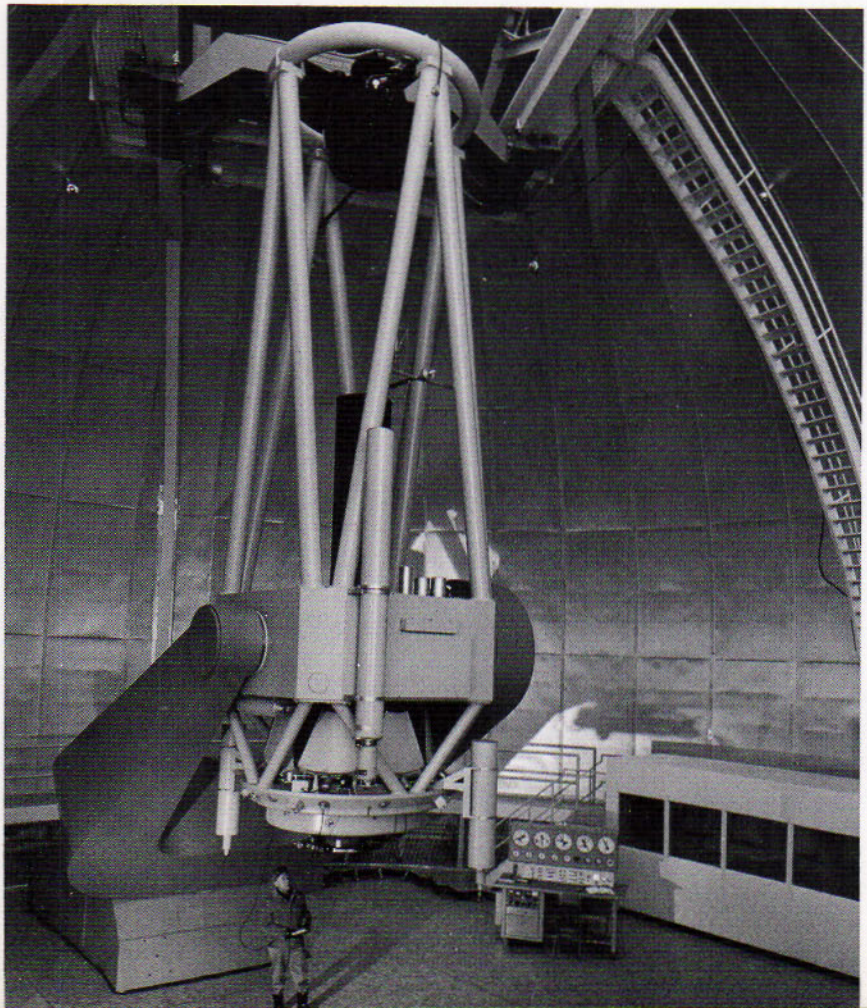
FOR FURTHER STUDY

Brouwer, Dirk and G. M. Clemence, *Methods of Celestial Mechanics,* Academic Press, 1961

Dziobek, Otto, *Mathematical Theories of Planetary Motions,* Dover, New York, 1962

McClusky, S. W., *Introduction to Celestial Mechanics,* Addison-Wesley, Reading, Massachusetts, 1963

Szebehely, Victor, *Theory of Orbits,* New York: Academic Press, 1967



The U.S. Naval Observatory's modern 61" astrometric reflector; an example of a modern telescope designed for a specific purpose. *(Official U.S. Navy photograph)*


8

PLANETS AND THEIR SATELLITES

MERCURY — VENUS — MARS, THE RED PLANET — THE ASTEROIDS — JUPITER, THE GIANT PLANET — SATURN AND ITS RINGS — URANUS AND NEPTUNE — PLUTO — TABLES OF THE PLANETS AND SATELLITES

Several centuries of patient study have pieced together the story of the planets. This study began with Galileo Galilei in 1609 with the application of the telescope to the moon, Venus, Saturn and Jupiter and has literally exploded during the past 10 years with the application of radio, radar and unmanned spacecraft to planetary study. Detailed knowledge of the planets provides a severe test for our theories on the origin of the solar system.

Mars' south polar cap region, from Mariner VI. *(NASA photograph)*

MERCURY

Mercury is the smallest principal planet; its diameter, 4640 kilometers, does not greatly exceed the moon's diameter. Its surface features seem to resemble those of the moon rather than the earth. The nearest planet to the sun, Mercury revolves around the sun once in 88 days and rotates on its axis in 59 days.

8.1 As Evening and Morning Star

Mercury is occasionally visible to the naked eye for a few days near the times of its greatest elongations, which occur about 22 days before and after the inferior conjunctions with the sun. Because the synodic period is only 116 days, three eastern and as many western elongations may come in the course of a year. They are not equally favorable for two reasons. (1) Mercury's apparent distance from the sun at greatest elongations ranges from 28° when the planet is also at aphelion to only 18° at perihelion. (2) Because the planet is always near the ecliptic, it is highest in the sky at sunrise or sunset when the ecliptic is most inclined to the horizon. This condition is fulfilled for us in middle northern latitudes (1.19) when the vernal equinox is setting and the autumnal equinox is rising.

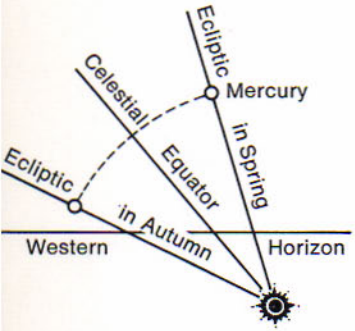


Figure 8.1
Mercury is most conspicuous as evening star near its greatest elongations in the spring.

For the second reason, Mercury is most likely to be visible as an evening star near its greatest eastern elongations that occur in March and April, and as a morning star near its greatest western elongations in September and October. It then appears in the twilight near the horizon, at times even a little brighter than Sirius, and twinkling like a star because of its small size and low altitude.

The terms *evening star* and *morning star* refer most often to appearances of inferior planets, particularly Venus, in the west after sunset and in the east before sunrise. The terms are applied as well to superior planets that are visible in the evening and morning sky, respectively.

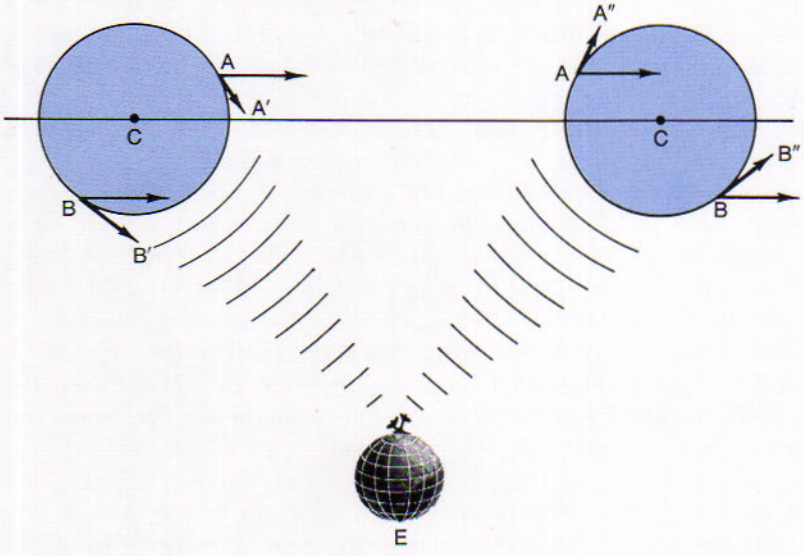
8.2 Viewed with the Telescope

Mercury shows phases, as an inferior planet should do (7.7). The phase is full at superior conjunction, quarter near greatest elongation, and new at inferior conjunction. The best views are likely to be obtained in the daytime, when the planet can be observed at higher altitudes. Faint dark blotches on the small disk, which may resemble the lunar seas, have been discerned by experienced observers and are recorded in some photographs. The conclusion, from these observations, that the period of Mercury's rotation is 88 days has proven to be erroneous. Radar studies have shown that the true rotation period is very nearly 59 days.

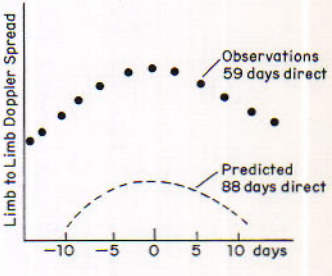
The radar technique consists of transmitting and receiving a series of radio pulses at the planet and measuring the spread in frequency in the reflected pulses. This spread in frequency is caused by the Doppler effect and arises from two sources, the true rotation (ω_1) of the planet and an apparent rotation (ω_2) introduced by a perspective effect. The instantaneous rotation (ω) giving rise to the frequency spread is the sum of the two rotations. When pulses of a precise frequency strike a rotating body the receding limb shifts the radiation to a lower frequency when it strikes and is reflected and the converse for the other limb.

$$\omega = \omega_1 + \omega_2$$

Figure 8.2
Apparent forward rotation of an inferior planet as it passes the earth.



The apparent rotation can best be seen with the aid of Figure 8.2. When the limb components are resolved into their radial components one limb appears to be approaching (or receding) faster than the other is approaching (or receding). This apparent difference in velocity gives rise to a spread caused by the Doppler effect and is the key to determining the sense of the planet's true rotation. For inferior planets the sense of ω_2 is always that of direct rotation and tends to zero at the planets' elongations. Thus, if ω_1 and ω_2 have the same sense, that is, if ω_1 is also direct, the sum (ω) will have a maximum as the planet goes through inferior conjunction. This is the case with Mercury. If the rotation of the planet would be retrograde, ω_1 would be opposite in sign to ω_2 and ω would go through a minimum at inferior conjunction.



Mercury Radar Rotation Curve

8.3 Mercury Resembles the Moon

This planet does not greatly exceed the moon in diameter and mass and, therefore, in surface gravity. Its low velocity of escape suggests scarcely better success than the moon has had in retaining an atmosphere. Its reflecting power is about as low as the moon's, and this small efficiency as a mirror probably has the same cause, namely, the reflection of sunlight from a rough surface having no atmosphere around it. That the planet is at least as mountainous as the moon is shown by the similar great increase in its brightness as the shadows shorten between the quarter and full phases. It is even probable that its surface is also heavily cratered.

Since Mercury has very little or no atmosphere, its surface is subjected to a much more intense heating by the sun. Early infrared measures by E. Pettit and S. B. Nicholson gave an average bright side temperature of about 610°K. Their efforts to delineate a temperature for the dark side were unsuccessful, and a temperature near absolute zero has long been assumed. Several investigations cast doubt upon this assumption, but none were convincing enough until K. I. Kellermann obtained a dark side temperature of 290°K at λ 11.3 cm by measuring from full phase to crescent phase. At the time, this result led to considerable theorizing centered on a possible atmosphere, but as we have just seen, it can now be explained by Mercury's rotation. Also, the 11.3 cm measurement integrates the disk temperature at an effective depth of about 11 cm below the surface. Later, M. A. Kaftan-Kassim and Kellermann obtained a temperature of about 210°K working at λ 1.9 cm. Further, their plot of phase against temperature showed a minimum prior to inferior conjunction which is consistant with forward rotation of the body. This result anticipated the radar results discussed previously.

8.4 Transits of Mercury and Venus

The inferior planets usually pass north or south of the sun at inferior conjunction. Occasionally they *transit,* or cross directly in front of the sun, when they appear as dark dots against its disk. The additional condition necessary for a transit is similar to the requirement for a solar or lunar eclipse; it is that the sun must be near the line of nodes of the planet's orbit.

The sun passes the intersections of Mercury's path with the ecliptic on May 8 and November 10. Transits are possible only within 3 days of the former date and within 5 days of the latter. This difference in the limits, which is caused by the eccentricity of the planet's orbit, makes the November transits twice as numerous as those in May.

About 13 transits of Mercury occur in the course of a century, the latest one, on May 7, 1970. Transits scheduled for the remainder of the century are given on Fig 8.4B and occur on November 10, 1973, November 13, 1986, November 6, 1993, and November 15, 1999, which will be a

Figure 8.4A
Transit of Mercury on 9 May 1970. This excellent photograph in Hα light shows Mercury on the left central portion of the sun. Accurate measurements of sunspot diameters can be made on such occasions. *(Courtesy of H. Caulk and R. W. Hobbs, NASA Goddard Space Flight Center)*

grazing transit. Transits of Mercury can be timed rather accurately and have been useful for improving our knowledge of the planet's motions. They cannot be viewed without a telescope.

Transits of Venus are possible only within about 2 days before or after June 7 and December 9, the dates when the sun passes the nodes of the planet's path. They are less frequent because the limits are narrower and also because conjunctions come less often. Transits of Venus now come in pairs having a separation of 8 years. The latest pair of transits occurred in 1874 and 1882; the next pair is due on June 8, 2004, and June 6, 2012. After a while there will be a long period when they occur singly. These transits are visible without a telescope, either by projection or by using a blackened piece of film.

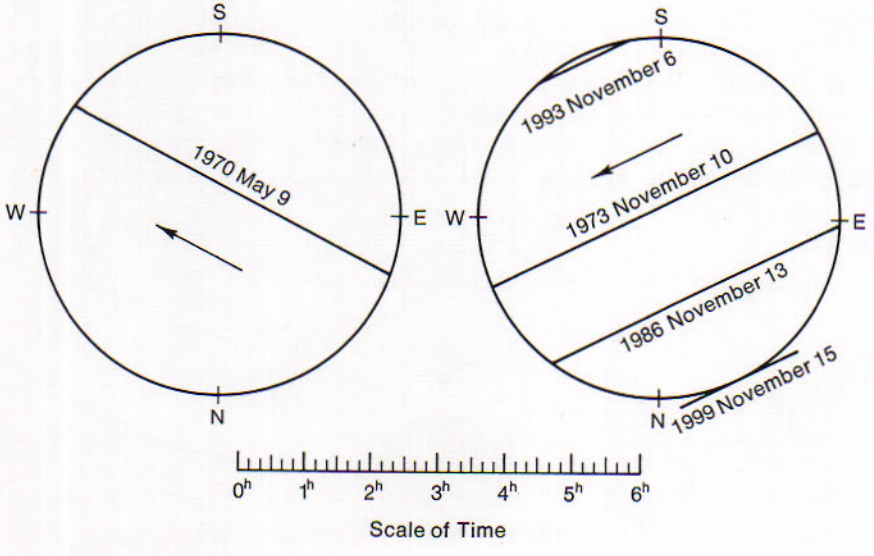


Figure 8.4B
Transits of Mercury, 1970–1999.

VENUS

Venus, the familiar evening and morning star, is the brightest planet. It outshines all the other celestial bodies except the sun and moon, and near the times of greatest brilliancy it is plainly visible to the naked eye at midday, when the attention is directed to it. The second in order from the sun, this planet revolves next within the earth's orbit at the mean distance of 108.9 million kilometers from the sun, completing its revolution once in 225 days. Its orbit is the most nearly circular among the principal planets. Although Venus resembles the earth in size, mass, and distance from the sun, it is quite dissimilar in its higher surface temperature, in the scarcity of free oxygen in its atmosphere, and is different from all of the major planets in that its rotation is retrograde.

8.5 As Evening and Morning Star

Because the orbit of Venus is within the earth's orbit and nearly in the same plane with it, this planet, like Mercury, appears to oscillate to the east and west of the sun's position. At superior conjunction its distance from the earth averages 260 million kilometers, or the sum of the earth's and its own distance from the sun. From this position Venus emerges slowly to the east of the sun as evening star; it comes out a little higher from night to night and sets a little later after sunset, until it reaches greatest eastern elongation 220 days after the time of superior conjunction.

The entire westward movement to greatest western elongation is accomplished in 144 days. Midway, the planet passes nearly between the sun and the earth into the morning sky. At inferior conjunction it averages only 41 million kilometers from the earth, or the difference between the earth's and its own distance from the sun. This is the closest approach of any principal planet, although some minor planets come at times still closer to the earth. Turning eastward again after greatest western elon-

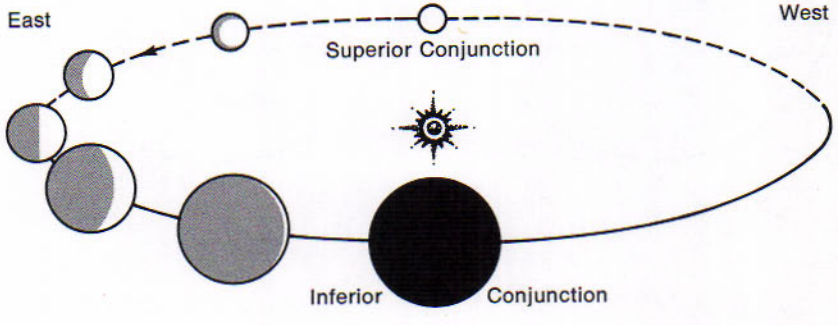


Figure 8.5
Changing phase and apparent size of Venus.

gation, Venus moves slowly back to superior conjunction, again requiring 220 days for this part of the journey. The synodic period is accordingly 584 days.

Greatest brilliancy as evening and morning star occurs about 36 days before and after the time of inferior conjunction. On these occasions Venus is 6 times as bright as the planet Jupiter and 15 times as bright as Sirius, the brightest star.

8.6 Through the Telescope; the Phases

As a visual object with the telescope the conspicuous feature of Venus is its phases, first seen by Galileo, in 1610. The phase is full at superior conjunction, quarter at greatest elongation, and new at inferior conjunction; but a thin extended crescent usually remains at the last-named aspect, because as a rule the planet crosses a little above or below the sun.

TABLE 8.I Configuration of Venus, 1970–1980

SUPERIOR CONJUNCTION	GREATEST ELONGATION EAST	INFERIOR CONJUNCTION	GREATEST ELONGATION WEST
1970 Jan. 25	1970 Sep. 1	1970 Nov. 10	1971 Jan. 20
1971 Aug. 27	1972 Apr. 8	1972 June 17	1972 Aug. 27
1973 Apr. 9	1973 Nov. 13	1974 Jan. 23	1974 Apr. 4
1974 Nov. 6	1975 June 18	1975 Aug. 27	1975 Nov. 7
1976 June 18	1977 Jan. 24	1977 Apr. 6	1977 June 15
1978 Jan. 22	1978 Aug. 29	1978 Nov. 7	1979 Jan. 18
1979 Aug. 24	1980 Apr. 5	1980 June 15	1980 Aug. 24

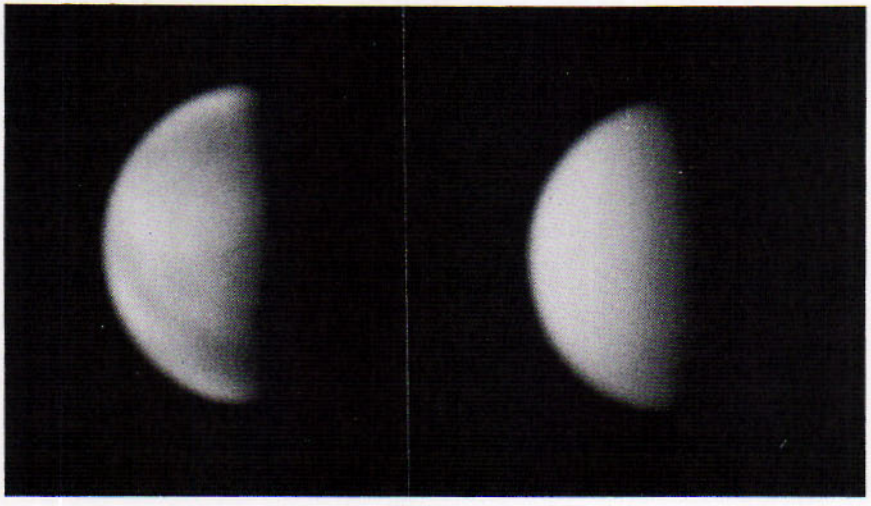


Figure 8.6
Venus near the quarter phase. In violet light (left) and red light (right). *(Photographs, April 8, 1950, by Gerard P. Kuiper, McDonald Observatory)*

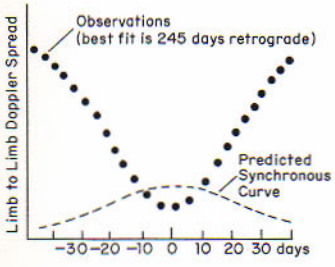
Unlike the moon, which is brightest at the full phase, Venus attains greatest brilliancy in our skies when its phase resembles that of the moon 2 days before the first quarter. At the full phase the planet's apparent diameter is 10″; at the new phase it is more than 6 times as great, because the distance from the earth is then reduced to about one sixth the former value. The increasing apparent size more than offsets the diminishing fraction of the disk in the sunlight, until the crescent phase is reached. At greatest brilliancy the crescent sends us 2½ times as much light as does the smaller, fully illuminated disk.

The physical diameter has been measured in many ways: micrometer measures, timing of transits, and radar. The micrometer measurements are difficult because it is difficult to delineate the true limb of the planet. The transit measures yield a radius of 6065 km, while the best radar value is 6,050 km. Because of the directness of the radar technique, we prefer the latter value.

8.7 The Atmosphere and Surface

The planet's surface is totally obscured by clouds although occasional streaks in the clouds can be seen through the telescope. As a result the rotation period of Venus can not be directly determined. V. M. Slipher, on the basis of his spectroscopic observations, maintained that its rotation was very slow and could even be retrograde, but others set the period from 30 days down to a day in the direct sense. Radar studies have settled this problem following the method explained in 8.2.

Radar studies give a period of 243 days, thus the rotation is retrograde. These results had been anticipated from temperature studies by F. D. Drake whose work at λ 10 cm showed the dark side temperature reached a minimum of about 580°K after inferior conjunction. This was confirmed by C. H. Mayer and his colleagues working at λ 3 cm giving a dark side temperature of 550°K. The spacecraft Venera 4 made a direct measurement on the ground, which gave 544°K. Thus we must conclude that any water on Venus is in the vapor form.



Venus Radar Rotation Curve

Fortunately, the same spacecraft sampled the primary constituents and water proves to be very scarce, certainly less than 1% (only 10^{-4} that of the earth) in agreement with the microwave studies mentioned above since even a small amount of water vapor would absorb the radiation. Carbon dioxide, as anticipated from ground based studies makes up 90% of the Cyrinthean atmosphere, nitrogen 7% and oxygen about 1%.

Venera 4 and Mariner V reveal that there is a very small or no magnetic field. Thus energetic particles from the sun and intergalactic space freely penetrate the atmosphere. This gives rise to a very thin ionosphere which was measured by these two spacecraft and by Mariner II several years

earlier. From these studies we conclude that the surface and atmosphere are not very hospitable to terrestrial beings.

There is a discrepancy between the two spacecraft concerning the surface temperature and pressure. An analysis of simultaneous radar and Mariner V data yields a surface temperature of 700°K and a surface pressure of 100 atmospheres, in contrast to the Venera 4 values of 544°K and 22 atmospheres. The discrepancy can only be reconciled by assuming that Venera 4 deployed its lander at twice the programmed altitude. If this is so, then there is excellent agreement with the temperatures deduced from the Mariner II data.

Precision studies of the radar scattering of Venus done by R. M. Goldstein and R. L. Carpenter revealed surface features confirming the retrograde rotation and giving a very accurate rotation period. Venus has several large areas that give a high backscattering and are thus not smooth, even probably highlands. The survey shows numerous small bright spots along with broad areas of low backscattering. The latter, dark to radar, must be relatively smooth, undulating surfaces, not unlike the moon's *maria* which present low backscattering to radar also.

MARS, THE RED PLANET

Mars is next in order beyond the earth. This red planet revolves around the sun once in 687 days and rotates on its axis once in $24^h\,37^m$. Its diameter is about 6760 Km, or slightly more than half the earth's diameter. Viewed with the telescope its surface exhibits a variety of markings. The persistent idea that Mars contains certain forms of life has made this planet an object of special interest, particularly at its closest approaches to the earth.

8.8 Oppositions of Mars

As a superior planet, Mars is best situated for observation when it is opposite the sun's place in the sky; it is then nearer us than usual and is visible through most of the night. Because of the considerable eccentricity, 0.09, of its orbit, Mars varies greatly in its distance from the earth at the different oppositions. The distance exceeds 97 million kilometers when the planet is near its aphelion, and may be slightly less than 56 million kilometers at the *favorable oppositions*, when it is also near its

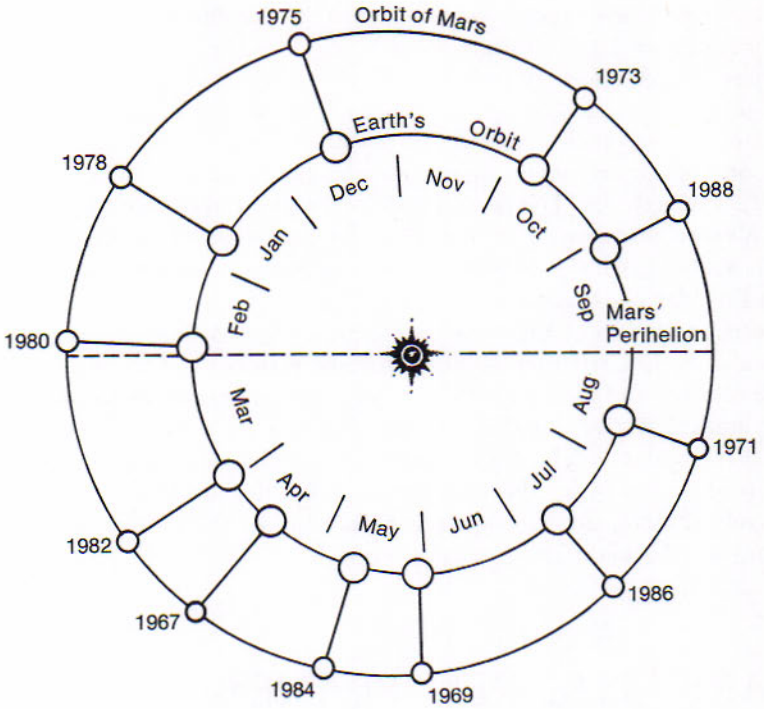


Figure 8.8
Varying distances of Mars from the earth at oppositions between 1954 and 1971, showing the favorable oppositions of 1956 and 1971.

perihelion. Oppositions of Mars recur at intervals of the synodic period, which averages 780 days, or about 50 days longer than 2 years. Thus they come in alternate years and each time about 50 days later than before.

Favorable oppositions occur at intervals of 15 or 17 years, usually in August or September, because on August 28 the earth has the same heliocentric longitude as the perihelion of Mars. On these favorable occasions Mars appears brighter in our skies than any other planet except Venus. It may then attain an apparent diameter of 25″, so that a magnification of only 75 times makes it appear with the telescope as large as the moon does to the unaided eye. Much of our knowledge of Mars has been gained around the times of the favorable oppositions. The dates of oppositions of Mars from the favorable opposition of 1971 to 1999 are given in Table 8.II. After the opposition of 1971, a pair about equally favorable will occur in 1986 and 1988.

TABLE 8.II
Oppositions of Mars 1971–1999

YEAR	OPPOSITION	NEAREST EARTH	MILLIONS OF KM	DIAMETER	MAGNITUDE
1971	Aug. 10	Aug. 12	56.3	24".9	−2.6
1973	Oct. 24	Oct. 17	65.2	21.4	−2.2
1975	Dec. 15	Dec. 9	84.5	16.5	−1.5
1978	Jan. 22	Jan. 19	97.7	14.3	−1.0
1980	Feb. 25	Feb. 26	100.7	13.8	−0.9
1982	March 31	April 5	95.1	14.7	−1.1
1984	May 11	May 19	79.7	17.6	−1.6
1986	July 10	July 16	60.5	23.2	−2.4
1988	Sept. 28	Sept. 22	58.9	23.8	−2.5
1990	Nov. 27	Nov. 20	77.4	18.1	−1.7
1993	Jan. 8	Jan. 3	93.7	14.9	−1.2
1995	Feb. 12	Feb. 11	101.2	13.8	−0.9
1997	March 17	March 20	98.6	14.2	−1.0
1999	April 29	May 1	86.6	16.2	−1.4

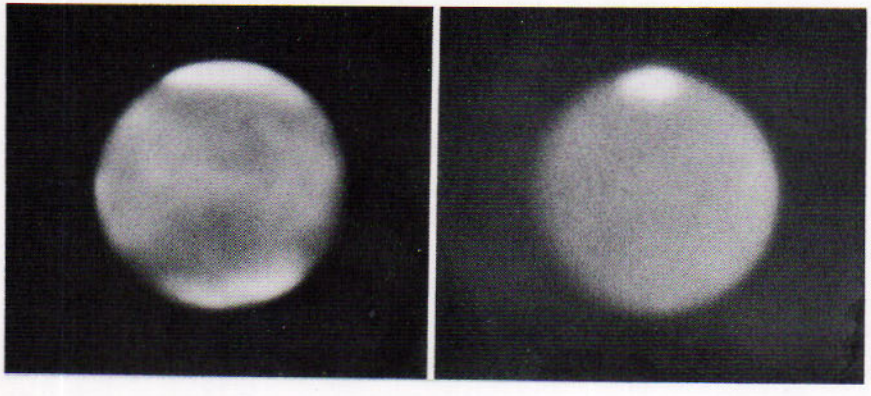
Abstracted from material furnished by Dr. R. Duncombe, U. S. Naval Observatory

8.9 Mars Through the Telescope

Viewed with the telescope in ordinary conditions, Mars is likely to be disappointing even when it is nearest us. The finer markings of its surface are often blurred by turbulence of our air and are also frequently dimmed by haze in the atmosphere of the planet itself. On rare occasions, when our air is unusually steady and the planet's atmosphere has cleared for a time, the surface features of Mars become surprisingly distinct with telescopes of only moderate size. (Plate II)

The white polar caps are the most conspicuous visual features; they are areas covered with snow to a thickness perhaps not exceeding several

Figure 8.9A
Seasonal changes in the south polar cap of Mars, observed in spring (left) and in summer (right). The pictures were taken on July 4, 1954 and August 23, 1956 respectively. *(Lowell Observatory photographs)*



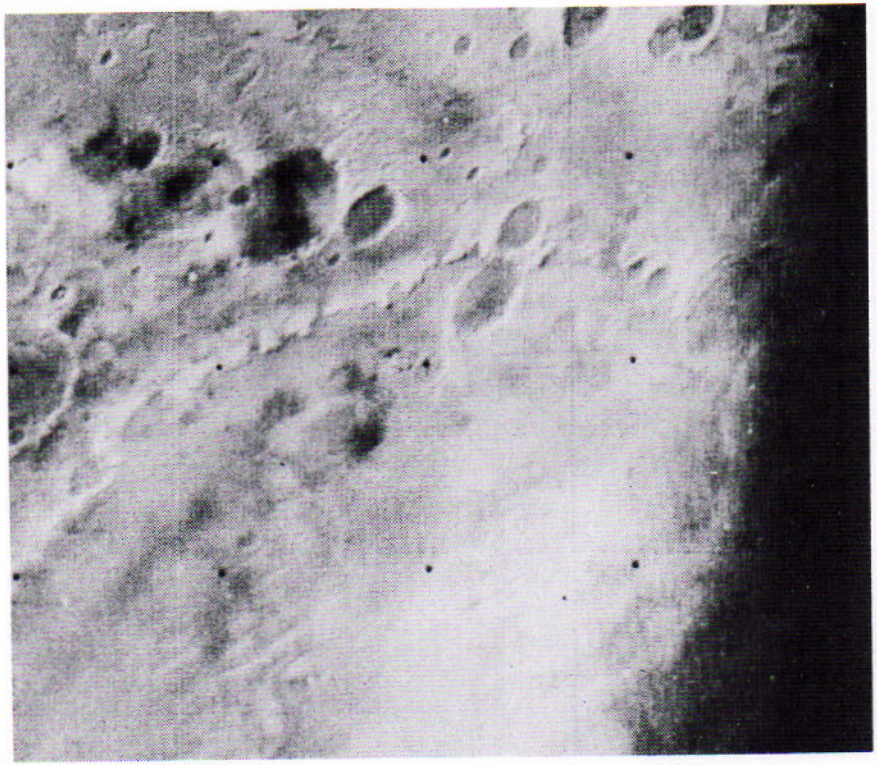


Figure 8.9B
Mars' south polar cap region, from Mariner VII. *(NASA photograph)*

centimeters on the average although drifts several meters deep occur. The snow appears to be of carbon dioxide instead of water.

Each snow cap is deposited in a large area around the pole during the winter season of its hemisphere. It shrinks with the approach of summer, and sometimes in its retreat toward the pole it leaves behind one or more white spots isolated for a time as though on the summit or colder slope of a hill. A cap normally forms under an atmospheric veil, but its disappearance can usually be followed without obstruction by haze or clouds.

Three fifths of the surface of Mars has a reddish hue that accounts for the ruddy glow of the planet in our skies. These brighter areas are desert expanses, the source of dust storms that obscure the darker markings. From their effect on the sunlight they reflect, the red areas are identified by Dollfus with pulverized limonite, the hydrated ferrous oxide.

The dark areas, originally supposed to be water areas, have been named seas, lakes, bogs, canals, and so on; and these designations have survived, like the lunar "seas."

The nomenclature of Mars was revised by the International Astronomical Union in 1958. The number of proper names for the large regions is reduced to 128, which are generally the same as before. Small details are designated in the revision only by their approximate Martian longitudes and latitudes. The official list of names and the maps for identifying the various features are shown in *Sky and Telescope* for November 1958.

One of the most controversial aspects of telescopic studies of Mars has been the existence of "canals." These were first reported in 1877 by G. V. Schiaparelli and have been seen and studied by many observers since, but especially by Percival Lowell. Recent observations (8.12) indicate that the markings observed must be related more to climatic conditions than to linear markings. Since their visibility correlates with the seasonal variations they may actually delineate mountain ranges whose relief is enhanced by a thin deposit of hoar frost on one flank of the range.

8.10 Rotation of Mars

The period of rotation is about $24^h\ 37^m$. The rotation of Mars has the same direction as the earth's rotation, and its period so nearly equals the earth's period that at the same hour from day to day almost the same face of the planet is presented to us, except that everything has stepped backward 10°. Thus the markings pass slowly in review, completing their apparent backward turning in about 38 days.

The inclination of the planet's equator to the plane of its orbit is nearly the same as the angle between the earth's equator and the ecliptic plane. The orientation of the axis differs at present about 90° from that of the earth; its northern end is directed toward the neighborhood of the star Alpha Cephei, not far from the position our own north celestial pole will have 6000 years hence.

8.11 The Seasons of Mars

Mars presents its poles alternately to the sun in the same way that the earth does, because of the similarity of its axial tilt. The seasons resemble ours geometrically, although they are nearly twice as long as ours. The winter solstice of Mars occurs when the planet has the same heliocentric longitude that the earth has about September 10 and not long after the time of its perihelion, when Mars has the same direction from the sun that the earth has on August 28. Summer in its southern hemisphere is therefore warmer than the northern summer, which comes when the planet is near the aphelion point in its orbit, and its southern winters are colder than the northern ones. For the same reason, the earth's southern hemisphere would have the greater seasonal range in temperature if there were no compensating factor (2.33).

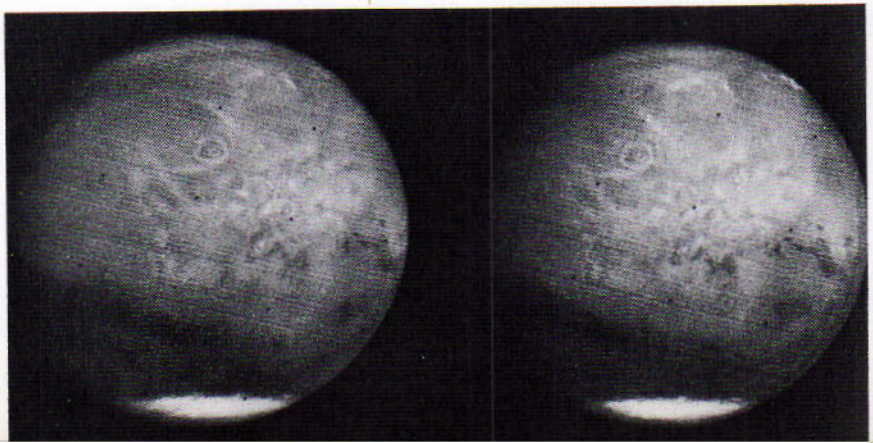
Although the whole variation in our distance from the sun is only 3 per cent, Mars in its more eccentric orbit is 20 per cent, or more than 42 million kilometers, farther from the sun at aphelion than at perihelion. The seasonal difference in the two hemispheres is noticeable. The south polar cap attains an area of the order of ten million square km and the northern cap about 8 million. It is the south polar cap that is toward the earth at the favorable oppositions, and this is accordingly the one that more often appears in photographs. Changes in features other than the polar caps are observed as well.

With the shrinking of each polar cap the region around the cap darkens, and the darkening gradually extends as far as the equator. The dark areas become more distinct and some of them take on a greenish hue. As summer progresses in that hemisphere, the markings fade. The times of the Martian year when the dark markings change in intensity and color are such as would be expected if the changes are caused by the growth and decline of vegetation. Kuiper's studies of sunlight reflected from the green areas seem to rule out familiar seed plants and ferns. They might be compatible with the presence of something like our lichens and hardy mosses, appearing sporadically in lava basins otherwise like those of the lunar maria.

8.12 Mar's Surface

The remarkable spacecraft Mariner IV after a historic journey, tape recorded 11 pairs of close up photographs as it swung past the planet and then transmitted them back to earth. The photographs revealed a surface heavily cratered like the surface of the moon. R. Baldwin in the 1950's hypothesized that this should be the case. Counts of craters as a function of size are in close agreement with similar counts on the moon. Also, many craters show a characteristic central peak which leads one to suspect that the origin is identical on both bodies. Age estimates based on the photographs give a value of approximately 3×10^9 years.

Figure 8.12
Mars from the approaching Mariner VII, showing the planet's features and perceptible rotation. Time interval between the two exposures: 42 minutes. *(NASA photograph)*



Mars, like Venus, is a dry planet essentially devoid of life. The notion persists that elementary living organisms thrive on Mars and in its atmosphere. Mariner IV failed to observe a magnetic field ($<3 \times 10^{-4}$ that of the earth) and hence protection for these primitive organisms from ultraviolet and particle radiation is lacking.

An even more exciting feat was performed by the spacecraft Mariner VI and VII. The two spacecraft were launched within two weeks of each other and performed flawlessly in achieving their goal of photographing Mars from a distance and close up. Their photographs (Figure 8.12) confirmed the Mariner IV results, provided great detail of the south polar cap (Fig. 8.9B) and removed all doubt that non-accidental linear markings exist on Mars. The canali had to be mental constructs due to the poor resolution of earth-based telescopes. Large craters and even concentrically ringed craters of the Mare Oriental type were photographed.

8.13 Mar's Atmosphere

The hazy atmosphere of Mars itself affects the view of the planet's surface. The haze is sometimes so thick that the surface markings become indistinct even in the red photographs. Kuiper concludes that the haze is caused by ice crystals and resembles the cirrus clouds of our own atmosphere. The haze is partly dissipated where the air of the planet is warmer; it is condensed into visible clouds where the air is colder. Thus the blue photographs show white clouds over the polar regions, and they also reveal small clouds forming near the sunset line and disintegrating near the sunrise line. Clouds of yellow dust from the deserts sometimes add to the indistinctness of the view.

These yellow and white clouds form at extremely high altitudes which makes them quite visible. Mars' low surface gravity is responsible for this. The surface pressure on Mars is only about 1.5% that on the earth, but at an altitude of 50 kilometers the density of Mars' atmosphere exceeds that of the earth's and consequently supports clouds at very high altitudes. By the same token, Mars' ionosphere, caused by solar radiation, must be proportionately higher than that of the earth and this has been confirmed by Mariner spacecraft observations.

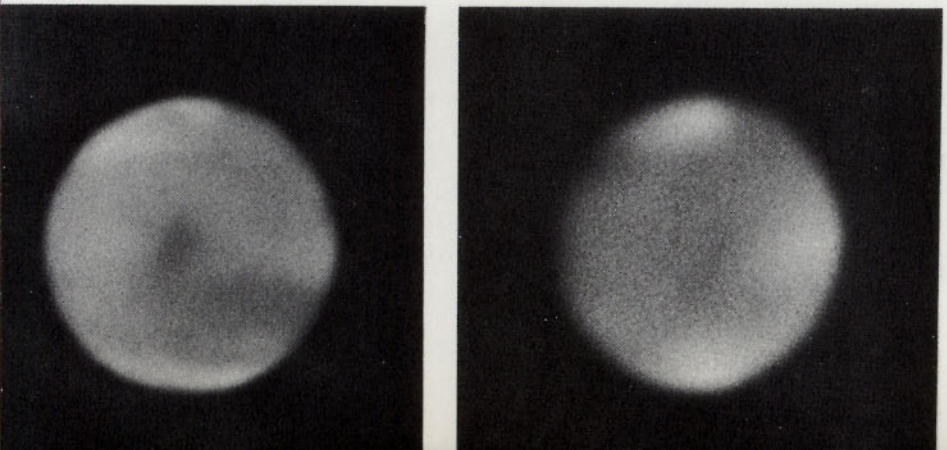


Figure 8.13
Mars through its hazy and clearer atmospheres. Photographed on May 28 and June 21, 1969. *(Lowell Observatory photograph)*

At times the haze clears rapidly over a large area of the planet; it cleared near the oppositions of 1909, 1924, and 1941, but disappointed the observers by failing to do so at the otherwise favorable opposition of 1956. The 1969 photographs in Fig. 8.13 were both taken in yellow light only 24 days apart. We note the cloud above the polar cap in the veiled view at the left and how much more distinctly the Syrtis Major and other surface markings appear through the clearer atmosphere in the view at the right.

8.14 The Climate of Mars

The atmosphere of Mars is considerably rarer than ours. The low reflecting power, 0.15, of the planet and the relative distinctness of the surface markings lead to this conclusion, while the low velocity of escape, 5 km a second, gives a reason for it. The chief constituent of the atmosphere is probably nitrogen, which is not expected to show in the spectrum. Kuiper finds that carbon dioxide is twice as abundant as in our atmosphere. He estimates that free oxygen must be less than 1 per cent and water vapor only a tenth of 1 per cent of our supply.

The surface temperature of Mars rises in the tropics at times above the ordinary freezing point of water, as radiometric measures show. The mean temperature for the entire surface is minus 50°C, compared with 15°C above zero for the earth. Mars is generally a cold desert expanse; its average climate might be approached by our bleakest desert if it could be raised into the stratosphere. Nevertheless, scientists have believed that the dark markings with their seasonal changes might be areas of very hardy vegetation.

8.15 The Satellites of Mars

Mars' satellites were discovered at the favorable opposition of 1877 by Asaph Hall at the Naval Observatory, who named them Phobos and Deimos (Fear and Panic, the companions of Mars). They are very small, probably not exceeding 10 miles in diameter, and are so near the planet as to be invisible except with large telescopes at favorable times. Their color, as measured photoelectrically by Kuiper, is nearly neutral gray, not yellowish like moonlight.

Phobos, the inner satellite, revolves at the distance of 9280 kilometers from the center of Mars and 5900 kilometers from its surface. It completes a sidereal revolution in only $7^h\,39^m$, a period less than one third of the period of the planet's rotation in the same direction. As viewed from the planet it rises in the west and sets in the east. No other known satellite in the solar system revolves in a shorter interval than the rotation period of its primary.

The distance of Deimos from the center of Mars is 23360 kilometers and its period of revolution is $30^h\,18^m$. It is smaller than the inner satellite and only a third as bright. One recent study of the periods suggested

that the satellites were spiralling in toward the planet. This raised considerable speculation that they were artificial objects. Extensive study by D. Pascu fails to reveal any such changes.

More than 100 years before the discovery of these satellites, Voltaire mentions them in the story of Micromegas; and Swift's Gulliver, in reporting the scientific achievements of the Laputans, refers to their observations of two satellites of Mars, "whereof the innermost is distant from the center of the planet exactly three of his diameters, and the outermost five; the former revolves in the space of ten hours, and the latter in twenty-one and a half."

THE ASTEROIDS

The asteroids, *or* minor planets, *are the thousands of small bodies which revolve around the sun mainly between the orbits of Mars and Jupiter. The term "asteroid" (starlike) describes the appearance of almost all of them with the telescope. With the single exception of Vesta they are invisible to the naked eye. Ceres, the largest asteroid, is 768 kilometers in diameter. The majority are less than 80 kilometers and some are known to be scarcely a mile in diameter. The combined mass of all the asteroids is probably not greater than 5 per cent of the moon's mass.*

8.16
Discovery of the Asteroids

Many asteroids have irregular forms, as shown by their periodic fluctuations in brightness in their rotations. This suggests that some smaller asteroids may be fragments of larger ones and that larger ones may have been chipped by collisions.

Ceres, the first known asteroid, was discovered accidentally by G. Piazzi in Sicily, on the first evening of the 19th century, because of its motion among the stars he was observing. It proved to be a minor planet revolving around the sun at a mean distance 2.8 times the earth's distance. This was the planet for which some other astronomers were searching because it seemed to be required by the Titius-Bode Relation (7.5). A second asteroid, Pallas, was found in the following year by an observer who was looking for Ceres; and this discovery promoted the search for others.

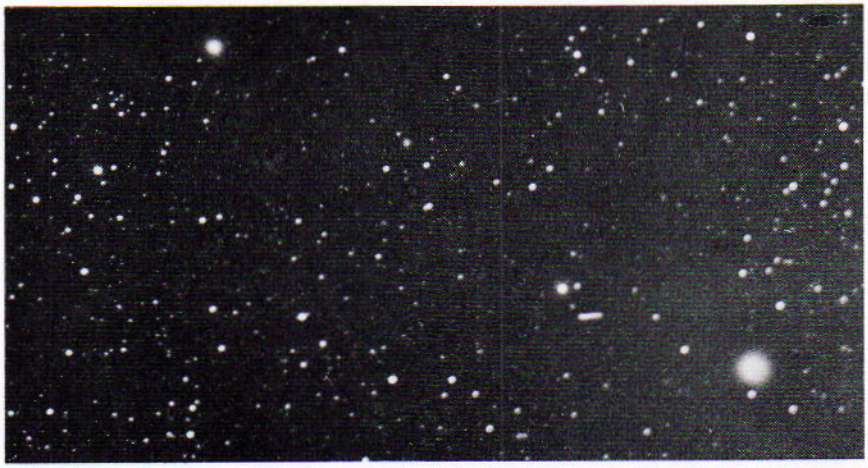


Figure 8.16A
Trails of three asteroids. *(Photographed at Königstuhl-Heidelberg)*

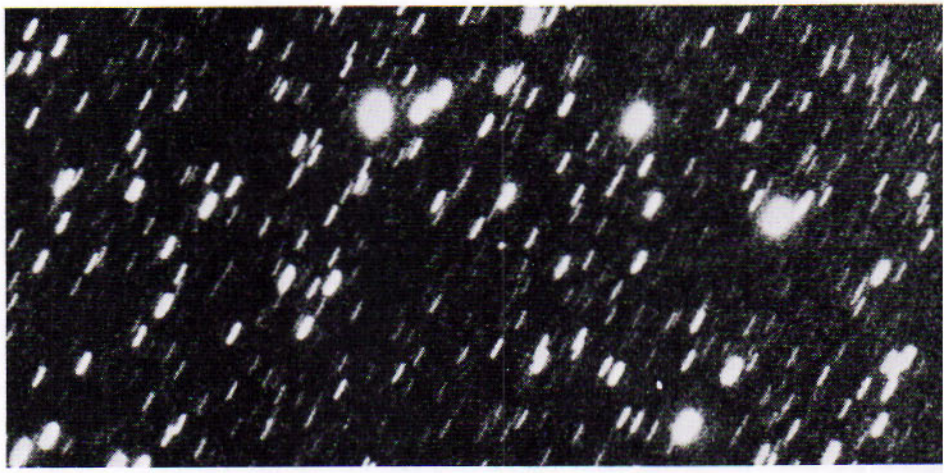


Figure 8.16B
Recovery of the minor planet 1322, Copernicus. The round image in the center is the minor planet. The star images are trailed. *(Photograph by Frank K. Edmondson, Goethe Link Observatory)*

The search was carried out visually for nearly a century. The observer at the telescope compared the stars in a region of the sky with a chart previously made of the region. If an uncharted star was seen, it was watched hopefully for movement among the stars that would reveal its planetary character. By this slow procedure, 322 asteroids had been discovered by 1891, when Max Wolf at Heidelberg was the first to apply photography in the search. In this modern method a time exposure of an hour or so is made with a wide-angle telescope. A fast-moving asteroid appears as a trail among the stars in the developed negative.

Hundreds of asteroids are now picked up each year in celestial photographs, often in the course of other investigations, and most of them are not observed thereafter. F. K. Edmondson and associates at the Goethe Link Observatory, Indiana University, have recovered many asteroids that might otherwise have been lost. Their procedure has been to shift the 10-inch Cooke telescope to follow the expected motion of the object

during each exposure. In their photographs the stars appear as short trails and the asteroids as points (Fig. 8. 16B), thus permitting the recovery of objects too faint to be detected if they were allowed to trail. Asteroids are designated by a circle with a number inside it. The number is usually the order in which it was discovered. Thus, Danae was the sixty-first asteroid discovered and is designated as (61).

8.17 The Orbits of the Asteroids

When its orbit has been reliably determined, an asteroid receives a permanent running number and a name that is less often used unless the object has unusual interest. The numbered asteroids exceed 1600. New asteroids, as their discoveries are reported, are given temporary designations by the central bureau under the direction of Paul Herget at Cincinnati Observatory.

The orbits have more variety than those of the principal planets. Although the majority are not far from circular and only slightly inclined to the ecliptic, some depart considerably from the circular form and are not confined within the bounds of the zodiac. As an extreme case, the orbit of Hidalgo has an eccentricity of 0.66 and is inclined 43° to the ecliptic; its aphelion point is as far away as Saturn. The revolutions of all asteroids are direct. The periods are mainly between 3½ and 6 years.

The asteroid orbits are not distributed at random through the region between the orbits of Mars and Jupiter. There are gaps in the neighborhoods of distances from the sun where the periods of revolution would be simple fractions, particularly one third, two fifths, and one half, of Jupiter's period. The avoidance of these distances, first announced, in 1866, by Kirkwood at Indiana University, is ascribed to frequent recurrences there of disturbances by Jupiter.

There are accumulations instead of gaps, however, at distances where the fractions are not far from unity. The situation is especially interesting where the period of revolution of asteroids is the same as Jupiter's period.

8.18 The Trojan Asteroids

Long ago, the mathematician J. L. Lagrange discovered a particular solution of the problem of 3 bodies and concluded that when the bodies occupy the vertices of an equilateral triangle the configuration may be stable. Although no celestial example was then known, he took as a hypothetical case a small body moving around the sun in such a way that its distances from Jupiter and the sun remained equal to the distance separating those two bodies. If the small body is disturbed, it will oscillate around its vertex of the triangle.

More than a dozen asteroids are examples of this special case. Achilles was the first of these to be discovered, in 1906. Named after the Homeric heroes of the Trojan War, they are known as the *Trojan group.* In their revolutions around the sun they oscillate about two points east and west of Jupiter, which are equally distant from that planet and the sun.

8.19 Asteroids that Pass Near Us

Several asteroids are known to come within the orbit of Mars and to pass nearer the earth's orbit than does any one of the principal planets. Some of these come in closer to the sun than the earth's distance. They are generally very small and are faint even when they are passing nearest us. The chance of discovering them as they speed by is so slight as to suggest that they are numerous.

Eros, discovered in 1898, is 168.9 million kilometers from the sun at its perihelion. Its least distance from the earth's orbit, also near the perihelion point, is less than 22.4 million kilometers. The most favorable oppositions of this 24-km asteroid are infrequent; the latest occurred in 1931, providing data for a re-examination of the value of the astronomical unit (7.16). The next close approach will come in 1975. Eros held the distinction of being the most neighborly of the planets until the discovery of Amor, in 1932. This asteroid comes to perihelion 16 million kilometers outside the earth's orbit.

Apollo, also discovered in 1932, has its perihelion inside the orbit of Venus and passed within 4.8 million kilometers of our orbit. Adonis, discovered in 1936, has its perihelion slightly farther than Mercury's mean distance from the sun; it passed about 1.6 million kilometers from the orbits of Venus, the earth, and Mars. Hermes, discovered in 1937, may have come even nearer the earth than did the other two. These three

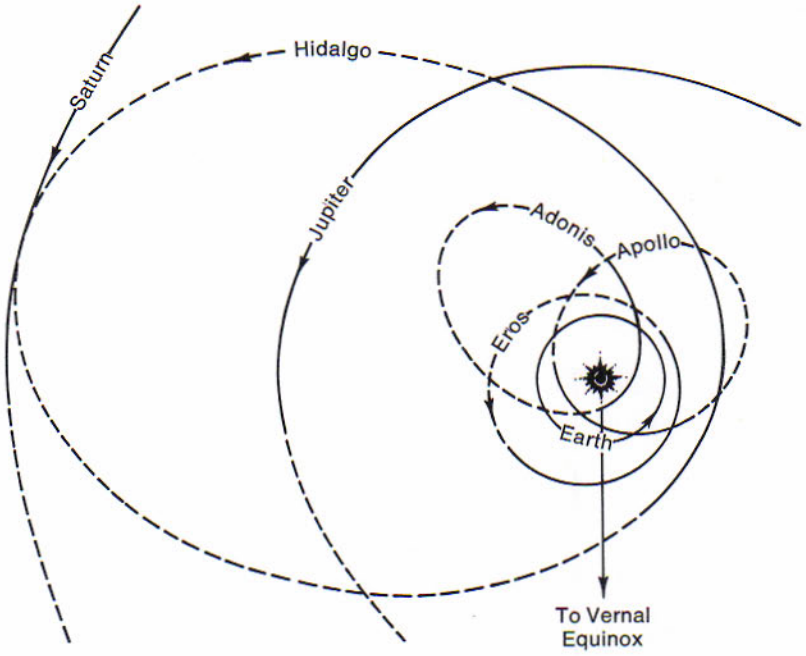


Figure 8.19
Orbits of four unusual asteroids. Broken lines represent the parts of the orbits south of the plane of the ecliptic. *(Adapted from a diagram by Dirk Brouwer)*

asteroids are about 1.6 kilometers in diameter. They were visible for such short intervals that their orbits were not very reliably determined. There is small chance of their being sighted again.

More recently, photographs by C. D. Shane and C. A. Wirtanen with the 20-inch astrographic camera at Lick Observatory revealed the long trails of other asteroids near the earth. In 1949, Walter Baade noticed a long trail in a photograph with the 48-inch Schmidt telescope; it proved to be the trace of the only known asteroid that crosses inside Mercury's orbit. Named Icarus, this object comes at its perihelion less than 32 million kilometers from the sun.

8.20 Irregular Shapes of Asteroids

Many asteroids have irregular shapes, so that they fluctuate periodically in brightness as they rotate and present varying cross sections for the reflection of sunlight to us. Eros is an example. Its brightness alternately increases for 79 minutes and then diminishes during an equal interval. The amplitude of the fluctuation varies conspicuously. At times the greatest brightness is three times the least; at other times the difference is slight.

The explanation is that Eros is shaped roughly like a brick 23 kilometers in length and 8 kilometers in width and thickness. It rotates on an axis through the middle of the brick from top to bottom, once around in $5^h\ 16^m$. When its edge is toward us, the asteroid presents its larger sides and smaller ends in turn, becoming brighter and fainter in the sunlight twice during each rotation. When Eros is in another part of its orbit where its top or bottom are more nearly toward us, its variation in brightness as it rotates is considerably less.

Photoelectric studies of representative asteroids by G. P. Kuiper and associates, mainly with the McDonald 82-inch telescope, have shown that over 90 per cent of the asteroids are variable in brightness. The rotation periods range from $2^h\ 52^m$ to 20 hours. The rotation axes appear to have random orientation. At least one asteroid, Eunomia, has retrograde rotation with its equator only slightly inclined to the ecliptic.

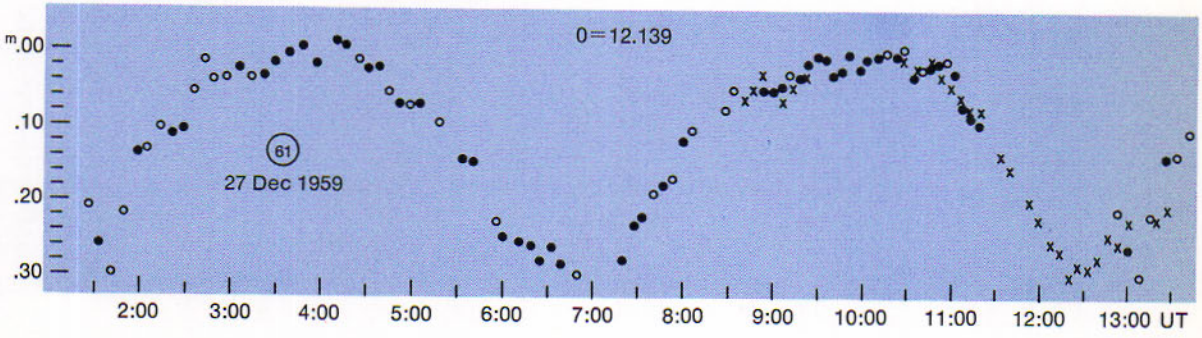


Figure 8.20
Light curve of Danae (Asteroid (61)). The almost regular repetition of the light variations indicates a roughly symmetrical object, but clearly not a spherical one. *(Adapted from* Astrophysical Journal, *Vol. 137, by permission. Observations by H. J. Wood III and G. P. Kuiper)*

Their irregular shapes suggest that the majority of asteroids are fragments resulting from collisions of larger bodies or have been nicked by collisions. Many fragments could have been thrown into orbits of higher eccentricity and inclination to the ecliptic than those of the bodies before collision. Some might then pass close to the earth and might even collide with the earth. Thus many meteorites may well be fragments of asteroids.

JUPITER, THE GIANT PLANET

Jupiter is the largest planet and is more massive than all the others combined. It is brighter in our skies than any other planet except Venus and occasionally Mars. At somewhat more than 5 times the earth's distance from the sun, it revolves in a period of nearly 12 years, so that from year to year it moves eastward among the stars through one constellation of the zodiac. Jupiter's banded disk and 4 bright satellites are easily visible with a small telescope. It has 12 known satellites in all, the largest number for any planet.

8.21 Jupiter's Cloud Markings

Viewed with a large telescope, Jupiter exhibits a variety of changing detail and color in its cloudy atmosphere. Brown bands parallel to the planet's equator appear on a yellowish background. The banded structure is associated with the rapid rotation. Jupiter's rotation is direct and its period is less than 10 hours, which is the shortest for all the principal planets; the speed of the rotation at the equator exceeds 40,000 km per hour.

Irregular cloud markings and bright and dark spots break the continuity of the bands. Some are short-lived and change noticeably from day to day, suggesting considerable turmoil beneath the cloud levels. Other spots persist for a very long time. Especially remarkable in this respect is the *Great Red Spot,* which has been visible for at least a century. This elliptical brick-red spot has been as long as 48,000 kilometers and as wide as 13,000 km. It drifts about like a solid floating in the near liquid lower atmosphere. (Plate III)

The positions of many markings vary as well as their forms, so that the rotation periods determined from two spots are not likely to agree precisely. A bright spot in a latitude somewhat south of the Great Spot has a period of rotation 20 seconds the shorter of the two; it drifts by the Spot and gains a lap on it in about 2 years.

8.22 The Constitution of Jupiter

Our understanding of the physical status of the planet is based mainly on observed features at and above the level of the obscuring clouds. The temperature at this level is −130°C as determined by the radiometric measures. It is slightly greater than the temperature that would be expected from heating only by radiation from the distant sun.

The light we receive from Jupiter is sunlight that has been reflected by the planet's clouds and has passed through the small amount of its atmosphere above the cloud level. The spectrum of this light is the solar spectrum with additional molecular bands of methane and ammonia, and fine traces of molecular hydrogen. No other molecules or elements have been reported. Methane is gaseous at this low temperature; its ordinary boiling point is −162°C. Ammonia freezes at −78°C; this constituent must be present as crystals, which partly sublime in the feeble sunlight. Methane and ammonia are among the impurities in Jupiter's atmosphere, which like the sun must be composed mainly of hydrogen and helium.

Radiation from Jupiter in the radio wavelengths was first noticed by B. Burke and K. Franklin, and is of three observed types: (1) thermal emission consistent with the temperature already stated; (2) nonthermal emission that issues in short blasts from the planet's ionosphere; (3) nonthermal emission of a different type, which is weak and only at microwave frequencies. It is believed to come from an equivalent of the terrestrial Van Allen belt. This radiation belt is presumably more extensive and the accompanying magnetic field is stronger than are those of the earth (2.7).

It is interesting to note that the nonthermal emission that issues as bursts is apparently related to the positions of the satellites, especially Io (Galilean satellite I). It seems almost 100% certain that bursts will occur when Jupiter's magnetic pole sweeps past Io and this body is in the proper position. Additional burst-like activity may be correlated with the solar cycle and solar flares.

Conditions below the cloud level of Jupiter are derived somewhat imperfectly by analysis from observed data about the planet as a whole. Wildt has remarked that the study of the internal structure of a cold planet presents many more difficulties than does the analysis of the interior of a star.

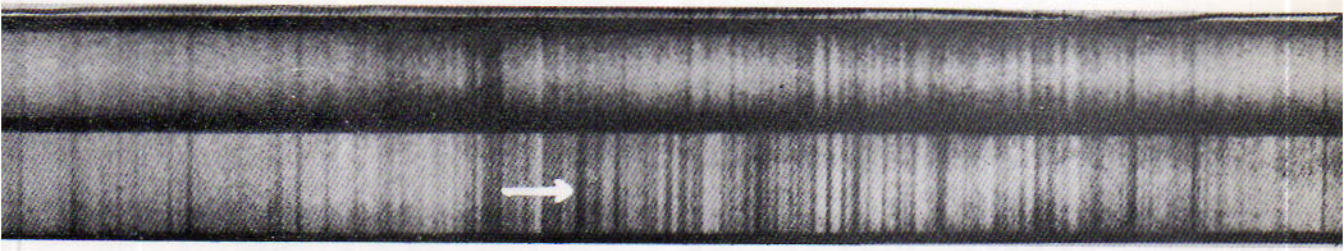


Figure 8.22
Near infrared spectra of Saturn (upper band) and Jupiter (lower). The dark lines of ammonia (arrow) are stronger in Jupiter's spectrum. *(Photograph by Theodore Dunham, the Hale Observatories)*

8.23 The Interior of Jupiter

The amount of the bulging of Jupiter's equator provides one clue to conditions in the interior. With its swift rotation the planet would be even more oblate than is observed if its mass were not highly concentrated toward its center. Other indications of what is hidden beneath the clouds are the low temperature and low average density, about 1.3 times the density of water, of the whole planet, which requires very light material in the outer parts.

Hydrogen is the predominant chemical constituent in recent theoretical models of Jupiter designed by R. Wildt, W. C. DeMarcus, and others. Conditions in the interior are considered highly uncertain. The compression below the cloud levels must eventually reach a critical value where there is no distinction between the gaseous and liquid states. At an undetermined distance below this level even the hydrogen should be converted to a solid state. The conclusion is that Jupiter has little resemblance to the earth. Indeed, some astronomers believe that Jupiter was an incipient star and only the presence of numerous other planets in similar nearly circular orbits kept it from being so. We will examine these views more carefully in Chapter 16.

Jupiter has a magnetic dipole field revealed by the nonthermal synchrotron emission. A detailed study shows that the magnetic axis is quite displaced from the center of the planet and tilted with respect to the axis of rotation as well. We refer to such an object as an "oblique rotator" and we will encounter this situation in the stars later.

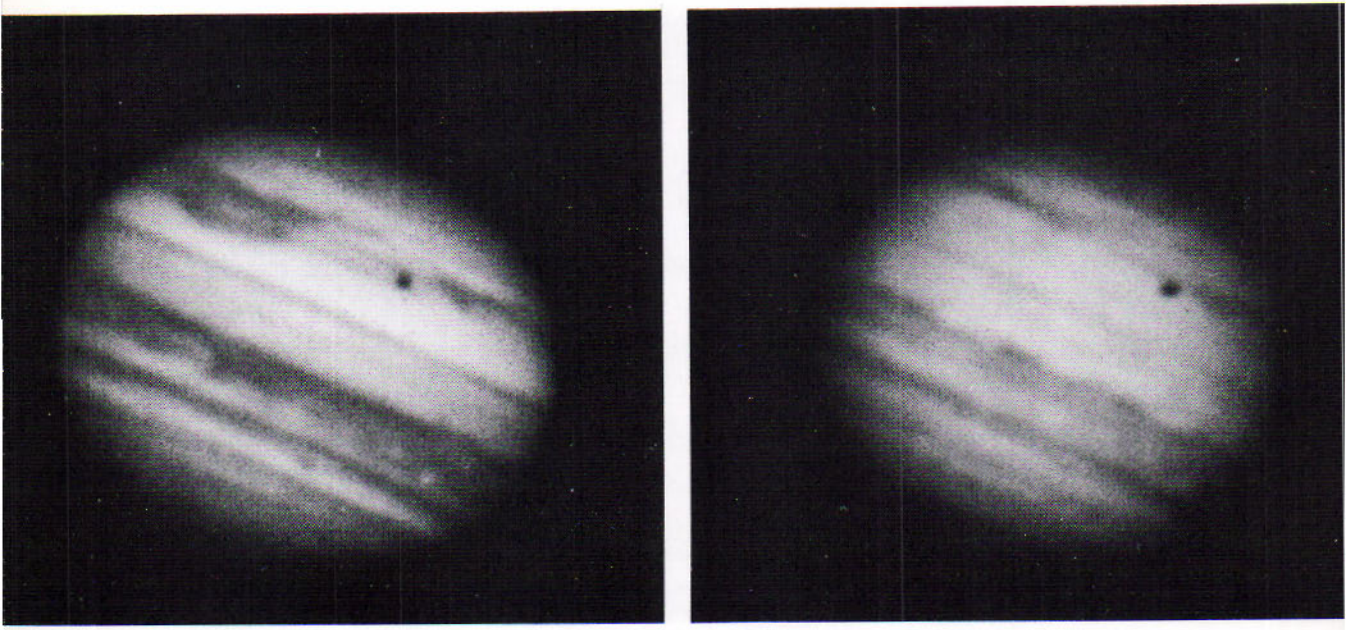


Figure 8.23
Jupiter in violet light (left) and infrared light (right). Notice the difference in the apparent diameter of the planet as seen in the two spectral regions. *(Lick Observatory photographs)*

8.24 The Inner Satellites

Jupiter's 12 known satellites are sharply divided into three groups: the inner satellites and the two groups of outer satellites. The 5 inner satellites have direct revolutions in orbits that are nearly circular and nearly in the plane of the planet's equator. Four of these are bright enough to be readily visible with a small telescope.

The four bright satellites were discovered by Galileo on January 7, 1610. They were independently discovered the following evening by the German astronomer Marius who gave them their names. They are generally designated, however, by numbers in order of their distances from the planet.

The 1st and 2nd satellites are about as large as the moon. The 3rd and 4th are 50 per cent greater in diameter; they are the largest of all satellites and are comparable in size with the planet Mercury. Spectrometer tracings in the infrared suggest that the 2nd and 3rd satellites may be covered with snow. At their greater distance from the sun the combined light of all four upon Jupiter is not more than 30 per cent of the light of the full moon on the earth. Their periods of rotation and revolution are the same. After the four bright ones, the numbering of the other satellites is in order of their discovery. The 5th satellite differs from the others of its group in its small size. Nearest of all to the planet and more difficult to observe on this account, it is the swiftest of all satellites, revolving at the rate of 1600 kilometers a minute.

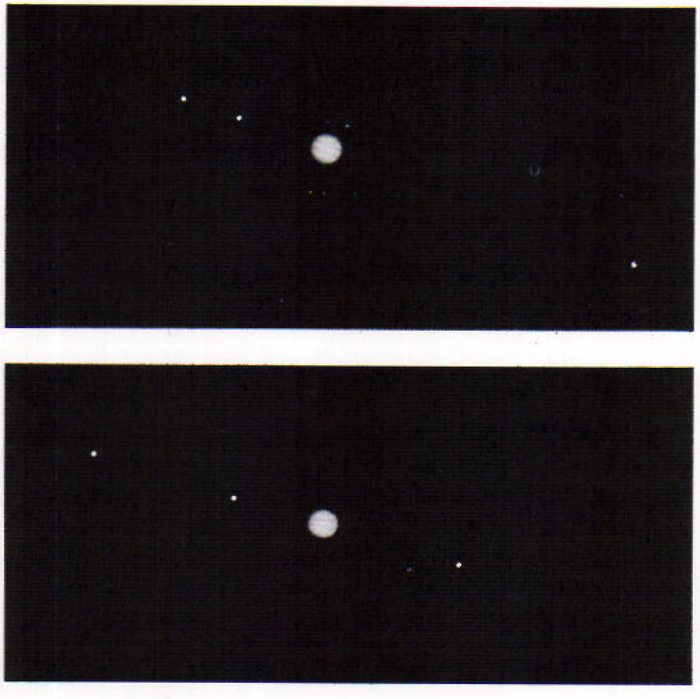


Figure 8.24
Jupiter and its four bright satellites. The lower photograph was taken four hours later than the upper one. *(Official Navy photographs of D. Pascu)*

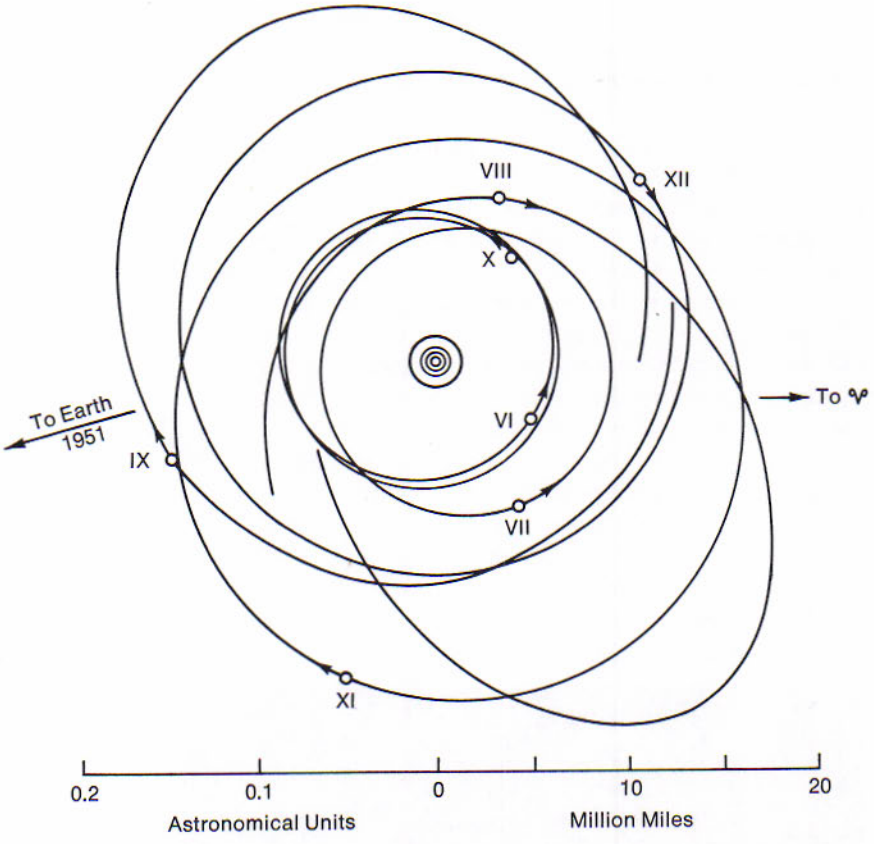


Figure 8.25
Orbits of Jupiter's satellites, showing marked changes in two orbits during a single revolution. *(Diagram by Seth B. Nicholson)*

8.25 The Outer Satellites

The orbits of the outer satellites show considerable eccentricity and inclination to the ecliptic. All seven are small and very faint. They were discovered photographically, the latest one by S. B. Nicholson in 1951; he remarks that the 12th satellite is slightly fainter than the 10th, which itself is not brighter than the light of a candle at the distance of 4800 kilometers. The 6th satellite is the only outer one that would be visible from Jupiter without a telescope.

These satellites are in two groups. One group contains the 6th, 7th, and 10th satellites, which have direct revolutions at the average distance of a little more than 11 million kilometers from the planet, and in periods around 260 days. The outer group contains the 8th, 9th, 11th, and 12th satellites. These have retrograde revolutions at the average distance of about 22 million kilometers and in periods around 700 days; they are the most distant of all satellites from their primaries. Jupiter's control over these remote satellites is disputed by the sun, which by its attraction greatly disturbs their orbits.

Data on the outer satellites (see Appendix Table 2 at the end of this book) are as given by Nicholson. The order of mean distance of the outermost four from the planet, as he points out, is subject to change in a few years by perturbations by the sun. The diameters of the faint satellites are estimated from their brightness, and the values he considers perhaps somewhat too great.

8.26 Eclipses and Transits of the Bright Satellites

The orbits of the four bright satellites are always presented nearly edgewise to us. As these satellites revolve around Jupiter, they accordingly appear to move back and forth in nearly the same straight line. The forward movement takes them behind the planet and through its shadow, although the 4th satellite often clears both; the backward movement takes them in front of the planet, when their shadows are cast upon its disk. These occultations, eclipses, transits, and shadow transits add interest to observations of Jupiter with the telescope. The times of their frequent occurrences are predicted in some of the astronomical almanacs.

Figure 8.26A
Phenomena of Jupiter's satellites as observed from the earth after Jupiter's opposition.

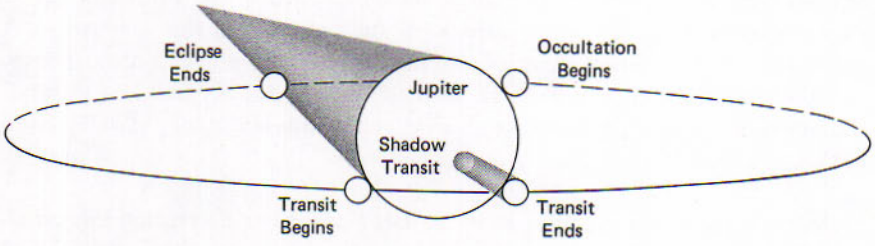
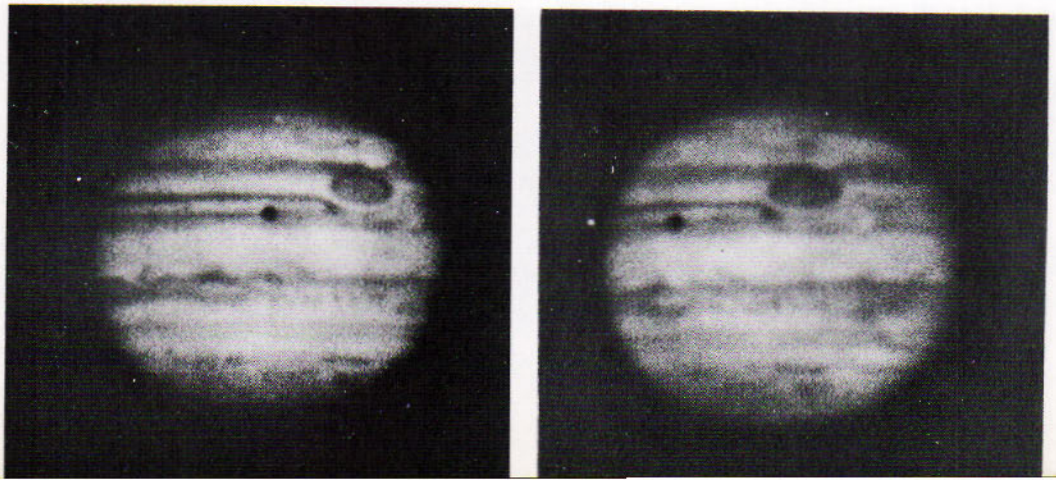


Figure 8.26B
Transit of Jupiter's tenth satellite. The shadow is clearly visible on both photographs, the satellite, only in the right one, shortly after having ended its transit. Notice also the rotation of the red spot. Time interval between exposures: 36 minutes. *(Lowell Observatory photograph)*



SATURN AND ITS RINGS

Saturn is the most distant of the bright planets from the sun and was the most remote planet known to early astronomers. At nearly twice the distance of Jupiter, it revolves around the sun in a period of 29½ years. This planet ranks second to Jupiter in size, mass, and number of known satellites. Saturn has the least mean density and the greatest oblateness of any principal planet. It is unique in the possession of a system of rings that encircle the planet and make it one of the most impressive of celestial objects viewed with the telescope.

8.27 The Constitution of Saturn

Saturn resembles Jupiter in many respects. The atmospheric markings are likewise arranged in bands, which are here more regular and less distinct. A broad yellowish band overlies the equator, and greenish caps surround the poles. The absorption of ammonia in the spectrum is much weaker, while that of methane is stronger than in Jupiter's spectrum. At the lower temperature of −150°C at the cloud levels of Saturn the ammonia gas is more nearly frozen out, so that the sunlight analyzed in the spectrum has penetrated farther down through the methane of the planet's atmosphere. Nonthermal radio energy has not been reported from Saturn. The thermal radiation at λ4cm agrees with the infra red temperature of −150°C cited above. (Plate IV)

Saturn's period of rotation at its equator has been determined spectroscopically as $10^h\,02^m$. Long-enduring spots, which can show the period more reliably, are rare. A period of about $10^h\,14^m$ has been derived from several white spots in the equatorial zone, and a second period of about $10^h\,40^m$ has been found for other spots around latitude 60°. T. A. Cragg at Mount Wilson Observatory concludes that Saturn rotates in these two basic systems instead of having a steady increase of period with increasing latitude. The rapid rotation combined with the large size and low density of the planet can account for its conspicuous oblateness. Hydrogen is the main constituent, as in the case of Jupiter.

8.28 Saturn's Rings

Saturn is encircled by three concentric rings in the plane of its equator. They are designated as the *outer ring,* the middle or *bright ring,* and the inner or *crape ring.* The rings are invisible to the unaided eye and were therefore unknown until after the invention of the telescope. The diameter of the entire ring system is 275,000 kilometers, or 2.3 times the equatorial diameter of the planet (119,000 kilometers). Because they have nearly

twice the diameter of Jupiter and are about twice as far from us, the rings have about the same apparent diameter as that of Jupiter.

The bright ring is 26,000 kilometers in width and its outer edge is as luminous as the brightest parts of the planet. It is separated from the outer ring by the 4,800-km. gap known as *Cassini's division* after the name of its discoverer. This is the only real division in the rings. There is no gap between the bright ring and the crape ring.

A surprising feature of the rings is that they are extremely thin; their appreciable thickness can scarcely exceed 16 kilometers.

8.29 Saturn's Rings at Different Angles

The rings are inclined 27° to the plane of the planet's orbit and they keep the same direction during the revolution around the sun. Thus their northern and southern faces are presented alternately to the sun and also to the earth; for as viewed from Saturn these two bodies are never more than 6° apart. Twice during the sidereal period of 29½ years the plane of the rings passes through the sun's position. It requires nearly a year on each occasion to sweep across the earth's orbit. As the earth revolves in the meantime, the rings become edgewise to us from 1 to 3 times, when they disappear with small telescopes and are only thin bright lines with large ones.

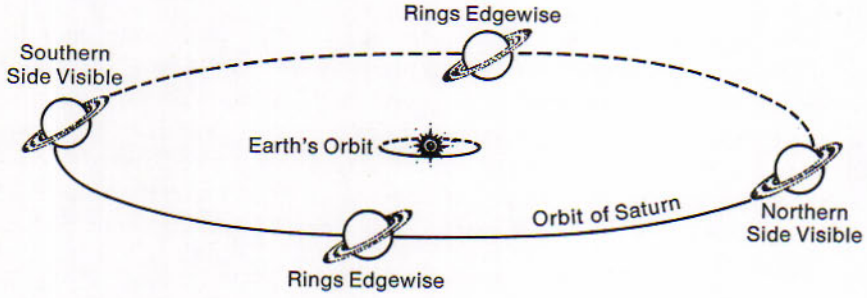


Figure 8.29
Cause of the different aspects of Saturn's rings. The plane of the rings is inclined 27° to the plane of Saturn's orbit.

The latest widest opening of the northern face of the rings occurred in 1959, when Saturn was near the position of the winter solstice. The latest edgewise presentation occurred late in 1965 and again in 1966. The following widest opening of the southern face of the rings will come in 1974.

When the rings are widest open, their apparent breadth is 45 per cent of the greatest diameter and one sixth greater than the planet's polar diameter. On these occasions Saturn appears brighter than usual, because the rings at this angle reflect 1.7 times as much sunlight as does the planet alone. When it is also near perihelion and in opposition, Saturn appears twice as bright as Capella.

8.30 Discrete Nature of the Rings

Saturn's rings consist of multitudes of separate particles revolving around the planet in nearly circular orbits in the direction of its rotation. They have the appearance of continuous surfaces because of their great distance from us. G. P. Kuiper's infrared studies with the lead sulfide cell of the sunlight reflected from the rings suggest that they are composed of ice particles, although L. W. Fredrick has discovered a broad band in the rings' reflection curve centered at 1.06 microns that is not present in the reflection spectrum of H_2O ice crystals. Ammonia ices have not been studied yet. Bright stars and Saturn's brighter satellites can be seen through the rings which confirms their multi particulate nature, but the most convincing argument is due to spectroscopy.

The Doppler effects (4.9) in the spectrum of Saturn's rings (Fig. 8.30A) first demonstrated by J. E. Keeler, show that the inner parts of the rings revolve around the planet faster than the outer parts, just as a planet that is nearer the sun revolves around it faster than one that is farther away. The reverse would be the case if the rings were continuous surfaces; for all parts would then go around in the same period, and the outside having farther to go would move the faster. The behavior of the spectrum lines is explained by Fig. 8.30B, where the slit of the spectroscope is placed along Saturn's equator. The upper parts of the planet and the rings are here approaching the observer, and the lower parts are receding from him.

The revolution periods of the outer edge of the outer ring and the inner edge of the bright ring are respectively $14^h\ 27^m$ and $7^h\ 46^m$, and the material in the crape ring goes around in still shorter periods. Because Saturn's equator rotates in about 10 hours, it is evident that the outer parts of the ring system move westward across the sky of Saturn. A considerable part of the bright ring, however, and all of the crape ring must rise in the west and set in the east as seen from the surface of the planet, duplicating the behavior of Phobos (8.15) in the sky of Mars.

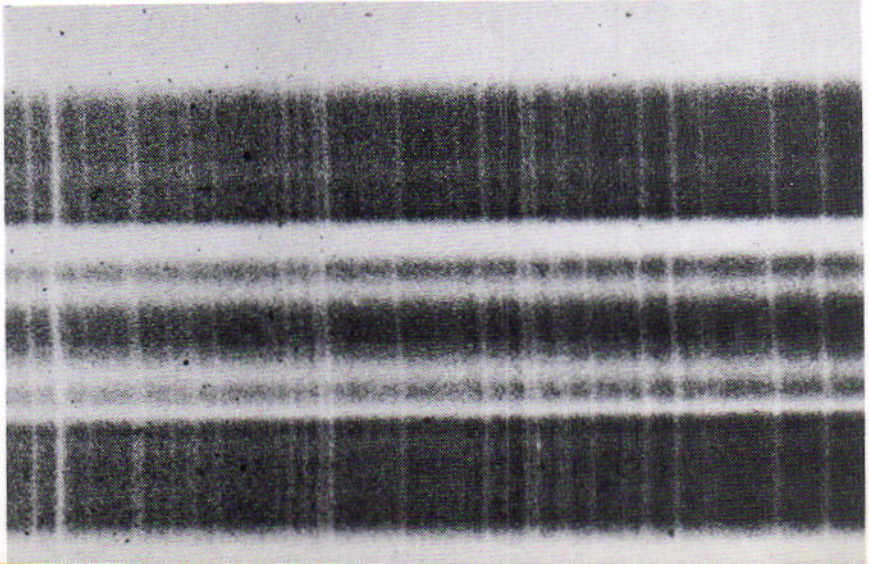


Figure 8.30A
The original discovery plate of the rotation of Saturn's rings by Keeler. The moon form's the comparison spectra. *(Courtesy of W. Beardsley, Allegheny Observatory)*

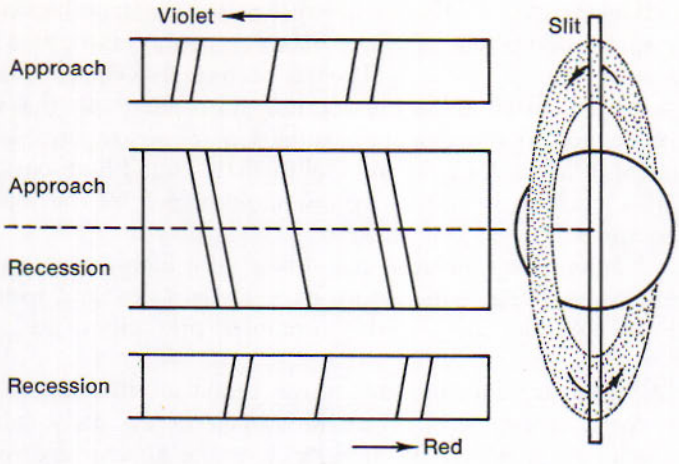


Figure 8.30B
Doppler effects in the spectrum of Saturn and its rings, explaining Figure 8.31A.

8.31 The Origins of Saturn's Rings

The rings' origin is associated with their nearness to the planet. According to a theory that was invoked in this respect long before the spectroscopic evidence was available, a solid ring so close to the planet would be shattered by the gravitational strain to which it would be subjected, whereas a ring of many small pieces would be reasonably stable. A liquid satellite of the same density as the planet would be broken into small fragments by the tide-raising force of the planet if its distance from the center of the planet is less than 2.4 times the planet's radius.

All parts of Saturn's rings are well within this critical distance, but the nearest satellite is safely outside. Because a stable satellite could not have formed at the distance of the ring, the ring must have formed directly as such. The mass of the ring system is not known from observation but may be similar to that of the innermost satellite of Saturn.

8.32 The Satellites of Saturn

Saturn has ten known satellites. The brightest, Titan, is visible with a small telescope as a star of the 8th magnitude. Five or six other satellites can be seen with telescopes of moderate aperture; they appear as faint stars in the vicinity of the planet and are easily identified by means of convenient tables in some of the astronomical almanacs. Phoebe, the most distant satellite, has retrograde revolution like Jupiter's outer satellites; all the others have direct revolutions. The tenth satellite, Janus, was discovered by A. Dollfus in 1967.

Two of Saturn's satellites are known to rotate and revolve in the same periods, as is shown by their variations in brightness in the periods of their revolutions. Evidently their surfaces are irregular in form or are uneven in reflecting power. Iapetus is the most remarkable in this respect; it is five times as bright at western as at eastern elongation.

Titan, the largest of Saturn's satellites, is remarkable in three respects. (1) It is the only satellite in the solar system definitely known to have an

atmosphere. G. P. Kuiper has recognized methane bands in its spectrum, and he points out another feature probably associated with the atmosphere: (2) The color of Titan is orange. It seems to him likely that the color is caused by action of the atmosphere on the surface material, analogous to the oxidation supposed to be responsible for the orange color of Mars. (3) Lyot and Dollfus have found limb-darkening on Titan's disk, such as is shown by the sun but not by the moon and Jupiter's satellites.

Kuiper also concludes from their very high reflecting power, 0.8, that the inner satellites of Saturn have icy surfaces, and from their low densities that they are probably composed primarily of ice.

8.33 Stability of Atmospheres

Whether a planet or satellite can retain an atmosphere depends on the velocity of escape (5.16) at the surface of the body. It also depends on the mean speed of the molecules in the atmosphere, whether they can become high enough to effect an escape. This speed increases with the temperature of the gas and decreases with increasing weight of the molecules. J. H. Jeans showed that an atmosphere is likely to be retained for astronomical periods of time if the mean speed of its molecules is less than 20 per cent of the velocity of escape. For some gases in our own atmosphere the mean speeds at 20°C are: 2.2 km/sec for H_2, 0.8 for CH_4 and NH_3, 0.6 for N_2, 0.5 for O_2 and CO_2.

TABLE 8.III Critical Velocities for Retention of Atmospheres

Jupiter	18.4 km/sec	Titan	1.0 km/sec
Saturn	12.8	Jupiter's satellite III	0.9
Neptune	11.8	Jupiter's satellite IV	0.7
Uranus	9.0	Jupiter's satellite I	0.7
Earth	2.3	Mercury	0.7
Venus	1.9	Jupiter's satellite II	0.6
Mars	1.1	Moon	0.5

Table 8.III, adapted from data by Kuiper, lists planets and satellites for which the critical velocities are not less than the lowest speed we have given for the gas molecules. These critical velocities are the escape velocities for each body multiplied by the 4th root of its distance in astronomical units from the sun to eliminate the temperature difference, and then reduced to 20 per cent.

Aside from hydrogen, which would be predicted in any considerable amount only around the four giant planets, the dividing line in the table comes below Titan; and this is in fact the division between bodies having known atmospheres and those for which atmospheres have not been detected. Pluto and Triton, where atmospheres are suspected from the strength of their ultraviolet reflections, are not included in the list because of uncertainties in some of the data.

URANUS AND NEPTUNE

These planets rank fourth and third respectively to Jupiter in size and mass. They obviously belong to the gaseous planets. Uranus is unusual in that its axis of rotation lies almost in the plane of the ecliptic.

8.34 Discoveries of Uranus and Neptune

Uranus was discovered in 1781 by William Herschel in England; he was observing a region in the constellation Gemini when he noticed a greenish object that appeared somewhat larger than a star. The object proved to be a planet more remote than Saturn and was given the name Uranus. An examination of the records showed that Uranus had been seen 20 times in the hundred years preceding its discovery; each time the position had been measured and set down as that of a star.

Because no orbit could be found to fit the older positions satisfactorily, it was necessary to wait for later ones. At length, in 1821, a new orbit was calculated with allowance for the disturbing effects of known planets. It was not long, however, before Uranus began to depart appreciably from the assigned course, until in 1844 the difference between the observed and calculated positions in the sky had increased to more than 2′, an angle not perceptible to the unaided eye but regarded as an intolerable discrepancy by astronomers. There seemed no longer any doubt that the motion of Uranus was being disturbed by a planet as yet unseen.

U. J. Leverrier in France discovered Neptune in 1846. From the discrepancies in the motion of Uranus he was able to calculate the place of the disturber among the stars. All that remained was to observe it. An astronomer at the Berlin Observatory searched with the telescope for the new planet and soon found it within a degree of the specified place in the constellation Aquarius. The discovery was acclaimed as a triumph for the law of gravitation, on which the calculation was based. J. C. Adams in England had successfully completed a similar calculation, but did not obtain effective telescopic cooperation.

8.35 Uranus

The first planet to be discovered telescopically, Uranus is barely visible to the naked eye. It has direct revolution in a period of 84 years at 19 times the earth's distance from the sun and rotates once in 10¾ hours, having its equator inclined nearly at right angles to the ecliptic. Nearly 48,000 km in diameter, Uranus appears with the telescope as a small disc on which markings are not clearly discernible. The spectrum shows dark bands of methane and also a broad absorption band in the near infrared, recorded by Kuiper and identified by G. Herzberg with molecular hydrogen. This was the first direct evidence of the hydrogen molecule in the

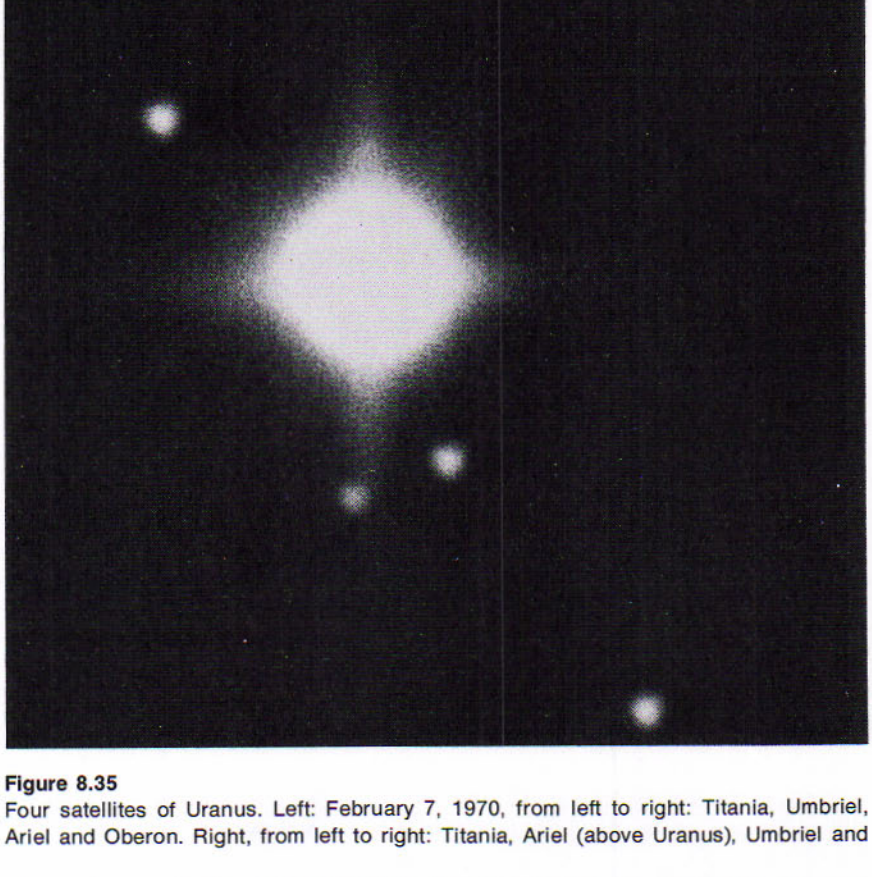


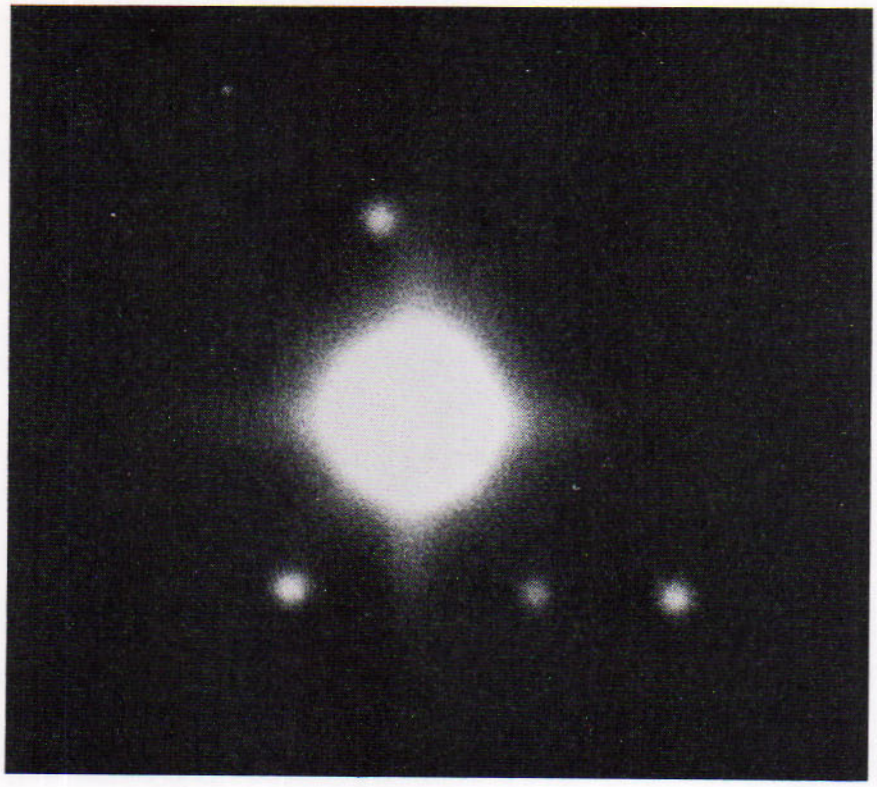
Figure 8.35
Four satellites of Uranus. Left: February 7, 1970, from left to right: Titania, Umbriel, Ariel and Oberon. Right, from left to right: Titania, Ariel (above Uranus), Umbriel and

atmospheres of the major planets. More recently, C. C. Kiess detected molecular hydrogen in the spectrum of Jupiter.

Uranus has five known satellites. The 5th, which is fainter and nearer the planet than the others, was discovered by Kuiper in 1948. The nearly circular orbits of the satellites are presented to the earth at various angles as the planet revolves; they were edge-wise to us in 1966 and will be flat-wise in 1987.

8.36 Neptune

This planet has nearly the same size as Uranus and seems to resemble it closely in other respects. Its diameter was recently measured by G. Taylor by means of stellar occultations to be 50,000 km. Neptune has direct revolution in a period of 165 years at a distance 30 times the earth's distance from the sun, and it rotates in the same direction once in 15.8 hours, according to the spectroscopic measures. Always invisible



Oberon. The fifth satellite, Miranda is inside the halation ring from the overexposed image of Uranus. *(Photographs by Elizabeth Roemer, Lunar and Planetary Laboratory, University of Arizona's Catalina Station)*

to the naked eye because it is so remote from the sun and earth, Neptune appears with the telescope as a star of the 8th magnitude. Very weak markings have been discerned on its small greenish disk.

Neptune has two known satellites. The first, Triton, is somewhat larger than the moon and is slightly nearer the planet than the moon's distance from the earth. Its mass, as determined by H. L. Alden, is 0.022 times the earth's mass, or nearly twice the mass of the moon, and it may have an atmosphere. Triton has a retrograde revolution around the planet, contrary to the direction of the planet's rotation.

The second satellite, Nereid, discovered by Kuiper in 1949, is much the smaller, fainter, and more distant from the planet; the distance ranges from about 1.6 to 9.6 million kilometers. The satellite has a direct revolution once in nearly a year in an orbit having an eccentricity of 0.75, the greatest for any known satellite.

Figure 8.36
Neptune and its inner satellite. *(Lick Observatory photograph)*

PLUTO

This most remote planet was discovered, in 1930, by C. W. Tombaugh at the Lowell Observatory as the successful result of a search for a planet beyond Neptune. The first two letters of the planet's name are the initials of Percival Lowell, who had initiated the search a quarter of a century before. From its small size and its unusual orbit for a principal planet, it is believed that Pluto was originally a satellite of Neptune, although recent studies have cast some doubt on this assumption.

8.37 Pluto

Pluto is visible with the telescope as a star now of visual magnitude 14.9, and 0.8 magnitude fainter in blue light. Its diameter is 5700 km, as measured by Kuiper with a disk meter on the 200-inch telescope. This figure is strengthened by a recent near occultation that set an upper limit of 6800 kilometers. Unless its density is greater than would be expected, its mass does not exceed a tenth of the earth's mass. Calculation of the mass from perturbations is difficult and the latest study sets the mass at 0.18 that of the earth; this is somewhat less than previous values, but is still believed large since it yields a minimum density of 6.5.

The noon equatorial equilibrium temperature (T_{max}) at the average distance of Pluto is only between 40° to 60°K. On the night side the temperature must drop to something like 10°K. If Pluto has an abundance of gases they must all be frozen out excepting molecular nitrogen and hydrogen. Assuming that our estimate of a mass of 0.1 earth masses is correct, the escape velocity is around 5 km/sec and the velocity critical to retaining an atmosphere is about 1.0 km/sec. At the temperature and pressures encountered at Pluto only the hydrogen could escape if even that. On the dark side the molecular nitrogen and oxygen rains or snows out. The pools of liquid and frozen nitrogen may well be the cause of the changing albedo of Pluto as they warm up and vaporize on the sunlit side. Marked changes correllating with the planet's heliocentric distance will determine whether or not this picture is correct.

$$T_{max} \simeq \frac{277}{\sqrt{d}}$$

d in a.u.'s

The period of Pluto's rotation is 6.390 days, as determined by M. F. Walker and R. Hardie and by Kuiper from photoelectric measures of its periodic fluctuations in brightness. Non-uniformities of its surface cause a variation of 0.1 magnitude in the brightness of the planet in the course of a rotation; the range is great enough to suggest that the axis is more nearly at right angles to than along the line of sight.

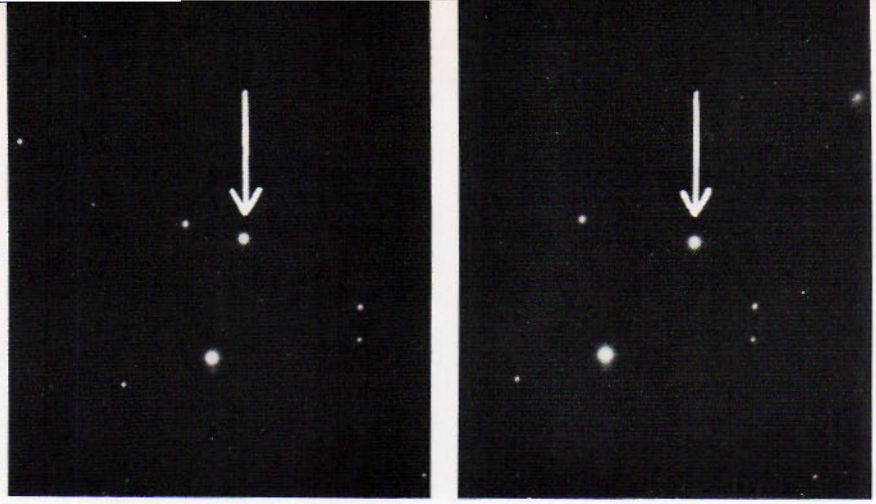


Figure 8.37
Two photographs of Pluto, showing motion of planet in 24 hours. *(Photograph from the Hale Observatories)*

8.38 The Orbit of Pluto

At its mean distance of 39½ astronomical units, or 5800 million kilometers, from the sun, Pluto has a direct revolution around the sun once in 248 years. Its orbit is inclined 17° to the ecliptic, the highest inclination for any of the larger planets, so that Pluto ventures at times well beyond the borders of the zodiac. With respect to its origin it might be included among the principal planets only for convenience in the descriptions.

The high eccentricity, 0.25, of Pluto's orbit introduces another feature that is unique among the larger planets. At aphelion Pluto is 2880 million kilometers beyond Neptune's distance from the sun, whereas at perihelion it comes 56 million kilometers nearer the sun than the orbit of Neptune. There is no danger of collision in our times, however; in their present orbits the two planets cannot approach each other closer than 384 million kilometers.

In Fig. 7.4 the plane of the page represents the ecliptic plane. The portion of Pluto's orbit south of this plane is indicated by the broken line in the figure. At the time of its discovery the planet was near the ascending node of its path. It will reach perihelion in 1989.

8.39 Tables of the Planets and Satellites

Some of the data on the planets and satellites in Tables 1 and 2 (Appendix) are taken from the *American Ephemeris and Nautical Almanac.* The adopted length of the astronomical unit is 149,600,000 kilometers. Data on the outer satellites of Jupiter are as given by Nicholson. Dimensions of some planets and satellites are derived by Kuiper from his measures of the apparent diameters with a disk meter, and the magnitudes of the satellites of Uranus are estimates by the same authority.

The diameters of a number of small satellites are uncertain, as is indicated by the question marks after their values in the table. Where the satellites do not show appreciable disks, the diameters are calculated from the observed brightness and assumed albedo.

REVIEW QUESTIONS

1. If we could observe Mars through opposition using radar, what would the observed frequency spread as a function of time look like?
2. If Mercury is at perihelion in the autumn, can we, as a rule, observe it at the end of astronomical twilight? What conditions are most favorable for observing Mercury?
3. Mercury's rotation period is ⅔ its period of revolution. Explain how this could have led the visual observers to deduce that the rotation period equalled the revolution period.
4. How does a phase lead or lag in temperature minimum near inferior conjunction indicate the sense of a planet's rotation. What assumptions are made?
5. We generally credit our atmosphere for protecting us from most meteors and erosion for eradicating the results from those that do strike the earth. What does the presence of craters on Mars tell us about Mars?
6. What conclusions can be obtained from observing the surface of Jupiter?
7. Explain the characteristics and three types of radio radiation from Jupiter.
8. What role does Jupiter play in the distribution of the asteroids?
9. What observations show that Saturn's rings are not solid? How do we know that the rings are very thin?
10. Why is Titan considered an unusual satellite?
11. What is unusual about Neptune?
12. Pluto's orbit has a major axis (a) of 39.52 a.u. and an eccentricity (e) of 0.25. At perihelion a planet's distance is given by $a \times (1 - e)$ and at aphelion by $a \times (1 + e)$. How much farther from the sun is Pluto at aphelion than at perihelion?
13. Determine the mass of Jupiter relative to the mass of the earth using data in Table 2 (Appendix). Check your results in Table 1 (Appendix). Do the same for Uranus using Arial.

REFERENCES

Moore, Patrick, *A Guide to Mars* (Macmillan, New York 1958)

Page, Thornton and Lou W. Page, eds., *Wanderers in the Sky* (Macmillan Co., New York, 1965)

Slipher, E. C., *Mars*, Flagstaff, Arizona: Northland Press, 1962

Slipher, E. C., *The Brighter Planets*, ed. J. Hall, Flagstaff, Arizona: Northland Press, 1964

Whipple, F. L., *Earth, Moon and Planets* (Harvard Univ. Press, Cambridge, Mass., 1963) rev. ed.

FOR FURTHER STUDY

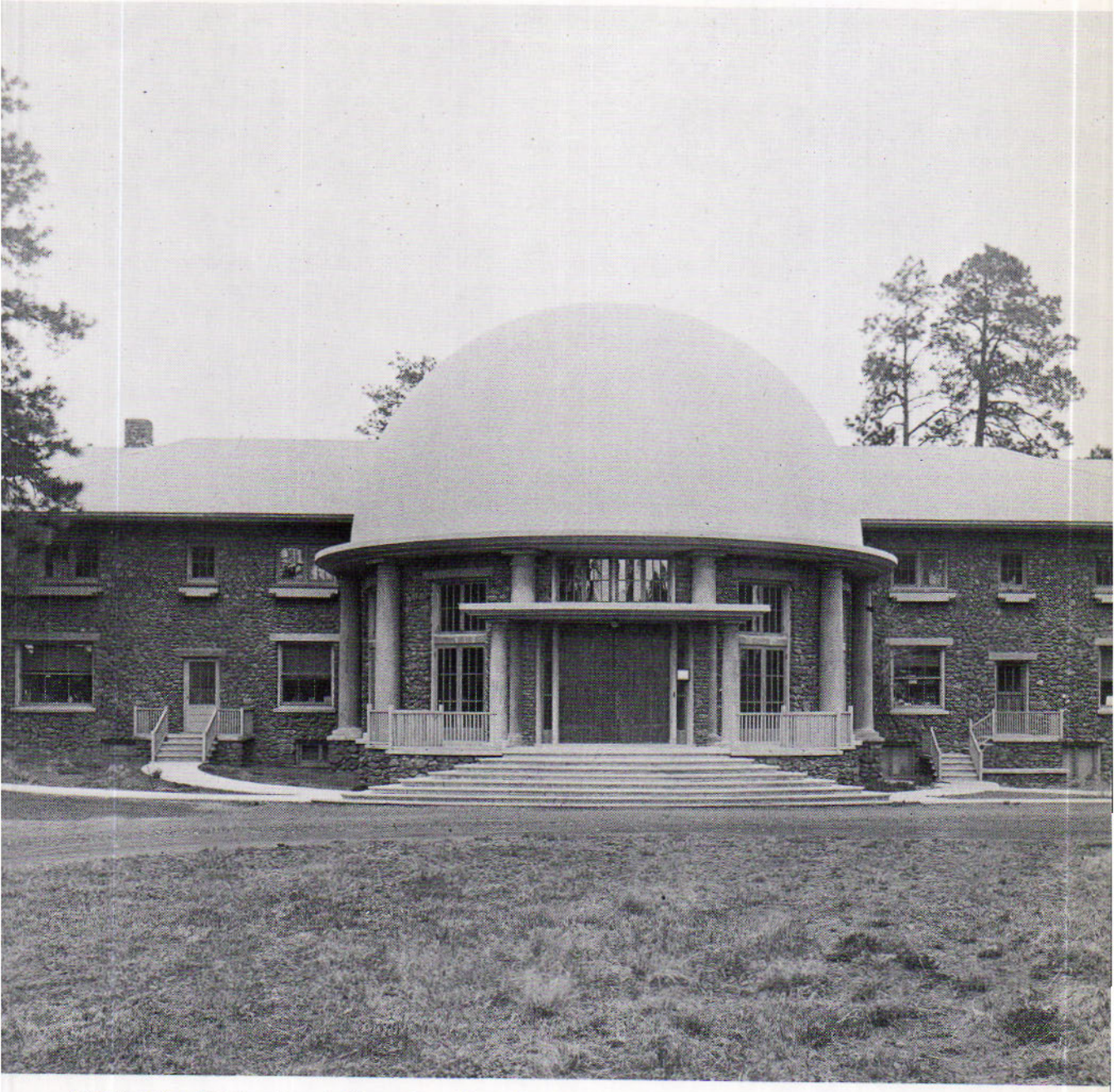
Brandt, John C. and Paul W. Hodge, *Solar System Astrophysics*, New York: McGraw-Hill Book Co., 1964

Glasstone, S., *Sourcebook on the Space Sciences*, Princeton, D. Van Nostrand Co. 1965

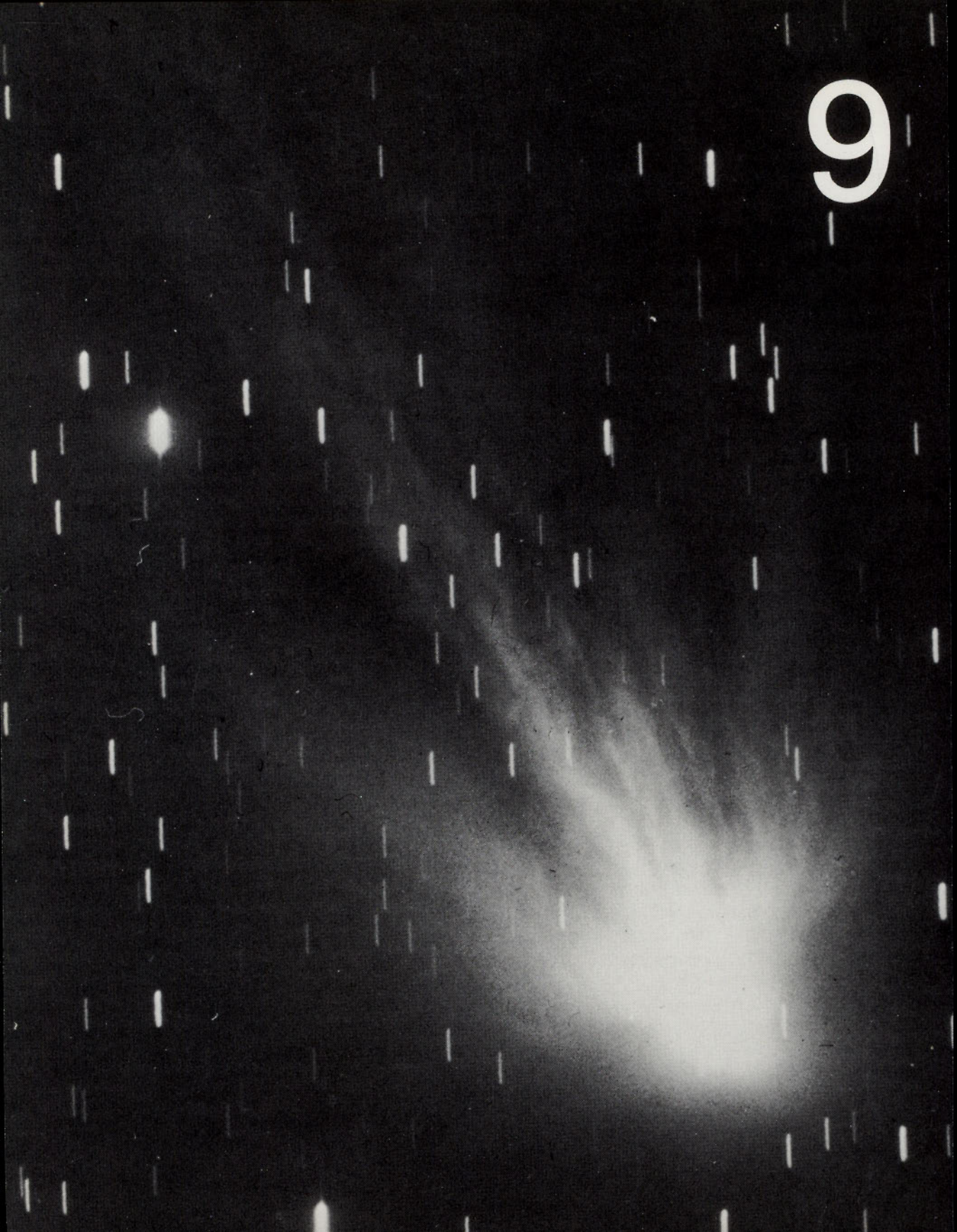
Kraus, John D., *Radio Astronomy*, New York: McGraw-Hill Book Co., 1966

Kuiper, Gerard P., ed., *The Atmospheres of the Earth and Planets*, Chicago: Univ. of Chicago Press, 1952

—— and B. M. Middlehurst, eds., *Planets and Satellites*, Chicago: Univ. of Chicago Press, 1962



Administration Building, Lowell Observatory, Flagstaff, Arizona. *(Lowell Observatory photograph)*


9

BETWEEN THE PLANETS

COMETS — METEORS AND METEOR STREAMS — METEORITES AND METEORITE CRATERS — THE INTERPLANETARY MEDIUM — THE PROBLEM OF THE ORIGIN OF THE SYSTEM

The description of the solar system continues with an account of the comets, meteors and the interplanetary medium. Comets revolve around the sun generally in highly eccentric orbits. Streams of meteoroids are products of the disintegration of comets. The meteors themselves make bright trails across the night sky when they chance to plunge into our atmosphere. Meteorites come through to the ground, and large ones may have momentum enough to blast out craters in the earth's surface. Some meteorites are believed to be fragments of shattered asteroids.

Comet Humason in 1962. *(Official U.S. Navy Photograph)*

COMETS

Characteristic of all comets, and the only conspicuous feature of many, is the foggy envelope of the coma. *The coma surrounds a rather small* nucleus *of frozen material. A comet's* tail *fans out in a direction away from the sun.*

9.1 Discovery of Comets

Comets are sometimes discovered at the observatories in examinations of photographs taken for other purposes. They have frequently been found by amateur astronomers who search for comets. The chief requirements for a comet hunter are a small telescope, much perseverance, and a catalog of nebulae and star clusters that could be mistaken for comets, although the motion of a comet among the stars will soon identify it. The western sky after nightfall or the eastern sky before dawn are the most promising regions for the search.

The report of the discovery of a comet, giving the comet's position and the direction of its motion among the stars, may be made to the Central Bureau of Astronomical Telegrams at the Copenhagen Observatory, or to the Smithsonian Astrophysical Observatory, Cambridge, Massachusetts, the central station in the United States for such astronomical news, which forwards the announcement to other observatories in the United States, Canada and Mexico. As soon as three positions of the comet have been observed at intervals of a few days, a preliminary orbit is calculated. Then it is usually possible to decide whether the comet is a new one or the return of a comet previously recorded, and what may be expected of it. Further observed positions provide data for the calculation of a more definitive orbit.

An average of 5 or 6 comets are picked up each year; usually two of these are returns of comets that have appeared before, and the rest are new ones. Comets that are bright enough to be visible without a telescope average less than one a year, and only rarely is a comet spectacular enough to attract the attention of those who are not astronomers.

A comet is designated provisionally by the year of its discovery followed by a letter in the order in which the discovery is announced; an example is Comet 1956 h. The permanent designation is the year (not always the year of discovery) followed by a Roman numeral in order of perihelion passage during that year; an example is Comet 1957 II. Many comets, especially the more remarkable ones, are also known by the name of the discoverer, or discoverers, or of the astronomer whose investigations of the comet entitle him by common consent to the distinction; Halley's comet (9.6) is an example.

Figure 9.1A
Comet 1969g, Tago-Sato-Kosaka. Note the detailed structure around the coma and in the tail. This plate was taken on December 28, 1969 when the comet was high out of the ecliptic in the southern hemisphere. *(Curtis Schmidt telescope of the University of Michigan at the Cerro Tololo Inter-American Observatory)*

Figure 9.1B
Comet Ikeya-Seki photographed by H. Giclas against the morning dusk. Notice the tail structure. *(Lowell Observatory photograph)*

9.2 The Orbits of Comets

A comet has no permanent individuality by which it may be distinguished from other comets. The only identification mark is the path it pursues around the sun. The orbits of 566 comets listed in J. G. Porter's catalog of 1961 are known with varying degrees of precision. The comets fall into two groups, with a somewhat indefinite dividing line between them:

1. Comets having *nearly parabolic orbits.* The orbits of many comets are so nearly parabolas that it is difficult to tell the difference, from the small portions of the orbits near the sun in which the comets can be seen. These orbits extend far out beyond the planetary orbits, and the undetermined periods of revolution are so long that only one appearance of each comet has thus far been recorded. In this sense they are "nonperiodic comets." The orbits are often highly inclined to the ecliptic. The revolutions of half of these comets are direct and of the other half are retrograde.
2. Comets having *definitely elliptic orbits.* The orbits of "periodic comets," having periods not exceeding a few hundred years, are more closely allied to the organization of the rest of the solar system. Although most of these orbits are also highly eccentric, they are more moderately inclined to the ecliptic, and the revolutions of the comets are mainly direct. The majority of all periodic comets are associated with the planet Jupiter (9.4).

9.3 Some Recently Observed Comets

Tables 9.I and 9.II contain selected lists of recently observed comets.

TABLE 9.I
Recent Bright Hyperbolic or Very Long Period Comets

COMET	YEAR	PERIHELION PASSAGE	PERIHELION DISTANCE (a.u.)	INCLINATION OF ORBIT
Arend-Roland	1956h	1957 Apr	0.32	120°
Mrkos	1957d	1957 Aug	0.36	94
Burnham	1959k	1960 Mar	0.50	160
Wilson	1961d	1961 Jul	0.04	24
Humason*	1961e	1962 Dec	2.13	153
Seki	1961f	1961 Oct	0.68	156
Seki-Lines	1962c	1962 Apr	0.03	65
Ikeya	1963a	1963 Mar	0.63	161
Alcock	1963b	1963 May	1.54	88
Pereyva**	1963e	1963 Aug	0.005	144
Tomita-Gerber-Honda	1964c	1964 Jun	0.49	162
Ikeya-Seki	1965f	1965 Oct	0.008	142
Mitchell-Jones-Gerber	1967f	1967 Jun	0.18	56
Tago-Sato-Kosaka	1969g	1969 Dec	0.47	76
Bennett	1969i	1970 Mar	0.54	90
Unnamed	1947n	1947 Dec	0.11	138
Unnamed	1948l	1948 Oct	0.14	23

* Comet Humason was not very bright but was a very interesting comet
** Comet Pereyva has estimated perihelion distances of as little as 0.003 A.U.

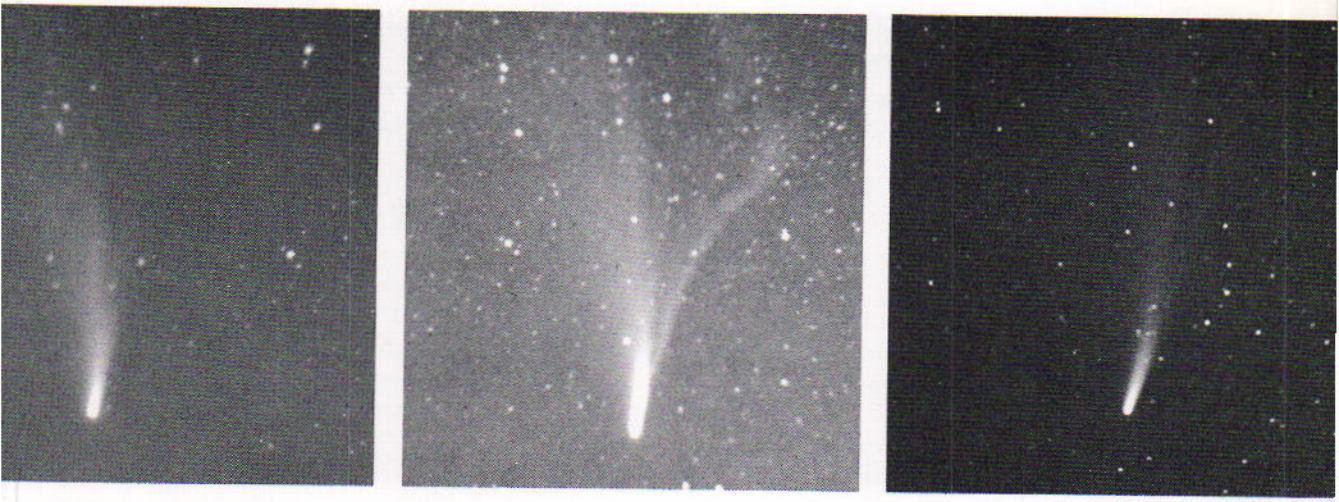
TABLE 9.II
Some Recently Observed Periodic Comets

COMET	FIRST SEEN	LAST SEEN	PERIOD (YEARS)	PERIHELION D, (a.u.)
Tempel-Tuttle	1366?	1965	33	0.99
Encke	1786	1967	3.3	0.34
Pons-Brooks	1812	1953	70.9	0.77
Crommelin	1818	1956	27.9	0.74
Pons-Winnecke	1819	1970	6.3	1.23
Faye	1843	1969	7.4	1.61
d'Arrest	1851	1963	6.7	1.37
Tempel 2	1873	1967	5.3	1.37
Holmes	1892	1964	7.3	2.35
Giacobini-Zinner	1900	1965	6.4	0.93
Daniel	1909	1964	7.1	1.66
Neujmin 1	1913	1966	17.9	1.54
*Schwassmann-Wachmann 1	1927	.....	16.1	5.51
Honda-Mrkos-Pajdusakova	1948	1969	5.2	0.56
Van Biesbroeck	1954	1965	12.4	2.41

* Comet Schwassmann-Wachmann 1 is due to encounter Jupiter soon, large orbit changes are expected.

The first table lists some of the brighter non-periodic comets for which parabolic orbits have been calculated. Note that about half of these comets have direct revolutions (i less than 90°) and the others retrograde (i greater than 90°). The brightest comets in this list were 1969 i, 1947 n,

Figure 9.3
Three views of Comet Bennett, taken on the first, fourth and tenth days of April, 1970, respectively and showing unusual activity in the tail. *(X-ray film photograph by W. Sopstyle, NASA Wallops Station)*



and 1948 l, which had the small perihelion distances of 0.54, 0.11, and 0.14 astronomical units, respectively. Two rather bright comets, 1956 h and 1957 d, were of considerable public interest because it is rather unusual for two comets bright enough to be observed with the naked eye to appear in the same year.

The second table lists some comets of short period and shows the dates through 1970 when their latest returns were observed. The dates do not appear for two entries because these comets, then having orbits of unusually small eccentricity, could be observed at every opposition with the sun.

The Schwassmann-Wachmann comet, 1925 II, revolves entirely between the orbits of Jupiter and Saturn. It was the first comet to be observed near its aphelion. Normally of the 18th magnitude, it is unique in exhibiting large and rapid flare-ups, for which the reason is unknown; an increase of 5 magnitudes or more in brightness has occurred within less than a day. Around the year 1974 the comet's orbit will be made even more nearly circular by close approach to Jupiter.

Oterma's comet had a nearly circular path around the sun between the orbits of Mars and Jupiter from 1943, the year of its discovery, until 1961. By prolonged close approach to Jupiter in the following few years, according to P. Herget, the eccentricity of the comet's orbit and the revolution period have been drastically increased.

9.4 Jupiter's Family of Comets

Two dozen or more comets of very short period have orbits closely related to the orbit of Jupiter. In each case the aphelion and one of the nodes are near the orbit of Jupiter, so that the comets often pass close to the planet itself. These are members of *Jupiter's family of comets.* Their close approaches to the giant planet result in such great perturbations of their orbits that the configuration of the family is not stable.

The periods of revolution of these comets around the sun are mostly between 5 and 9 years, averaging a little more than half of Jupiter's period with a gap between 5.5 and 6.2 years (similar to the asteroid gaps). The orbits are generally not much inclined to the ecliptic and the revolutions are all direct. The comets themselves are never conspicuous objects; a few become faintly visible without the telescope when they pass near the earth.

Encke's comet, discovered in 1786, was the first member of Jupiter's family to be recognized, in 1819. Its period of revolution, 3.3 years, is the shortest of any known comet. Its aphelion point has been drawing in toward the sun and is now a whole astronomical unit inside Jupiter's orbit. Forty-eight appearances of this comet have been recorded as observed between 1786 and 1968.

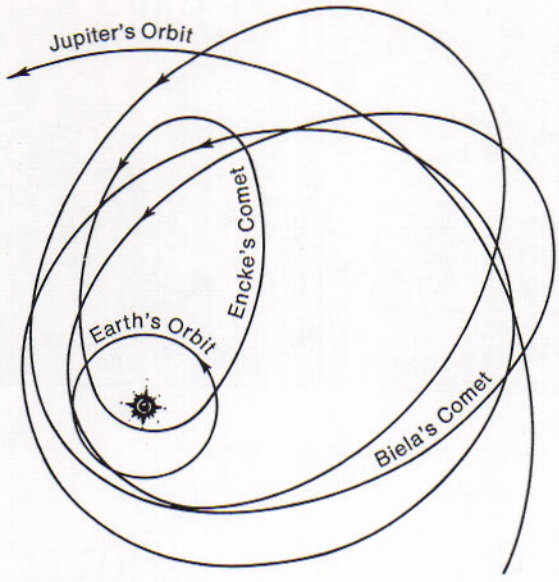


Figure 9.5
Orbits of four comets of Jupiter's family. Encke's comet has the smallest known orbit.

9.5 The Capture of Comets

The relation between Jupiter and its family of comets makes it seem probable that the planet has acquired the family by capturing some of the comets that chanced to be passing by. The low inclinations of the orbits and the direct motions of all comets in the family suggest that the process is selective. Comets having original orbits of sufficiently large perihelion distance and low inclination, so that they moved parallel to Jupiter for a time, are the most likely to be captured; their orbits may be made progressively smaller at successive encounters until they become members of the family.

The capture process was invoked in former times to account for all periodic comets. It was supposed that comets were casual visitors from outside the solar system and that only those captured by planets prolonged their stay with us. It now seems probable that all comets we see are natives of this system.

9.6 Halley's Comet

This famous comet, the first known periodic comet, is named in honor of Edmund Halley, who predicted its return. Halley calculated as a parabola the orbit of the bright comet of 1682 and noted its close resemblance to the orbits that he had similarly calculated for earlier comets of 1531 and 1607 from records of their places in the sky. Concluding that they were appearances of the same comet, which must therefore be moving in an ellipse, Halley predicted that it would return again in 1758. The comet was sighted that year according to prediction; it

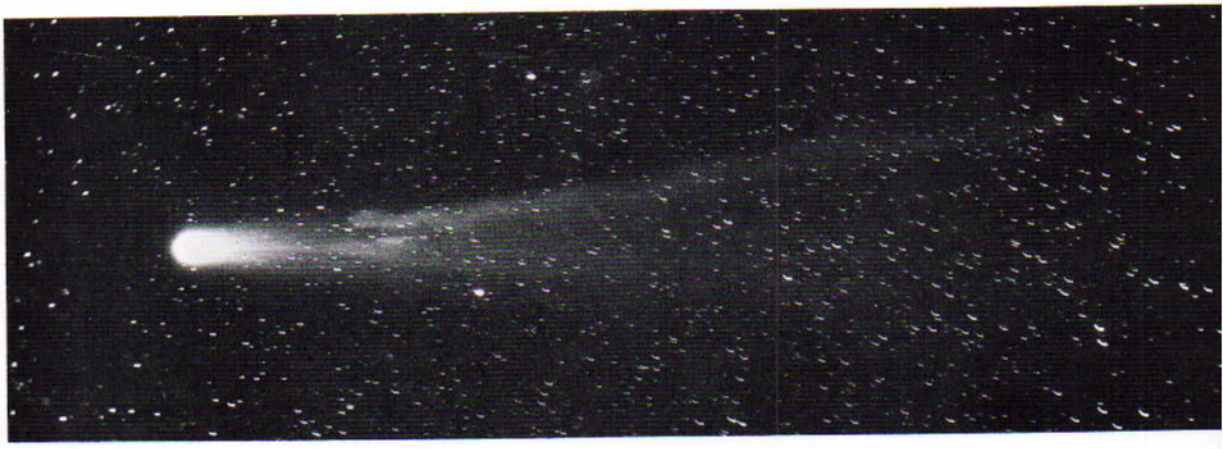


Figure 9.6A
Halley's Comet at its latest appearance in 1910. *(Lick Observatory photograph)*

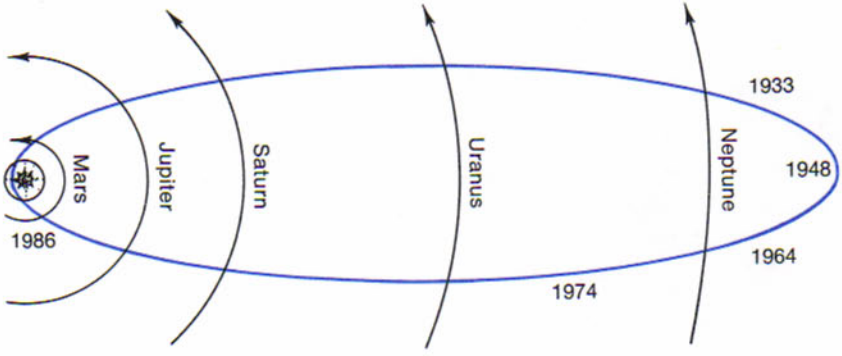


Figure 9.6B
Orbit of Halley's Comet, which passed aphelion in 1948 and will return to perihelion in 1986.

returned again in 1835 and 1910. Halley's comet is the only conspicuous comet having a period less than 100 years. The revolution is retrograde.

Twenty-eight observed returns of this comet have been recorded, as far back as 240 B.C. It was Halley's comet that appeared in 1066, at the time of the Norman conquest of England. The period has varied nearly 5 years meanwhile, because of disturbing effects of planets; the average interval between perihelion passages is 77 years. The comet, which has now passed its aphelion beyond the orbit of Neptune, is one of about ten comets forming a family associated with this planet. It will return to perihelion in 1986.

9.7
A Remarkable Group of Comets

This includes the great comets of 1668, 1843, 1880, 1882, and 1887. These comets passed unusually close to the sun and their orbits seemed practically identical. They were evidently parts of a single comet, which was disrupted at a previous close approach to the sun. The separate parts were dispersed in orbits of different sizes and, therefore, with periods of different lengths. All the orbits closely resemble the orbit of the original comet in the vicinity of the sun.

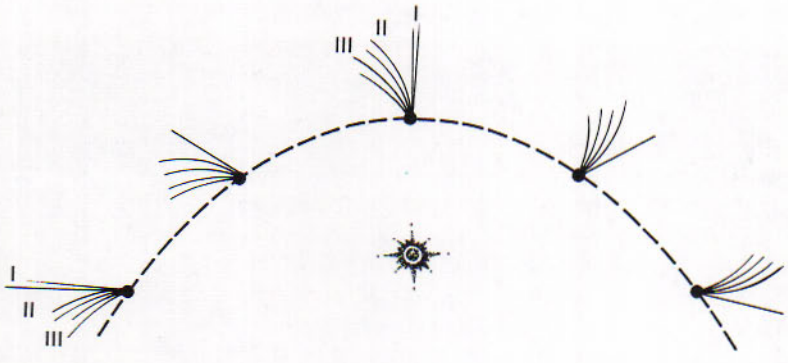
Comet 1882 II, the most spectacular of the group, was one of the brightest comets of modern times, plainly visible in full daylight. It passed through the sun's corona, within 500,000 km of the sun's surface, with a speed exceeding 1.5×10^6 km/h. Effects of tidal disruption during the close approach were evident soon afterward. The nucleus of the comet divided into four parts, which spread out in the direction of the revolution. These are expected to return as four comets between the 25th and 28th centuries.

9.8
Formation of a Comet's Tail

Comet tails are classified into three types: Type I, relatively straight tails consisting of ionized molecules; Type II, curved tails consisting of dust; and Type III, strongly curved fine dust tails. A comet's tail develops as it approaches the sun, the dust tails reflecting more light due to the inverse square law and the gaseous tails glowing as the sun's radiation becomes more intense. Comets generally show only dust tails until they come within three, or even two, astronomical units of the sun. Then a gaseous tail develops as well. Comet Mrkos exhibited both types of tails quite clearly.

The tail of a comet generally points directly away from the sun, being repelled by a force exceeding that of the sun's attraction. The repulsive

Figure 9.8
Tail of a comet directed away from the sun.



force is usually ascribed to the pressure of the sun's radiation, perhaps increased irregularly by collision with streams of high-speed particles emerging from the sun. By Kepler's third law the material of the tail revolves around the sun at a slower rate as it moves outward, falling more and more behind the head of the comet. Thus the tail is generally curved, the dusty part the more strongly because this material is likely to be repelled less rapidly than the gases of the tail.

$P^2 \propto a^3$

Meanwhile, some heavier meteoroidal products of the comet's disintegration may fan out behind the comet along the orbit plane, as they did conspicuously from a bright comet in 1957 (Fig. 9.10). In most cases the dust and meteoroidal material remain close to the orbital plane.

9.9 The Spectrum of a Comet

The spectrum of a comet is characterized by bright bands that are produced by gases set glowing by the sun's radiation. The gases are composed mainly of ionized molecules of carbon (C_2), methyne (CH), hydroxyl (OH), ammonia radicals (NH_2 and NH), and cyanogen (CN). These are rather unstable and are soon transformed into more durable molecules, such as carbon monoxide, carbon dioxide, and nitrogen, as they are driven from the coma into the tail. The unstable constituents of the coma are formed by action of sunlight on parent molecules of methane, ammonia, and water in the nucleus. These materials remain frozen there when the comet is far from the sun, and begin to evaporate as it approaches the sun.

Bright lines of sodium may become prominent in the spectrum, and lines of iron and nickel have been seen, when a comet is sufficiently heated by close approach to the sun. A faint replica of the solar spectrum also appears, showing that a comet shines partly by reflected sunlight. The dark-line spectrum is generally the only feature when a comet is more than 3 astronomical units from the sun.

Figure 9.9
Spectrum of Halley's Comet. *(Lowell Observatory photograph)*

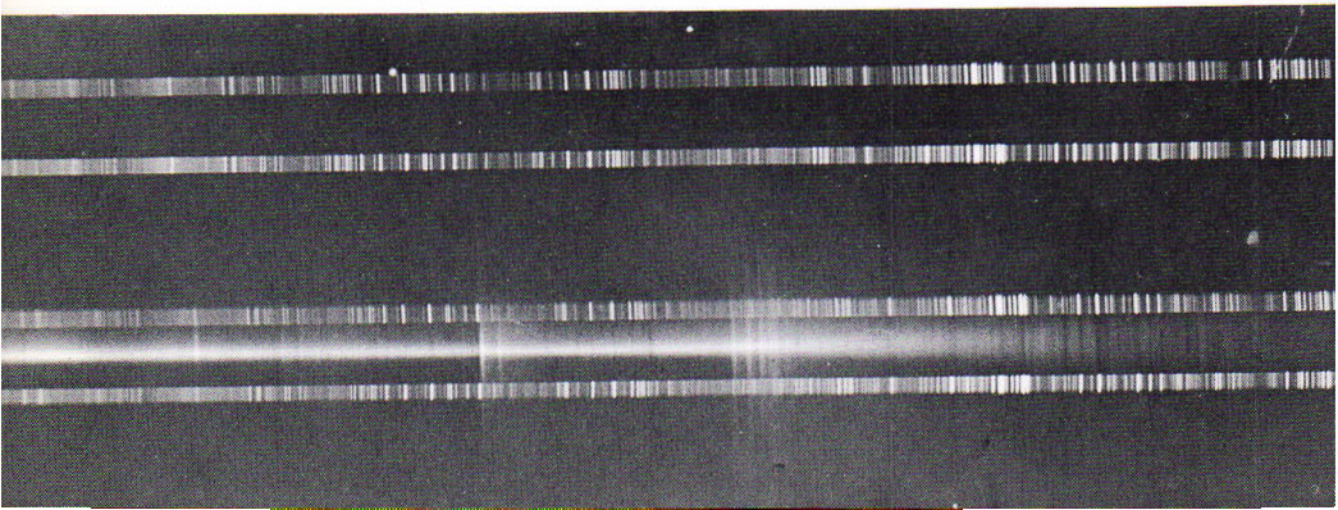


Figure 9.10
Comet Arend-Roland on April 27, 1957. Notice the sunward antitail of meteoric material. *(H. L. Giclas, Lowell Observatory photograph)*

9.10 Comets with Sunward Antitails

Comet Arend-Roland (1956 h) was of special interest because it showed a sunward antitail in addition to an ordinary tail that became 20° to 30° long to the unaided eye. This nonperiodic comet has a retrograde motion in an orbit inclined 60° to the ecliptic. It reached perihelion on April 8, 1957, at one third of the earth's distance from the sun. On April 25 the earth passed through the plane of the comet's orbit. For a week around that date the antitail was conspicuous; it progressed from a stubby fan to a narrow spike as long as 15° on April 25, then reverted to a short, broad fan, and soon disappeared.

The sunward antitail was produced by meteoric material fanning out in the plane of the comet's orbit. The material in the fan was too diffuse to be easily visible far from the comet's head when presented broadside to us. When the layer was seen nearly on edge, it appeared as a long, narrow spike. Whipple has pointed out that Comet 1862 III similarly spread meteoroidal material along the orbit plane to produce the stream of the Perseids, or August meteors.

Comet Seki-Lines (1962 c) also showed an antitail. This nonperiodic comet became a bright object to the unaided eye because its head passed on April 1, 1962 only 4 million km from the sun's photosphere. In Alan McClure's unusual photograph the ordinary tail appears 15°, or 50 million km, in length and the antitail is a stubby fan.

9.11 The Nature of a Comet

A comet's nucleus is a very porous structure of ices having meteoric material embedded in them, according to the theory by F. L. Whipple. The ices are chiefly of methane, ammonia, and water. The meteoric material is generally in small particles and is composed of iron, nickel, calcium, magnesium, silicon, sodium, and other elements. The nucleus of a comet is only a mile or so in diameter.

Some of the ices evaporate at each approach to the sun. The gases issue explosively into the coma, carrying the meteoric fragments with them, and are then swept out through the tail or scattered along the orbit. Whipple estimates that a comet evaporates $\frac{1}{200}$ of its mass around each perihelion passage. That it is not dissipated more rapidly is because the surface of the nucleus becomes increasingly gritty; the meteoric material protects the ices underneath until the structure collapses. After many returns, the comet disappears. The meteoric material originally embedded in the ices continues to revolve around the sun as a meteor stream.

An important feature of the theory accounts for the spiral motions of some comets. The issuing of gas from the sunward side of the comet causes a jet propulsion action on the nucleus in the direction of the sun. If there is delay in setting the evaporation going, the comet's rotation swings a component of the propulsion along the orbit. Where the rotation is in the opposite direction to the revolution, the rotation diminishes the speed in the orbit, so that the comet begins to spiral in toward the sun. This effect seems to explain the hitherto perplexing behavior of Encke's comet (9.4). The opposite effect, where the comet rotates in the same direction as its revolution, is illustrated by d'Arrest's comet.

9.12 The Origin of Comets

Until 1577 comets were generally believed to be an awesome atmospheric phenomenon. Later comets were believed to be interlopers from interstellar space, some being captured to form the Jupiter family as we have seen. However, as the techniques of celestial mechanics became more refined and as the observational material became more accurate, it was discovered that no comet orbit was clearly hyperbolic. If all the comets belong to our solar system then we must explain their origin and this explanation must fit into any theory about the origin of the solar system.

J. Oort has proposed an icebox theory to explain the comets. In this theory there is a belt of debris stretching out as far as one light year from the sun. The debris is the remnant of the nebula that formed the solar system. Since the nebula formed in the interstellar medium, the primary constituents are frozen molecules and, consequently, dust. Clumps of this material form and probably enlarge by accretion.

The clumps are moving at velocities on the order of only 0.1 km/sec. A transient disturbance is all that is required at these distances and speeds to alter the direction and start the clump moving inward on a

long journey toward the sun. The disturbing force could be a passing star or, more likely, the sum of forces from the major planets. If this theory is correct and these clumps are remnants of the solar nebula, we are nearly at the point where we can collect a piece of this nebula. While not easy, it will be possible to have a spacecraft intercept and analyze the contents of a parabolic comet or better still, return it to earth for analysis.

METEORS AND METEOR STREAMS

Meteoroids *are stony and metallic particles revolving around the sun, which become separately visible as meteors only when they plunge into the earth's atmosphere. Melted and vaporized then by impact with the air molecules, they produce luminous trails across the night sky. Unusually bright ones are called* fireballs. *In addition to the sporadic meteors that seem to be moving independently, there are swarms and streams of meteoroids revolving around the sun, which cause "meteor showers" when they encounter the earth. A number of streams are identified with the orbits of comets from which they originated.*

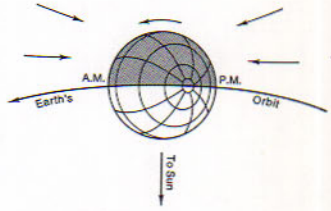


Figure 9.13
Meteor flights are swifter after midnight. In the morning we are on the forward side of the earth in its revolution.

9.13 The Influx of Meteors

The total number of meteor trails brighter than visual magnitude +5 over all the earth's surface is determined by G. S. Hawkins as 90 million in a 24-hour period. The meteors producing these trails add several metric tons a day to the earth's mass. The number of trails visible to a single observer in this period is usually very small. The recorded hourly rate is likely to increase somewhat through the night. In the evening we are on the following side of the earth in its revolution, where we are protected except for the meteoroids that overtake the earth. In the morning we are on the forward side, more exposed to the incoming meteoroids.

The swiftness of the meteor flights across the sky also increases through the night. At our distance from the sun the speeds of the meteoroids in their eccentric orbits approach the parabolic value of 42 km a second (7.25). Their speeds relative to the earth accordingly range from 42 minus km/sec for meteors that overtake us to 42 plus km/sec for head-on collisions, with some addition for the earth's attraction.

1 metric ton = 1000 kg

9.14 Meteor Trails and Trains

The *trails* of meteors are the bright streaks they produce in the flights through the air. The faster meteors generally appear in the photographs at a height of about 130 km and disappear at 95 km; the slower ones begin at 95 km and have been followed to a height of 40 km.

As its flight is resisted by the air molecules, the meteor is heated to incandescence and fused. The brightness of the trail depends on the kinetic energy of the meteor, that is, on its mass and the square of its velocity. The brightness also depends on the density of the air and the compactness of the meteor itself. Many meteoroids, presumably from the outer parts of comet nuclei, are extremely porous and fragile; these are likely to be shattered in their flights through the air and to collapse with bursts of added brightness.

In addition to their momentary trails, the meteors sometimes leave bright *trains* along their paths. These are cylinders of expanding gases, which remain visible from a few seconds to generally not longer than half an hour, and are often twisted by air currents. A rarer type of meteor train appears in the *noctilucent clouds* recorded especially by Canadian observers. They are dust trains in the wakes of very bright fireballs, and are best seen in the twilight when they are high enough to be illuminated by sunlight.

9.15 Meteors as Members of the Solar System

When the velocity of a meteor as it enters the atmosphere is known, it is possible to calculate its orbit around the sun and to determine among other results whether the meteor was a member of the solar system or came from outside it. Measuring the linear speed and the actual direction of the trail involves a parallax problem requiring positions of the trail among the stars as observed from two stations. The Harvard program, directed by Whipple, employs for this purpose two Baker-Schmidt cameras, having apertures of 30 cm and focal lengths of 20 cm, at stations in New Mexico 35 km apart. A similar pair is operated north of Edmonton by Canadian observers. The two cameras are directed toward the same point in space 80 km above the ground. The trail of any bright meteor that passes through this region is recorded by both cameras and is interrupted in the photographs at intervals of a small fraction of a second by rotating shutters in front of the films so as to facilitate the timing.

Several thousand trails have been doubly photographed in New Mexico by such procedure, and many of these have been analyzed for velocities and orbits under the leadership of L. G. Jacchia. Some of the sporadic meteors had orbits of the Jupiter-family comet-type (9.4) with direct motions and low inclinations to the ecliptic. Some others were of the long-period type with random directions and inclinations. None of these had a reliably determined hyperbolic orbit. Like the meteors in streams the sporadic meteors had elliptical orbits and so were members of the solar system.

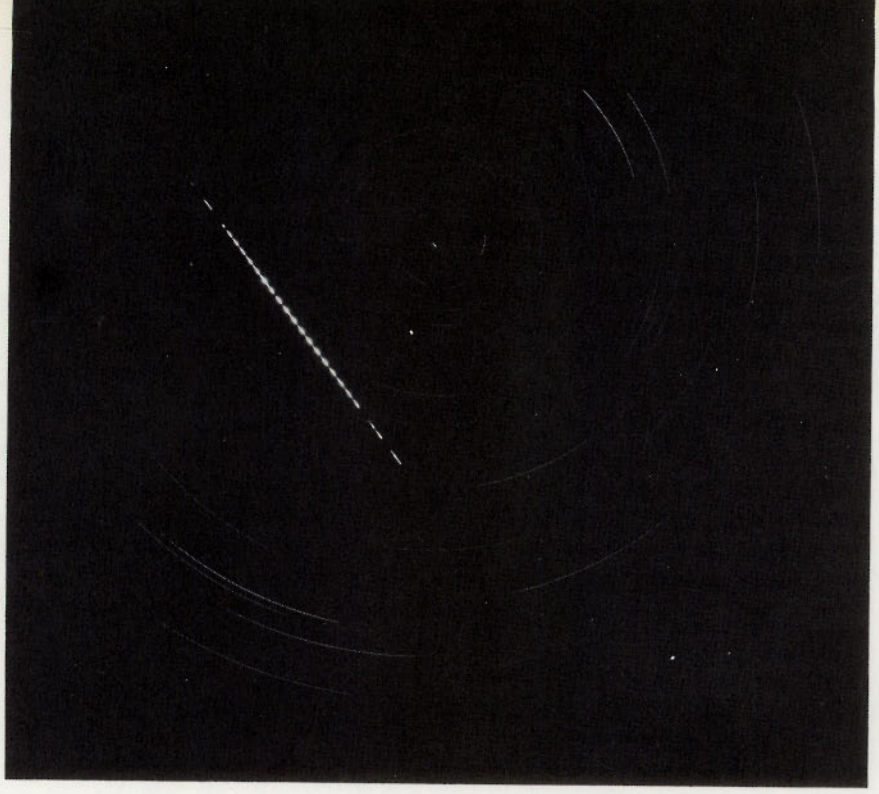


Figure 9.15
Trail of a very bright meteor. Polaris is the short bright star trail near the center of the concentric trails. *(Smithsonian Astrophysical Observatory photograph)*

The speed by itself is enough to determine the status of the meteor in this respect. If the speed on entering the atmosphere, with allowance for the motions and attraction of the earth, is as great as 42 kilometers a second, the velocity of escape from the sun at the earth's distance, the meteor may have come from outside the solar system. If the speed is less than this critical value, the meteor is a member of the system.

The echoes of radio beams returned from the ionized trails of meteors have been observed at Ottawa to determine the speeds of meteors. The beams are emitted either continuously or in pulses, and the echoes are automatically recorded. D. W. R. McKinley's records of more than 10,000 meteors down to the 8th visual magnitude include not one where the original speed was certainly as great as the velocity of escape. Thus the fainter meteors as well as the brighter ones appear to be members of the solar system.

9.16 Spectra of Meteor Trails

The spectra of meteor trails have been photographed in considerable numbers in recent years, particularly by P. M. Millman in Canada and J. A. Russell in California. These objective prism and grating spectra give an idea of conditions to which the meteors are subjected in their flights through the air and of the chemical compositions of the meteors themselves.

A meteor spectrum is an array of bright images of the trail in the various wavelengths of the emitted light. The character of the lines depends on the temperature to which the meteor is raised by the air resistance. The spectrum of a swifter meteor is likely to contain prominent H

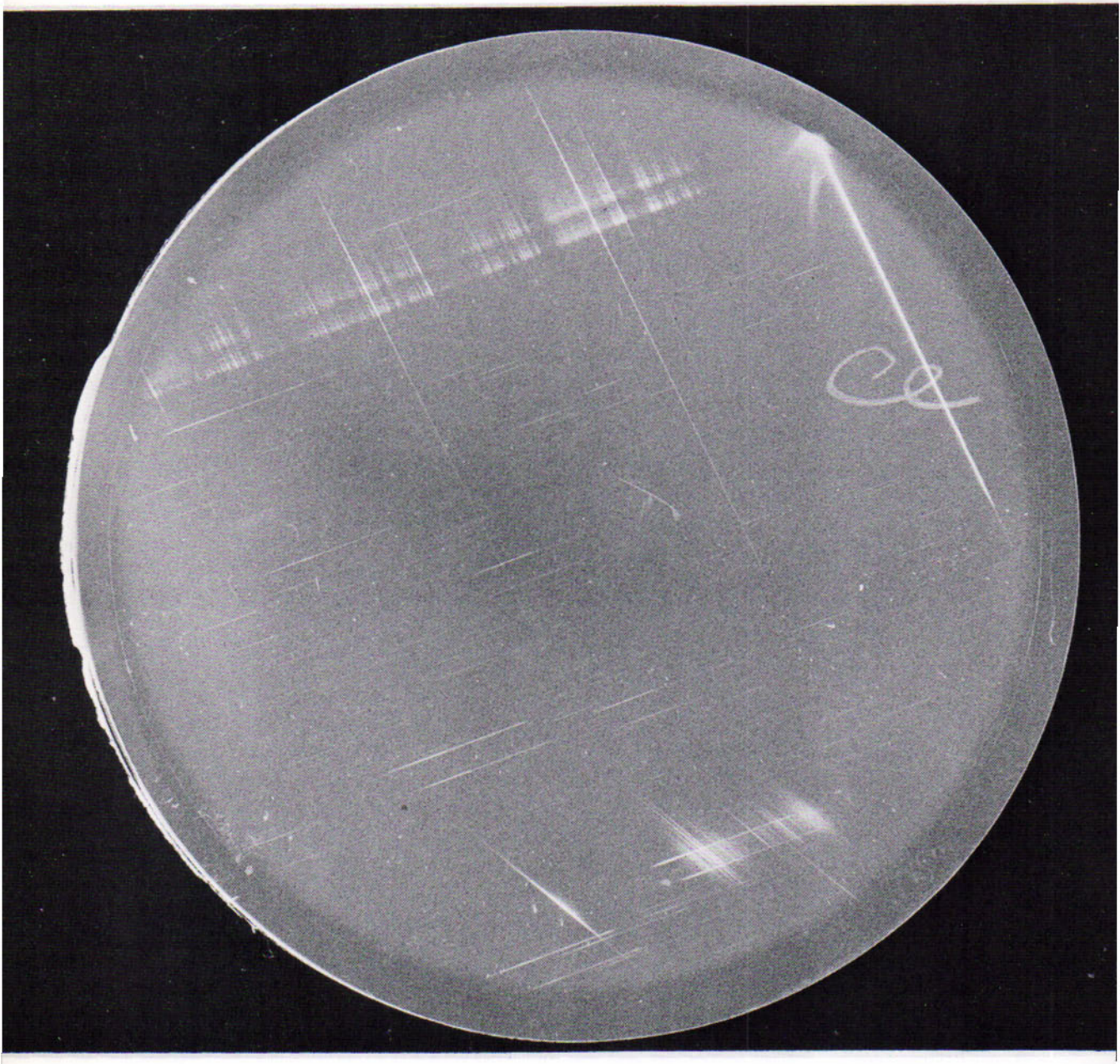


Fig. 9.16
Spectrum of meteor. *(Photograph by G. Harvey, NASA Langley Research Center)*

and K lines of ionized calcium. In the spectrum of a very bright Perseid meteor of 1968, G. Harvey's photograph (Fig. 9.16) shows more than 100 bright lines, from which he identified the presence of nine neutral elements: Fe, Mn, Ca, Si, Al, Mg, Na, N, and H, and also four ionized elements: Fe, Ca, Si, and Mg. Nitrogen and some oxygen lines are frequently produced in the air that is itself heated by the meteor's flight.

9.17 Meteoroid Streams and Showers

Multitudes of meteoroids revolving together around the sun constitute a *meteoroid swarm.* When the meteoroids are considerably extended over a similar part of their orbits in the sun's vicinity as is often the case, they form a *meteoroid stream.*

Because the meteoroids in a stream are moving in nearly parallel paths when they encounter the earth, their bright trails through the air are nearly parallel. Just as the rails of a track seem to diverge from a distant point, so the trails of meteors in a shower appear to spread from a point or small area in the sky. The *radiant* of a meteor shower is the vanishing point in the perspective of the parallel trails; it is located by extending the trails backward.

Meteor showers and the streams that produce them are generally named after the constellation in which the radiant is situated, or else after the bright star near the radiant at the maximum of the display. Examples are the Perseids and the Beta Taurids. The place of the radiant in the sky drifts from day to day during the progress of the shower, as the earth keeps changing its position with respect to the orbit of the stream. A stream is sometimes named after the comet with which it is associated. Thus the Draconids are also known as the Giacobinids.

Figure 9.17
Leonid meteor shower of 1966. Notice the well-defined radiant, and a non-shower meteor (upper left, the longest of all trails in the picture). *(Photograph by D. McLean)*

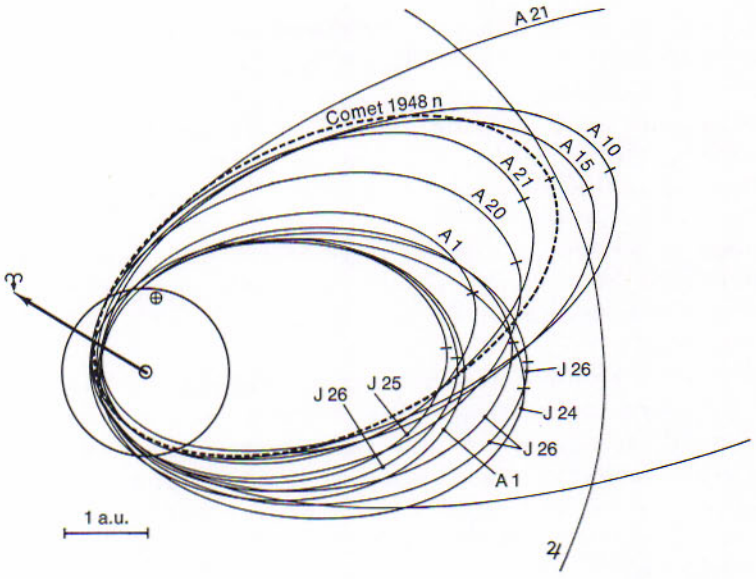


Figure 9.18
Orbits of Alpha Capricornid Meteors. The trails were photographed on the days in July and August shown on the orbits, which are projected on the ecliptic plane. *(Diagram by Frances W. Wright, L. G. Jacchia, and F. L. Whipple, Harvard Observatory Meteor Program)*

9.18 The Orbits of Shower Meteors

The orbits of a dozen meteors of the Alpha Capricornid stream (Fig. 9.18) show how widely the aphelia of a stream may be scattered in space. These are determined from double-station trails photographed by the Harvard meteor program between July 16 and August 22. This example may be somewhat extreme because there may be more than one stream involved, as the investigators explain. Viewed as a single stream, the shower at maximum display has its radiant northeast of Alpha Capricorni, and it may be associated with Comet 1948 n.

A meteor shower occurs only when the orbit of the swarm intersects or passes near the earth's orbit and when the swarm and the earth arrive together at the intersection. The position of this point on the earth's orbit determines the date of the shower. If the swarm is condensed, the interval between showers depends on the period of revolution of the swarm and its relation to the length of the year. Streams may spend more than a year in crossing the earth's orbit, or may be so scattered, as in the case of the Perseids, that we encounter some of their members every year. Some streams that formerly produced spectacular showers no longer do so because their orbits have been altered by disturbing effects of planets.

9.19 Some Noteworthy Meteor Showers

The Perseids furnish the most conspicuous and dependable of the annual showers. Their trails are visible through 2 or 3 weeks, with the greatest display about August 12. Their orbit is inclined 65° to the ecliptic plane and passes near the orbit of no other planet. Next in order of numbers and annual reliability are the Orionids and the Geminids. Among the

less frequent showers, the stream of the Leonids, revolving around the sun in a period of 33 years, produced in 1833 and 1866–67 the most spectacular showers of modern times. At later returns, however, the Leonids appeared only in sprinkles until 1967, when a spectacular shower occurred late one morning over Mexico, the United States and Western Canada.

The Draconids gave the most remarkable showers of the present century. They revolve in the orbit of the Giacobini-Zinner comet, a member of Jupiter's family (9.4) first noticed in 1900; the comet itself revolves around the sun in a period of $6\frac{1}{2}$ years. The main body of the swarm and the earth have arrived together at the intersection of their paths in October at intervals of 13 years.

After nightfall on October 9, 1933, the skies of western Europe were streaked with meteor trails coming from the head of Draco. Before night began in America, that shower was over. America's turn came on October 9, 1946, 2 weeks after the comet had passed by. For a few hours on that evening there were fine displays wherever the skies were clear. At the height of the shower some observers counted as many as 100 trails a minute despite the bright moonlight, which obscured the fainter ones. A recurrence of the Draconid shower on the night of October 9, 1959, was watched for, but only a few trails were observed. Indeed, an abundant shower was not expected, because a close approach to the planet Jupiter had altered the orbit of the stream.

9.20 Daytime Meteor Showers

Observers at the Jodrell Bank Station of the University of Manchester in England have made systematic surveys of meteor activity by means of radar echoes. With this technique they are unhampered by cloud or daylight. They not only have investigated radiants conforming closely to previous results of evening observers, but also have discovered radiants of showers observable only in the daytime.

The daytime activity they have observed includes a number of showers of very short period. Among the daytime showers of June are the Zeta Perseids, the Arietids, and the Beta Taurids. An unexpected display of Draconids was recorded for 2 hours on the afternoon of October 9, 1952. Evidently an advance contingent of the meteor stream was encountered by the earth half a year before the associated comet was scheduled to reach the intersection with our orbit.

9.21 Association of Meteoroid Streams and Comets

The Draconids are by no means unique in their association with a comet. When meteoroidal material is released from the ices of a comet's nucleus (9.10), the material may become a meteoroid stream extending along the comet's orbit. Table 9.III, which is taken from more extended data by F. L. Whipple and G. S. Hawkins in the *Handbuch der Physik*, volume 52, 1959, lists major meteor showers and the comets associated with them.

The table gives in each case the date of maximum display in universal time, the position of the radiant so that it may be located in the star maps, and the name of the parent comet when it is known.

Two meteor showers are associated with Encke's comet. They are the daytime Beta Taurids of June and the Taurids of November; they come from the same extended stream that the earth encounters before and after the perihelion passage of the stream. Two showers are associated less certainly with Halley's comet; they are the Eta Aquarids of May and the Orionids of October. The radiants assigned in the table to the Delta Aquarids and the Taurids are each the means of two radiants several degrees apart.

A remarkable case of the dispersal of a comet into a meteor stream is that of Biela's comet, a member of Jupiter's family having a period of 6½ years. At its return to the sun's vicinity in 1846, the comet had divided into two. The divided comet reappeared with greater separation in 1852 and has not since been seen. Traveling in the orbit of the lost comet, the Andromedids, or Bielid meteors, gave fine showers in 1872, 1885, and 1898, and nearly disappeared thereafter. Perturbations by planets have shifted the orbit of the main body of the swarm, according to Hawkins, so that it now passes 3.2 million km from the earth's orbit at the closest point.

TABLE 9.III
Meteor Showers and Associated Comets

SHOWER	MAXIMUM DISPLAY (UT DATE)	RADIANT AT MAX. (EQUINOX OF 1950)			ASSOCIATED COMET (AND NOTES)
		R. A.		*Decl.*	
Quadrantids	Jan. 3	15^h	20^m	+48°	
Lyrids	Apr. 21	18	0	+33	1861 I
Eta Aquarids	May 4	22	24	0	Halley(?)
*Arietids	June 8	2	56	+23	(= Delta Aquarids)
*Zeta Perseids	June 9	4	8	+23	
*Beta Taurids	June 30	5	44	+19	Encke
Delta Aquarids	July 30	22	36	−11	(two streams)
Alpha Capricornids	Aug. 1	20	36	−10	1948 n
Perseids	Aug. 12	3	4	+58	1862 III
Draconids	Oct. 10	17	36	+54	Giacobini-Zinner
Orionids	Oct. 22	6	16	+16	Halley(?)
Taurids	Nov. 1	3	28	+17	Encke (two streams)
Andromedids	Nov. 14	1	28	+27	Biela
Leonids	Nov. 17	10	8	+22	Temple
Geminids	Dec. 14	7	32	+32	
Ursids (Ursa Minor)	Dec. 22	13	44	+80	Tuttle

* Shower in daytime.

METEORITES AND METEORITE CRATERS

A meteorite *is a mass of stony or metallic material, or both, which survives its flight through the air and falls to the ground. The fall is frequently accompanied by flashes of light and by explosive and roaring sounds, effects that may be magnified in the reports of startled observers. Most meteorites are found on the ground or buried a little way in it. A few very large ones have blasted out the pits we call* meteorite craters. *Many meteorites are believed to be fragments of shattered asteroids.*

9.22 The Fall of Meteorites

Meteorites are often found in groups. Either they entered the atmosphere as a compact swarm, or they arrived as single bodies which were shattered in the air by the shock of their reduced speed or when they struck the ground. The individuals of a large group may be distributed over an elliptical area several miles long in the direction of the forward motion. Whenever meteorites are picked up almost immediately after landing, they are likely to be cool enough to be handled comfortably. In their brief flights through the air the heat has not gone far into their cold interiors, and the melted material has been mostly swept away from their surfaces.

Individuals from nearly 1600 falls have been recovered, according to F. C. Leonard's catalog of meteoritic falls. The majority of these were not seen to fall, but their features definitely distinguish them from the native rocks. Meteorites are generally named after the locality in which they were found; examples are the Canyon Diablo, Arizona, meteorites, and the Willamette, Oregon, meteorite. Collections are exhibited in the Chicago Natural History Museum, the American Museum of Natural History, New York City, the Arizona State University at Tempe, and many other places.

9.23 External Appearance of Meteorites

Characteristic of most meteorites is the thin, glassy, and usually dark crust. It is formed from the fused material which was not swept away and which hardened quickly near the end of the flight through the air. The surface is often irregular, being depressed in places where softer materials melted faster, and raised in other places where drops of molten spray fell and hardened.

If the meteorite was shattered shortly before reaching the ground, the fragments have irregular shapes. If the individual had a longer flight through the air, it was shaped by the rush of hot air. Some individuals

turned over and over as they fell and were rounded. Others kept the same orientation and became conical, a common form.

9.24 Structure and Composition

Meteorites are essentially of two kinds, the stones (aerolites) and the irons (siderites). There are gradations between them (siderolites) from stones containing flecks of nickel-iron to sponges of metal with stony fillings. Inside their varnish-like crusts the *stony meteorites* are often grayish, having a characteristic granular structure that serves to establish their celestial origin. The rounded granules are crystalline; they are mainly silicates, of which the most common are iron-magnesium silicates, such as olivine and enstatite, similar to those in our igneous rocks. Quartz is very rarely a constituent.

1 ton = 907.2 kg
1 oz = 28.3 g
1 lb = 0.45 kg

The largest known stony meteorite, weighing somewhat more than 907 kilograms, fell on February 18, 1948, in Furnas County, Nebraska (Fig. 9.24A); it is preserved at the University of New Mexico. Another unusually large individual, now in the Chicago Natural History Museum, fell at Paragould, Arkansas, in 1930; it weighs 340 kilograms. Stony meteorites do not often have the large dimensions of many irons, partly because they offer less resistance to fracture and erosion.

Iron meteorites are silvery under their blackened exteriors; they are composed mainly of nickel-iron alloys, especially kamacite and taenite. Where they occur in crystal forms, a characteristic pattern of intersecting crystal bands (Fig. 9.24B), parallel to the faces of an octahedron, may be etched with dilute nitric acid on a plane polished section.

Micrometeorites are so minute that they are not greatly altered when they enter the atmosphere. The presence of micrometeoroids in interplanetary space is indicated by the scattered sunlight of the zodiacal light (9.31). Meteoritic dust is also produced by the larger meteoritic bodies themselves, when they are either crumbled by collision or are partly melted to form oxidized droplets in the air.

Figure 9.24A (left) Furnas County, Nebraska, meteorite. A stony meteorite weighing at least a ton. *(Official photograph of the Institute of Meteoritics, University of New Mexico)*

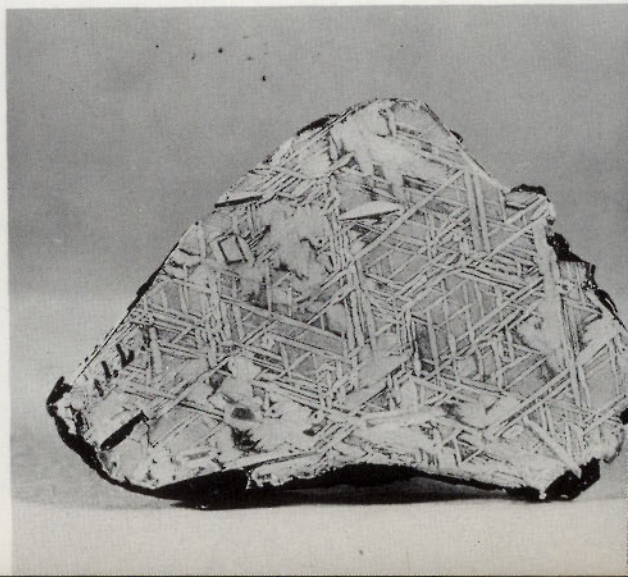


Figure 9.24B (right) Etched section of Knowles, Oklahoma, meteorite. The banded pattern characterizes most iron meteorites. *(American Museum of Natural History, New York)*

Figure 9.25
Ahnighito meteorite on scale at the Hayden Planetarium, New York City. *(American Museum—Hayden Planetarium)*

9.25 Large Iron Meteorites

About 35 known individual meteorites weight more than 907 kilograms (2000 lbs.), according to Leonard. Almost all are nickel-iron meteorites. With the exception of the Furnas County stone, two stony-irons, and a 1800 kilogram individual piece from the 1947 Siberian fall, their falls were not observed. The two largest irons on record are the Hoba meteorite and the Ahnighito meteorite (Fig. 9.25).

The Hoba meteorite lies partly buried in the ground in the Grootfontein district, Southwest Africa. Its rectangular exposed surface measures 2.8 × 3 meters and its greatest thickness is more than one meter; its weight is unknown. The Ahnighito meteorite is the largest of three irons that the explorer R. E. Peary found near Cape York in northern Greenland, in 1894, and brought back to New York. It measures about 3.5 × 2.1 × 2 meters and weighs a little more than 31,000 kilograms. This meteorite is exhibited in the American Museum—Hayden Planetarium, New York City, on the scales with which it was first accurately weighed, in 1956.

The Willamette meteorite, also in the Hayden Planetarium, is the largest found in the United States. A conical mass of nickel-iron, weighing about 13,600 kilograms, it was discovered in 1902 south of Portland, Oregon. Large cavities in the base of the cone, which lay uppermost for a long time, were formed by weathering. Three large iron meteorites, each weighing more than 9070 kilograms, were found in Mexico. They are the Bacubirito (26,300 kg), the Chupaderos (19,000 kg, in two pieces that fit together), and the Morito (10,000 kg). The last two may be seen in the School of Mines, Mexico City.

9.26 Two Siberian Falls

Two Siberian Falls of large meteorites in the present century have attracted much attention. The first occurred on June 30, 1908, in the forested area of the Tunguska River in central Siberia. The meteorite appeared as a brilliant fireball in full daylight from hundreds of miles away. Its mass is unknown. The trees were felled within a radius of 30 to 50 kilometers; they lay without bark and branches, and with their tops pointing away from the center of the area. Many craters were blasted out near the center, the largest one 50 meters in diameter. Although larger remnants of the Tunguska meteorite have not been recovered, soil samples from the region are said to contain microscopic chips and spherules of nickel-iron dust.

The second fall occurred on February 12, 1947, over a 2.5-square km-area on the western spurs of the Sikhote-Alin mountain range in southeast Siberia. The original body broke into many pieces in the atmosphere and came down as an "iron rain," the initial pieces being further fractured as they struck the ground. The area was pitted with 200 small holes and larger craters, the largest one 30 meters in diameter. The field was strewn with an estimated 91,000 kg of nickel-iron fragments. An interesting feature of the fall, described by E. L. Krinov, was a wide gray band of dust in the wake of the brilliant meteorite; the dust remained visible for several hours in the daytime sky and gradually settled to earth as microscopic droplets.

9.27 The Barringer Meteorite Crater

This famous crater, located near Canyon Diablo in northeast Arizona, is a circular depression in the desert 1280 meters across and 174 meters deep; the depth is measured from the rim, which rises to an average height of about 40 meters above the surrounding plain. Thirty tons of meteoritic iron have been picked up within a distance of 10 kilometers around the crater. The largest individual unit, weighing more than 640 kg, is exhibited in the museum on the north rim of the crater. Samplings in a survey directed by J. S. Rinehart indicate that the total amount of crushed meteoritic material around the crater weighs 10,900 metric tons.

1 metric ton = 1000 kg = 2204.6 lb.

The rocks below the crater floor are crushed to a depth of several hundred feet and give evidence of having been highly heated. Millions of kilograms of limestone and sandstone were displaced outward to form the crater wall, and loose blocks of rock weighing up to 6300 metric tons lie around the rim.

The meteorite that produced the Barringer crater is estimated to have had a diameter of 60 meters and to have weighed at least 900,000 metric tons. When an object of this size encounters the earth, its speed is not greatly reduced by air resistance. When it strikes the ground, at least

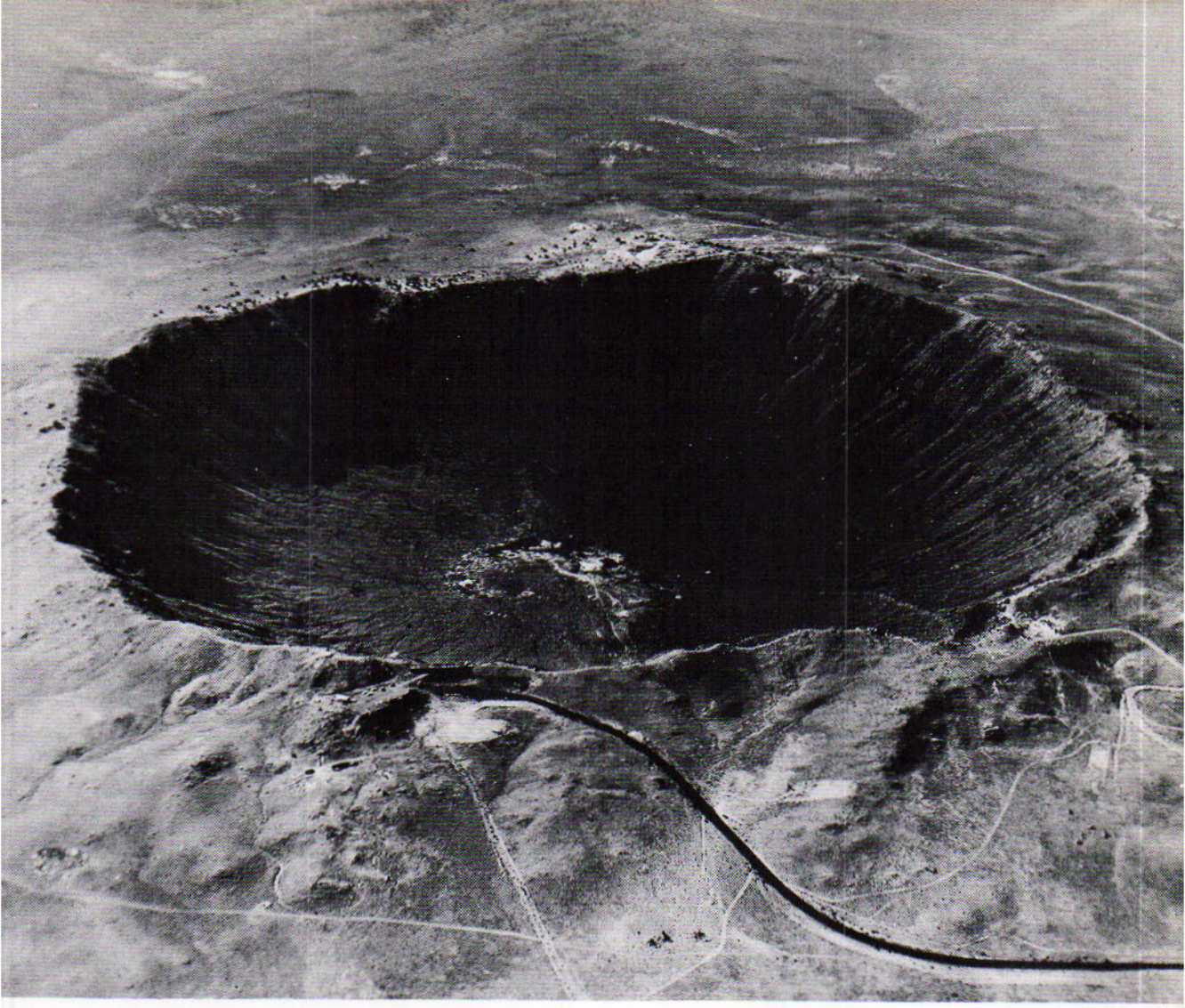


Figure 9.27
Barringer meteorite crater near Winslow, Arizona. *(Meteor Crater Society photograph)*

the outer parts of the meteorite and the ground in contact are intensely heated and fused. The gases expand explosively, blasting out the crater and scattering whatever may be left of the meteorite over the surrounding country.

As fragments of colliding asteroids (8.20), the meteorites would usually revolve around the sun from west to east, as do all asteroids, so that they would be overtaking the earth when they encounter it. Thus the majority of crater-forming meteorites strike the ground with relative speeds not great enough to cause the vaporization of their entire masses.

Only an occasional meteorite would be likely to have its orbit so greatly altered that it could land on the earth's advancing hemisphere. Then its higher relative speed might cause an impact of such violence as to leave only microscopic remnants of the intruder for the collectors. The Tunguska fall of 1908 is cited as an example.

9.28 Other Meteorite Craters

More than a dozen other craters or groups of craters in various parts of the world have been definitely considered to be of meteoritic origin. Among the largest is the Wolf Creek crater in West Australia, having a diameter of 850 meters at the bottom and a depth of 50 meters. These were formed by impacts of meteorites within the last million years. The scars of earlier falls have gradually vanished as readily recognized features of the landscape. Erosion has reduced the heights of their rims, and sedimentation has filled their depressions. A few "fossil craters" scarcely noticeable on the ground have been observed in aerial photographs. Examples are the Brent and Holleford craters in Ontario, described by C. S. Beals. Their diameters are 3 and 2.5 kilometers, respectively, and their ages are estimated as 500 million years.

The geologist R. S. Dietz has undertaken to increase the list of recognized meteorite craters, where the surface structure itself is inconclusive and where no meteoritic remnants have been reported. He points out in *Scientific American* for August 1961 that the shock generated by the impact of a meteorite in producing a large terrestrial crater transcends that of any volcanic or other earthly explosion. Conclusive evidence of such superintense shock waves are recognized by the unusual fracture pattern in rock fragments known as *shatter cones* and also by a form of silica known as *coesite,* that is produced under extremely high pressure.

An example of a region having shatter cones is the Steinheim Basin in southern Germany, formerly believed to be the site of a great volcanic explosion. Another of several examples mentioned by Dietz is the Vredefort Ring in the Transvaal of South Africa; the entire deformation is 200 Kilometers in diameter, comparable with the largest lunar craters. Coesite, the second shock-wave product, is reported present in the vicinities of the Giant Kettle in southern Germany, the Wabar craters of Arabia, the large Ashanti crater in Ghana, and elsewhere.

V. E. Barnes believes that the glassy stones known as *tektites* are solidified droplets of molten rock that were splashed into the air by the impact of large meteorites at these and other sites. But the distribution of tektites along long linear areas has led other astronomers to believe that these bodies were actually splashed out of the moon as molten rock resulting from meteoritic impacts. The composition of lunar material returned by Apollos 11 and 12 is quite different from that of tektites.

INTERPLANETARY MEDIUM

Gases, particles, radiation and dust occupy the spaces between the planets. Active sampling of this space by various spacecraft have greatly expanded our knowledge of the interplanetary medium over the past decade.

9.29
Evidence for the Interplanetary Medium

Long before the space age, evidence for the existence of the interplanetary medium accumulated and yielded some clues as to its nature. Screwlike motions in the tails of comets—for example, Comet Whipple-Fedtke-Tevzadze—give clear evidence of an *interplanetary magnetic field.* Similar conclusions were reached by studying the behavior of the earth's magnetic field. Scintillations in radio intensities led to the belief that a knotty stream of electrons and ionized particles are constantly flowing in the solar system and are associated with the "solar wind" (glossary).

All of the interplanetary space probes have carried detectors for measuring the composition of the interplanetary medium, its densities, its magnetic field, etc. *Plasma clouds* originating at the sun have been measured moving at 2,000 kilometers per second and the magnetic field strength has been measured at about 10^{-4} gauss.

9.30
Composition of the Interplanetary Medium

The primary constituents of the medium are electrons, protons, helium nuclei and dust. It appears that the particles and magnetic field have their origin in the sun while the dust component is debris left from the original nebula from which the solar system formed or is material being swept up as the sun moves through the Galaxy. The dust concentration increases between the earth and Mars and decreases between the earth and Venus as shown by the various Mariner missions. This observation can be explained by *radiation pressure* and may be related to the asteroid belt as well. Even so, the larger dust grains, called *micrometeroids,* are not nearly so numerous as predicted. The general dust component is the cause of the *Zodiacal Light* and probably the *Gegenschein* as well (9.31).

9.31
The Zodiacal Light and the Gegenschein

The faint triangular glow of the *zodiacal light* in the sky is best seen in middle northern latitudes in the west after nightfall in the spring and in the east before dawn in the autumn. In corresponding southern latitudes it is best seen after sunset in September and before sunrise in March. Broadest near the horizon, where it is then 30° or more from the sun, it tapers upward to a distance of 90° from the sun's place below the horizon. Because the glow is almost symmetrical with respect to the ecliptic, it

Figure 9.31
The Zodiacal Light and Comet Ikeya-Seki, October 31, 1965. The bright star at the top is Regulus in the constellation of Leo. Note how much brighter the Comet's tail is as compared to the Zodiacal Light. *(Photograph by H. Gordon Solberg, Jr., Las Cruces, New Mexico)*

reaches a higher altitude and is easier to observe at those seasons when the ecliptic is most inclined to the horizon.

In the tropics, where the ecliptic is more nearly perpendicular to the horizon, the zodiacal light is visible throughout the year in both the evening and morning. The glow is said to extend in a narrow band along the ecliptic entirely around the heavens. It also becomes visible at total solar eclipse immediately around the sun, where it blends with the sun's true corona (10.31).

The zodiacal light is attributed to sunlight scattered by a ring of meteoric dust particles that are revolving around the sun near the ecliptic plane. Whipple suggests that the supply of dust may well be continually replenished by disintegrating comets; else it would be depleted as the dust grains fall into the sun or are dispersed by pressure of the sun's radiations.

The *gegenschein,* or *counterglow,* is a very faint, roughly elliptical glow extending about 20° along the ecliptic and 10° in celestial latitude. It is centered 2° or 3° west of the point in the sky opposite the sun and is variable in form and position. This glow is visible to the unaided eye in the most favorable conditions, especially when it is viewed in the fall projected against the dull region of the constellations Pisces and Cetus. It is readily photographed with a very wide-angle camera or recorded with a photoelectric cell. Among the several explanations of the glow, the hypothesis that it is sunlight scattered by a dust-tail of the earth seems plausible at present but is not at all final, according to J. C. Brandt.

THE ORIGIN OF THE SYSTEM

Once the general characteristics of the solar system became known mankind has tried to explain the systems origin. Any theory must not only explain the regularities of the system, but must also provide for any major irregularities. Some 22 major theories have been advanced since the first one by Descartes in 1644.

9.32 The Nature of the Problem

Theories of the origin of the solar system are often known as hypotheses in deference to the style set by the distinguished pioneer P. S. de Laplace. These accounts are generally related to the problem of the birth of the sun itself. The theory must ultimately meet certain "boundary conditions" such as: the system must be older than 5 billion years (the age of the earth by radioactive dating) but younger than the Universe, it must produce a sun that is extremely stable in its energy output and, hopefully, it will produce a system in accord with the Titius-Bode Relation (7.5), although the latter is not a very strong boundry condition. By a current theory, which we examine later (16.15), the sun and the other stars evolved from contracting clouds of cosmic gas. How could a planetary system have formed around a youthful sun? The question is specifically about the solar system because this is the only known system of its kind. A successful answer to the question must conform to physical principles and must result in the system as it is known today.

The larger planets, as we have seen (7.4), have direct revolutions around the sun in nearly circular orbits, which are nearly in the same plane; they generally rotate in this direction as well. The larger satellites revolve around their planets and rotate on their axes in a similar manner. The less massive members of the system—the smaller satellites, the asteroids, and the comets and meteoroids—follow these regularities less faithfully, a fact that must be accounted for.

Theories of the origin of the solar system have been devised almost entirely during the past two centuries. The earlier of these theories were concerned with the mechanical features of the system, the later ones with characteristics of individual members as well, such as their compositions and densities. All the theories encounter problems of quantity and distribution of angular momentum.

9.33 Angular Momentum

The angular momentum of a revolving body, such as a planet revolving around the sun or a particle of a rotating globe, is the product of the mass, the square of the distance from the center of motion, and the rate of angular motion. The principle of the conservation of angular momentum asserts that the total angular momentum (the sum of all these products) of an isolated system is always the same.

Consider a rotating globe of gas. The total angular momentum is found by summing the products of mass, square of distance from the axis of rotation, and angular velocity for all molecules of the gas. Suppose now that the globe shrinks, the mass remaining the same. Because the distances of the molecules from the axis are diminishing, the angular velocities must increase to maintain the same angular momentum. Thus the shrinking of the globe increases the speed of its rotation.

$j = mr^2\omega$
$J = \Sigma j_i = K$
K, a constant

A successful theory of the evolution of the solar system must account not only for the present total angular momentum of the system but also for the distribution of the momentum. The sun's rotation contributes only 2 per cent to the total angular momentum of the solar system, whereas Jupiter's revolution accounts for 60 per cent and the four giant planets carry nearly 98 per cent of the total.

Most theories of planetary evolution have begun with a *solar nebula*, a rotating gaseous envelope surrounding the primitive sun in the course of its development. As examples we select the earlier hypothesis of Laplace and the recent accounts of C. F. von Weizsäcker and G. P. Kuiper.

9.34 Nebular Hypothesis

The nebular hypothesis of Laplace, published by him in 1796, is the most celebrated although not the first of the early attempts. Another hypothesis of the nebular origin of the solar system, proposed by Immanuel Kant in 1755, had not received much attention and was apparently unknown to Laplace.

Laplace's account begins with a gaseous envelope surrounding the primitive sun and having direct rotation around it. As the envelope contracted, it therefore rotated faster and bulged more and more at the equator. When the centrifugal effect of the rotation at the equator became equal to the gravitational attraction toward the center, an equatorial ring of gas was abandoned by the contracting mass. Successively smaller rings were subsequently left behind whenever the critical stage was repeated. Each ring gradually assembled into a gaseous globe having its circular orbit around the sun the same as the ring from which it was formed. Many of the globes developed satellites in a similar manner as they condensed into planets.

This process was intended to result in a system having the orderly motions we have noted for the principal members of the solar system, where the rings of Saturn might seem to have remained to demonstrate the correctness of the theory.

9.35 Value of the Nebular Hypothesis

The simplicity of the nebular hypothesis and the weight of its distinguished authorship combined to elevate it to a leading place among astronomical theories throughout the 19th century. It served as a powerful stimulus to scientific thought not only in astronomy but in allied

sciences as well; and it led the way to ideas of orderly development in other fields. As its deficiencies were gradually recognized, they gave warning of situations to be avoided in subsequent theories. These are chiefly as follows:

1. Some of the steps outlined in the hypothesis could scarcely have occurred because of the tendency of gas to disperse. The gaseous rings, for example, would not be expected to assemble into planets.
2. The hypothetical solar system would not have conformed to the actual one in the amount and distribution of angular momentum. Calculations show that in order to have abandoned the ring from which Neptune was formed the system must have possessed a total angular momentum 200 times as great as the present total. In addition, the hypothesis requires that the greater part of the present angular momentum should appear in the sun's rotation where only 2 per cent is actually found.
3. The organization of the solar system is more complex than was known in the time of Laplace. Many exceptions have since been discovered, as we have seen, to the regularities that his theory undertook to represent. Moreover, the compositions of the planets and their atmospheres were unknown at that early time.

9.36 Some Later Theories

The failure of Laplace's nebular hypothesis to reproduce the present solar system acceptably suggested eventually the trial of a different process. About the year 1900, the geologist T. C. Chamberlin and the astronomer F. R. Moulton proposed the planetessimal hypothesis. They imagined that another star passed close enough to the sun to cause the emergence from it of much gas by excessive prominence activity. The gaseous envelope thus formed around the sun condensed into small solid particles, called planetessimals, which assembled to produce members of the planetary system. This and an alternate tidal hypothesis were dismissed later as unprofitable. The interest of scientists subsequently reverted to the nebula surrounding the primitive sun as a more promising approach to the evolution of the system.

In 1945 the physicist C. F. von Weizsäcker proposed a revised version of the nebular hypothesis. The nebula was divided for planetary development—not by rotational instability, as in Laplace's theory, but by turbulent vortices caused by its rotation in accordance with Kepler's third law. The nebula was assigned an original mass a tenth of the sun's mass, or 100 times the present combined mass of the planets, and a composition mainly of light chemical elements like the composition of the sun. After their lighter gases had mostly escaped, the planets such as the earth were left with their present increased percentages of heavier elements.

9.37 The Protoplanet Hypothesis and the Accretion Hypothesis

The protoplanet hypothesis, published in 1949 by G. P. Kuiper, is a more recent development of the nebular theory. In it, the author avoids the difficulties of the Laplace hypothesis by allowing the solar nebula to contain enough material of solar composition for the present planets with their selective compositions to have formed. The excessive mass causes the rotating nebula to break up and protoplanets to form in the fragmented nebula. The envelopes of the planets were largely composed of hydrogen, helium, water vapor, ammonia, methane and neon, as will always be true for gases of solar composition kept at low (planetary) temperatures. This point later becomes essential to the development of star formation and agrees well with the observation results.

Kuiper's theory explains the spacing of the planets fairly well. It treats the major satellites' formation in the same way as that of the planets and therefore explains the ordered process of rotations and revolutions. The primary difficulty in this theory is that the various protoplanet nebulae overlap as do the various protosatellite nebulae. To overcome this problem and solve the distribution of angular momentum, W. H. McCrea has proposed an *Accretion theory.*

In McCrea's theory the nebula that forms the solar system is in the form of density concentrations that he calls *blobs.* The sun consisted originally of several blobs that coalesced and eventually became the center of the solar system. The more rapidly moving blobs tore away, in a sense, and are lost from the system, while the slower moving blobs slowly coalesced into planets and lesser bodies. The remaining lesser material enlarged the planets by accretion.

9.38 The Role of Magnetic Fields

Kuiper's theory is consistent with most of the facts, but he has also called attention in his papers to observations which his general theory does not incorporate. This theory is primarily explained in terms of hydro-dynamics, but a rigorous application leaves the sun with more angular momentum than it actually has.

Hoyle has proposed a more elaborate theory involving magnetic fields which follows earlier work by E. Fermi and Chandrasekar. This theory calls upon the *magnetic field* of the *protosun* to transfer its angular momentum to the planets. It is entirely possible that *magnetohydrodynamics,* as it is called, may add a refinement (in order) to the original von Weizacker theory that will in turn sharpen the Kuiper theory—in itself, an extension of the von Weizacker theory as we have pointed out.

The existence of magnetic fields in planets is not, per se, fundamental (except in so far as life on earth is concerned). Such fields depend more on the phase of substances within the planet, i.e. the temperature and pressure, and its rotation than on any fundamental characteristic of the nebula itself.

REVIEW QUESTIONS

1. What are the general characteristics of the periodic comets?
2. What are the two major families of comets?
3. Describe the spectrum of a comet.
4. Describe the three types of comet tails.
5. Why are there more meteors after midnight then before?
6. What is the difference between a meteor and a meteorite?
7. Discuss the spectrum of a meteor.
8. Explain the Zodiacal light.
9. What are the primary constituents of the solar wind?
10. Describe Kuiper's protoplanet hypothesis. Does Oort's theory for the origin of comets fit into this hypothesis?

REFERENCES

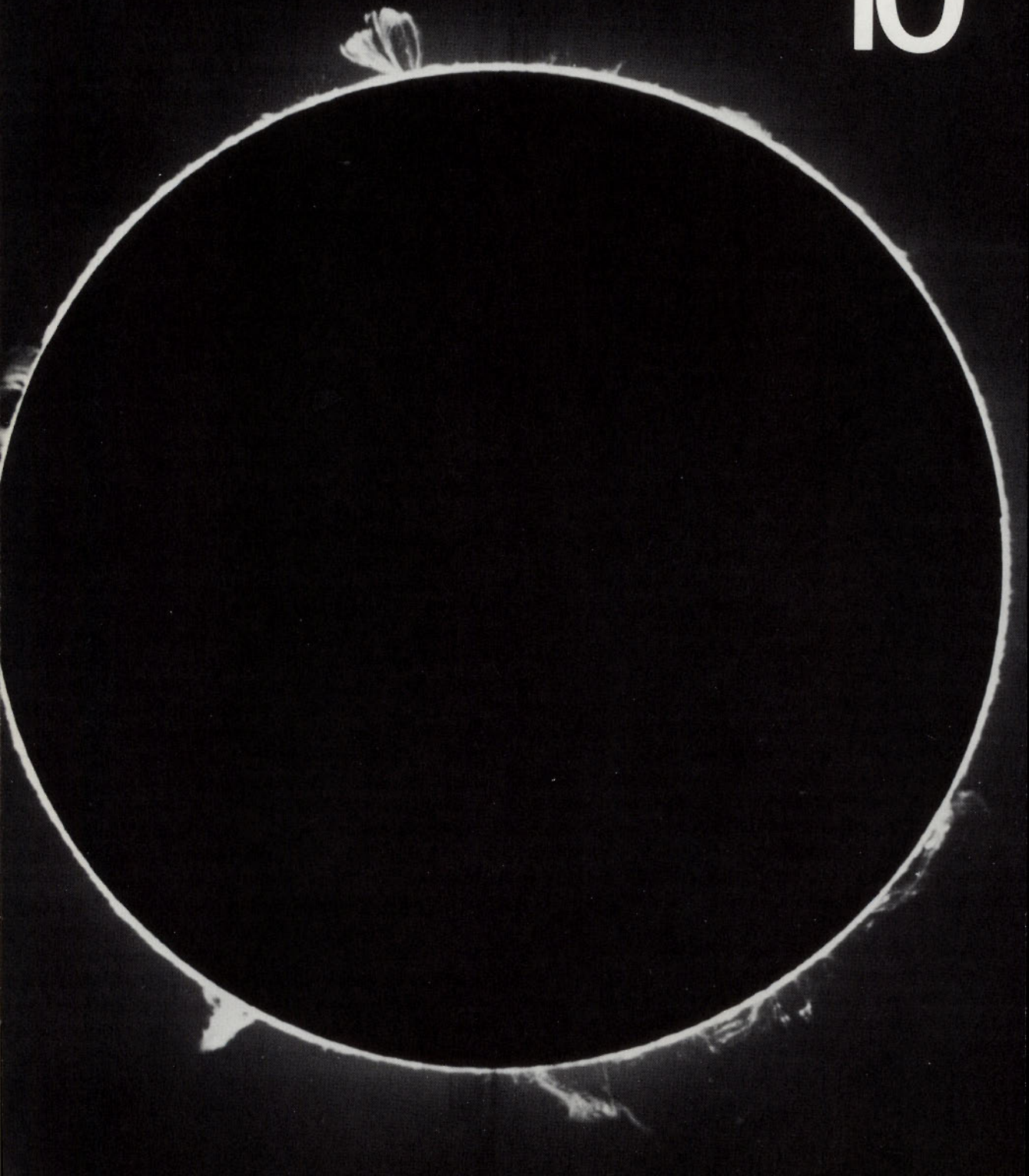
Heide, Fritz, *Meteorites,* Chicago, University of Chicago Press, 1964.
Lovell, A. C. B., *Meteor Astronomy,* Oxford: Clarendon Press, 1954.
Richter, N. B., *The Nature of Comets,* London: Methuen & Co., 1963.
Watson, F. G., *Between the Planets,* Cambridge, Mass.: Harvard University Press, 1956.
Wood, John A., *Meteorites and the Origin of Planets,* New York: McGraw-Hill Book Co., 1968.

FOR FURTHER STUDY

Jastrow, Robert and A. G. W. Cameron, eds., *Origin of the Solar System,* New York, Academic Press, 1963.
Liller, William, ed., *Space Astrophysics,* McGraw-Hill, New York, 1961.
McKinley, D. W. R., *Meteor Science and Engineering,* New York: McGraw-Hill Book Co., 1961.
Middlehurst, Barbara and Gerard P. Kuiper, eds., *The Moon, Meteorites and Comets,* Chicago: University of Chicago Press, 1963.

The 140-foot (42.6 m) steerable antenna of the National Radio Astronomy Observatory at Green Bank, W. Virginia. *(NRAO photograph)*

10



THE SUN

INNER STRUCTURE AND MOTION OF THE SUN — THE SUN'S RADIATION AND TEMPERATURE — THE SUN'S VISIBLE SURFACE; SUNSPOTS — THE CHROMOSPHERE AND CORONA — ASSOCIATED IONOSPHERIC DISTURBANCES

The sun is the dominant member and the power plant of the solar system. It is also the only star that is near enough to us to be observed in detail. Our account of the sun is accordingly associated with the descriptions of the planetary system which precede and with those of the stars in the following chapters.

Solar prominences. *(Photograph from the Hale Observatories)*

INNER STRUCTURE AND MOTION OF THE SUN

10.1 Structure of the Sun

The sun is a globe of very hot gas having a visible surface 1,390,000 km in diameter, or 109 times the earth's diameter. The sun is therefore 1⅓ million times as large in volume as the earth, and its mass is one third of a million times as great as the earth's mass. Its average density is one fourth the earth's mean density, or 1.4 times the density of water.

The *interior* of the sun, below the visible surface, is known to us only indirectly from theoretical researches. Its temperature increases from 6000°K at the deepest visible level to thirteen million degrees at the center. The outer layers of the sun merge gradually one into another. Conditions in the sun's interior are discussed in Chapter 12, and the problem of the sun's evolution is considered in Chapter 16.

The *photosphere* is the visible surface, the shallow layer from which the

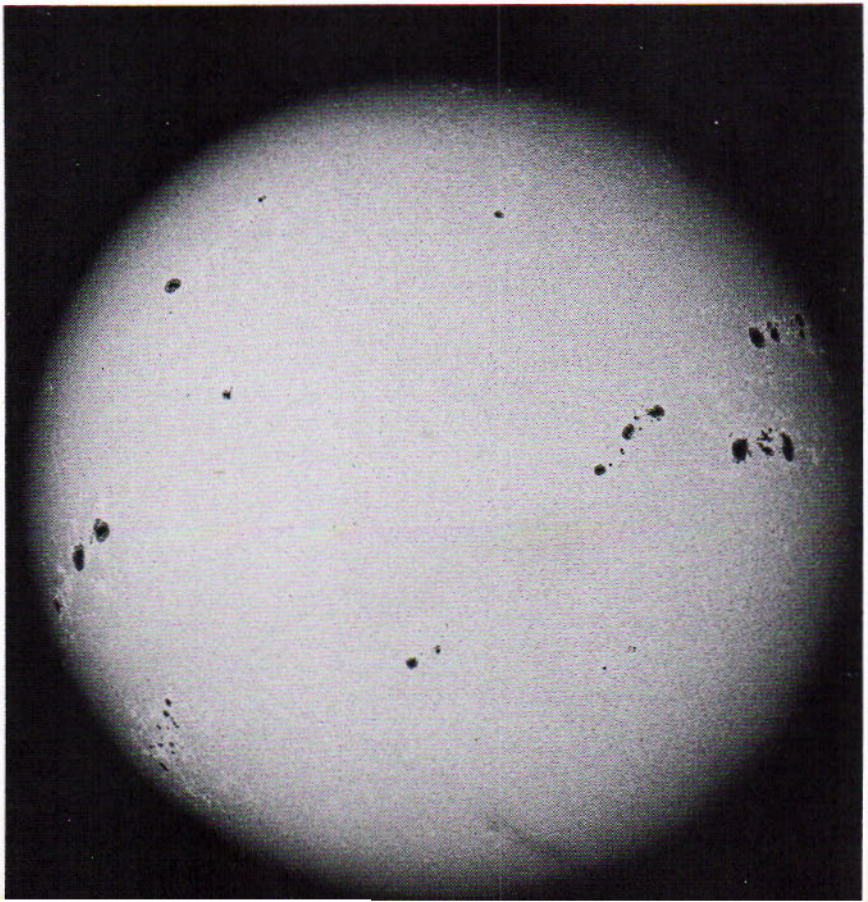


Figure 10.1
The sun on September 15, 1957. Note the large number of spots and also the limb-darkening effect. *(Photograph from the Hale Observatories)*

continuous background of the solar spectrum is emitted. It is mottled by brighter granulations and faculae and is often marked by darker sunspots. The gases above the photosphere constitute the sun's *atmosphere.*

The *chromosphere* is so named because of its red color, caused mainly by the glow of hydrogen. It extends several thousand miles above the photosphere. Its lowest stratum, sometimes called the reversing layer, is the source of most of the dark lines of the solar spectrum. The *prominences* appear above the chromosphere, at times attaining heights of many hundred thousand kilometers. They are visible during total solar eclipses and with special apparatus on other occasions.

The *corona* appears around the sun during total eclipses as a filmy halo of intricate structure. Its inner parts can be observed with the coronagraph at other times.

10.2 The Sun's Rotation

The gradual movement of sunspots across the disk of the sun is a well-known effect of the sun's rotation. A spot now near the center of the disk will disappear at the edge in about a week. Two weeks later the spot will reappear at the opposite edge if it lasts that long, and after a week more it will again be seen near the center of the disk. The interval after which a long-lived spot comes around again to the same place on the disk of the sun is about 2 days longer than the actual period of the sun's rotation in that latitude, because of the earth's revolution in the meantime.

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

The sun's rotation is in the same sense as the earth's rotation and revolution, from west to east. Thus the spots on the hemisphere which faces us are carried across the disk from east to west with respect to directions in the sky. Because the sun's equator is inclined 7° to the ecliptic plane, the paths of spots are slightly curved. The curvature is greatest in March, when the southern end of the sun's axis is most inclined toward us, and again in September, when its northern end is most inclined toward us. In June and December the spots move across the disk in lines more nearly straight.

Unlike the earth, where the surface is constrained to rotate all in the same period, the period of the sun's rotation is longer as the distance from its equator is greater. The period at the equator is about 25 days. It has increased to 27 days in latitude 35°, beyond which the spots rarely appear. The further lengthening of the period toward the poles is established by spectroscopic observations.

The speed of the sun's rotation in a particular latitude can be determined by photographing on the same plate the spectra of the west (receding) and the east (approaching) edges, and comparing the two (Fig. 10.2B). The solar lines in the first spectrum are displaced toward the red by the Doppler effect; in the second spectrum they are displaced

toward the violet. Half the difference in the positions of the lines represents the apparent speed of the rotation in that latitude, and from it the actual period can be derived. The period of the sun's rotation as determined from the spectra increases progressively from less than 25 days at the equator to as much as 33 days in latitude 75°.

We have no direct information about how the interior of the sun rotates. R. H. Dicke has recently measured the *oblateness* of the sun and found it larger than the usual rotating models would predict it to be. He explains this by assuming that the relatively small radiative core of the sun is rotating very rapidly, perhaps as fast as a complete revo-

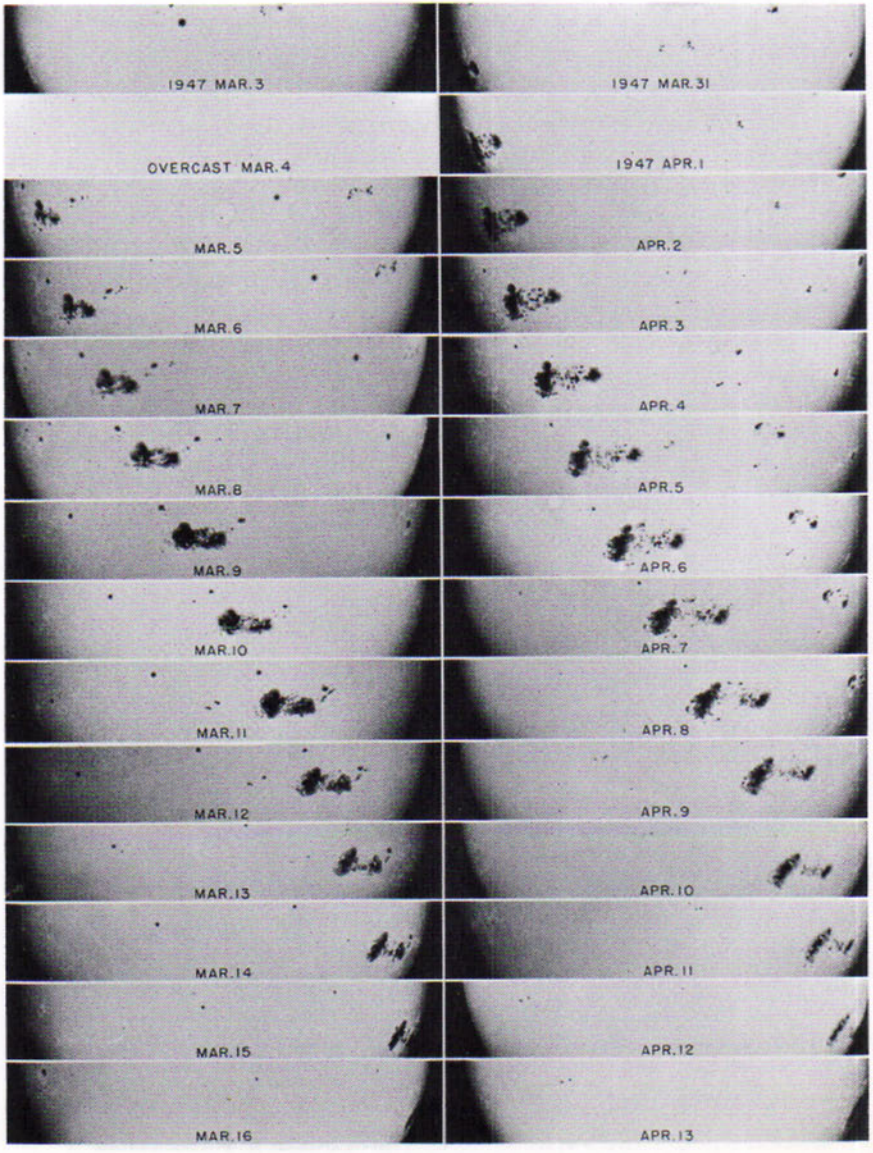
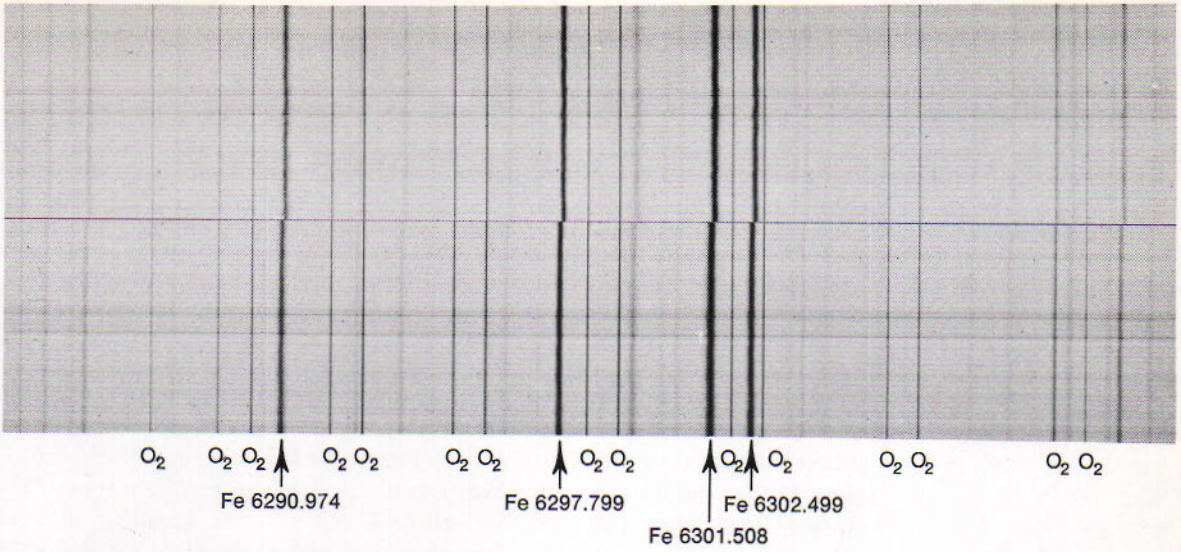


Figure 10.2A
Sunspots show the sun's rotation. This large group of 1947 lasted more than 3 months. *(Photograph from the Hale Observatories)*



lution every six or eight hours. This interpretation has been objected to, but it would help to explain a few problems at the price of raising a few more. For example, such a solar interior would explain most of the excess in the advance of perihelion of Mercury. The advance is presently explained by relativity theory. A rapidly rotating core does help explain why the sun's photosphere is still rotating.

The sun's magnetic field and the solar wind act as a brake on the rotating surface of the sun by transferring angular momentum away. The half-life of rotation is approximately 4 billion years so we require some internal source of rotation to keep the surface rotating at an essentially constant rate.

Figure 10.2B
Effect of the sun's rotation on its spectrum. The Doppler effect causes the spectrum of the west limb (top) to be displaced to the red and the east (bottom) limb to be displaced to the blue. The very sharp undisplaced lines are due to oxygen molecules in the earth's atmosphere. *(Courtesy of A. K. Pierce, Kitt Peak National Observatory).*

THE SUN'S RADIATION AND TEMPERATURE

The sun radiates energy to space at a prodigous, but controlled rate. The energy output, called the solar constant, has been essentially constant for more than 10 billion years based upon biological evidence.

10.3 Radiation Received from the Sun; the Solar Constant

The rate at which we receive the sun's radiation is measured by the rate of its heating of a substance that completely absorbs the radiation. The *pyrheliometer* is the instrument that has been employed for this purpose at stations of the Smithsonian Astrophysical Observatory. It contains a

thermometer for recording the rate at which the radiation raises the temperature of a blackened silver disk. Allowance is made for the radiation scattered in the earth's atmosphere; this requires the addition of at least 30 per cent to the observed rate of heating, depending on the sun's distance from the zenith.

The *solar constant* is the rate at which the solar radiation is received by a surface exposed at right angles to the sun's direction just outside our atmosphere, when the earth is at its mean distance from the sun. Its value, as determined by R. Tousey, taking into account the results of rocket records, is 2.0 cal/min cm^2 (calories per minute per square centimeter); this differs from the Smithsonian value, 1.95 cal/min cm^2 by scarcely more than the probable error. A *calorie* is the quantity of heat required to increase by 1°C the temperature of 1 gram of water at 15°C.

The equivalent value of the solar constant is 1.4×10^6 ergs/sec cm^2. Multiplying by the number of square centimeters in the surface of a sphere having as its radius the earth's mean distance from the sun, we have the rate at which the energy is intercepted by the sphere. This is practically the rate of the sun's total radiation; it equals 4×10^{33} ergs/sec. The rate of the sun's radiation is accordingly 6.5×10^{10} ergs/sec cm^2.

$S = 4\pi d^2$

The *erg* is the unit of work, or energy; it is about the impact of a slow-flying mosquito. The *horsepower* is a common unit of the rate of doing work; it equals about 7.5×10^9 ergs/sec. Thus, the sun produces about 8.8 hp/cm^2, or a total of 5×10^{23} hp.

10.4 Laws of Radiation

The relations between the temperature of a body and the quantity or quality, or both, of the radiation it emits are expressed by the *radiation laws*. These laws apply to a *perfect radiator*, called a "black body" because it has the greatest possible efficiency as a radiator at any particular temperature and is also a perfect absorber of radiation. They also define temperature in the sense that we are here using the term. Despite their ideal character the laws serve reasonably well for the sun, stars, and planets.

Planck's Law. M. Planck in a series of papers arrived at the general relation for a black body which relates the spectral radiance at a given wavelength (ρ_λ) to the absolute temperature (T).

$$\rho_\lambda = \frac{2\pi h^2 c}{\lambda^5} \left(e^{hc/\lambda kT} - 1\right)^{-1}$$

Stefan's Law. (often referred to as the Stefan-Boltzmann law) was given much earlier than Planck's and may be derived from Planck's Law by integrating over all wavelengths. It relates the total energy (E) in ergs emitted by a square centimeter per second to the fourth power absolute temperature. The value of the constant in the formula is obtained from laboratory experiments. It is interesting to note that if the temperature of a body is doubled its total radiation increases by a factor of 16.

$$E = \sigma T^4$$
$$\sigma = 5.669 \times 10^{-5}$$

Wien's Law also preceded Planck's law and relates the maximum wavelength (λ_{max}) to the inverse of the absolute temperature of the black body. Wien's Law can be obtained from Planck's law by differentiating it to wavelength and setting the derivative equal to zero (one must assume that T is reasonably large). In the common form of Wien's law the wavelength is obtained in centimeters and must be multiplied by 10^8 to give the results in Angstrom Units. If, for example, the temperature is 4000°K, the most intense part of the spectrum has the wavelength of 7242×10^{-8} cm, or 7242 angstroms (10.20), in the red. If the temperature is raised to 8000°, the greatest intensity is shifted to 3621 angstroms, in the ultraviolet. Thus when a piece of metal is heated to incandescence, it first has a dull red glow, which brightens and changes to bluish white as the metal is further heated.

$$\lambda_{max} T = 0.2897$$

Planck's law is the most general of the radiation laws. By means of this rather complex formula, which may be found in treatises on physics, it is possible to calculate for a perfect radiator at any assigned temperature the rate of radiation in the different wavelengths.

10.5 Spectral Energy Curve

This curve, calculated from the general law for particular temperatures (Fig. 10.5), shows how the intensity of the radiation varies along the spectrum. At a higher temperature the curve is higher at all points; and the increase is greater for shorter wavelengths, so that the peak of the curve is shifted toward the violet end of the spectrum.

Stefan's law and Wien's law refer to special features of the spectral energy curve. The former relates to the area under the curve, which represents the total amount of energy radiated from a square centimeter at

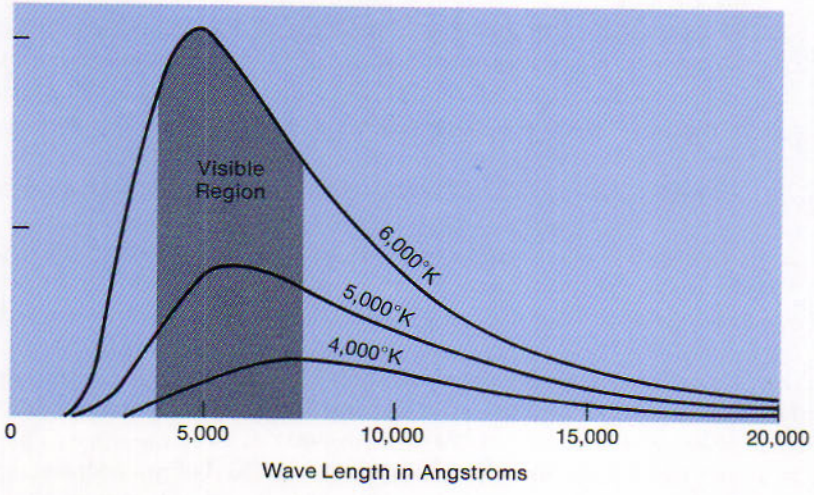


Figure 10.5
Spectral energy curves for a perfect radiator. The heights are proportional to the intensity of the radiation. As the temperature is increased, the total radiation represented by the area under the curve increases (Stefan's law), and the peak of the curve is shifted to shorter wavelengths (Wien's law). The curve for each temperature is calculated by Planck's formula. The shaded area indicates the radiation to which the eye is sensitive.

a particular temperature; the latter gives the wavelength of the most intense radiation at that temperature.

On combining the data of observation with these radiation laws, the sun's *effective temperature* (T_e) can be determined, that is, the temperature that the sun's surface must have, if it is a perfect radiator, in order to radiate as it does.

10.6 The Sun's Temperature

The effective temperature of the sun's visible surface is about 5750°K. This value is obtained very nearly by substituting the rate of the sun's radiation (6.5×10^{10} ergs/sec cm^2 from 10.3 in the formula of Stefan's law. Since this rate is determined from the solar constant, which is derived from the radiation from all parts of the disk, this effective temperature is the average for the whole disk; it is the value to be used when the temperatures of the sun and stars are to be compared. It will be noted later that the sun's disk is less luminous near the edge than at the center.

The effective temperature at the center of the sun's disk is about 6000°, as determined by three methods:

1. *By Stefan's law.* The average radiation from all parts of the disk, as calculated from the solar constant, is increased by 16 per cent to represent the radiation from the center of the disk. The corresponding temperature is about 5980°.
2. *By Wien's law.* The wavelength of the most intense radiation is around 4750 angstroms, in the blue-green of the spectrum. The resulting temperature is something like 6100°. This determination has the least reliability of the three, because of uncertainty in assigning the position of maximum intensity in the spectrum.
3. *By Planck's law.* The observed intensities of the solar spectrum in the various wavelengths are compared with the energy curves calculated by the formula for different temperatures. A fair general agreement is found between the observed curve and the one calculated for the temperature of 6000°.

   Exact agreement in the values obtained from the three laws would scarcely be expected. Differences between the observed and calculated curves are caused by the thousands of dark lines throughout the solar spectrum, and by general absorption in the sun's atmosphere which is not the same for the different regions of the spectrum.

10.7 Temperatures in Different Parts of the Sun

The sun's effective temperature of 5750°K is the average for all parts of the disk, as we have noted. The apparent surface temperature decreases from 6000° at the center of the disk to 5000°K near the edge. The temperatures of sunspot umbras are around 4600°K. Below the photosphere the temperatures rise so rapidly with increasing depth that most of the

sun's interior is above 1,000,000°K. The central temperature is now believed to be of the order of 13,000,000°, a considerable reduction from values formerly assigned it. Throughout the sun the heat is sufficient to keep all the material in the gaseous state.

The temperature diminishes with increasing height above the photosphere; it is 4500° in the lower chromosphere, and it would be 3000° in the corona as measured by the heating of a perfect absorber at these distances from the photosphere. The *temperature* in this sense is as defined by the radiation laws and we refer to this as the color temperature. As derived from the random motions of the particles and their degrees of ionization, however, the *kinetic temperature* of the upper chromosphere is 100,000°, and that of the corona is 1,000,000°.

THE SUN'S VISIBLE SURFACE: SUNSPOTS

What appears to be a smooth uniform surface is, upon careful scrutiny a highly structured, mottled sphere with complex motions. The surface occasionally shows dark blotches called sunspots that are of great interest because they occur in a cyclic period measured in years. Highly specialized instruments and telescopes have been built to study sunspots and the sun's surface.

10.8 Observing the Sun

It is unsafe to look at the sun with more than a quick glance on a clear day with unprotected eyes; to look at it through a telescope without special precaution invites immediate and serious injury to the eye, because the objective acts as a burning glass. A convenient way to observe the sun with a telescope is to hold a sheet of smooth white cardboard back of the eyepiece, racking the eyepiece out beyond the usual position of focus until the sun's image is sharply defined. This also has the advantage that many can observe at once. Finer details may be viewed through the telescope with a special solar eyepiece which admits to the eye only enough light to form a clear image of the sun.

Photographs provide permanent records of the changing features of the sun's surface and its surroundings. They are taken frequently at a number of observatories, often with special solar telescopes. Photographs

of the sun and its spectrum from high-altitude stations, balloons, and rockets are giving exciting new information. A Princeton University project, directed by M. Schwarzschild, to photograph the sun from an unmanned balloon made a successful beginning in 1957. The photographs have been taken with a 12-inch (30 cm) telescope at the height of about 24.5 km (Fig. 10.12). Other "photographs" of the sun in the UV and X-ray regions are being obtained from orbiting spacecraft.

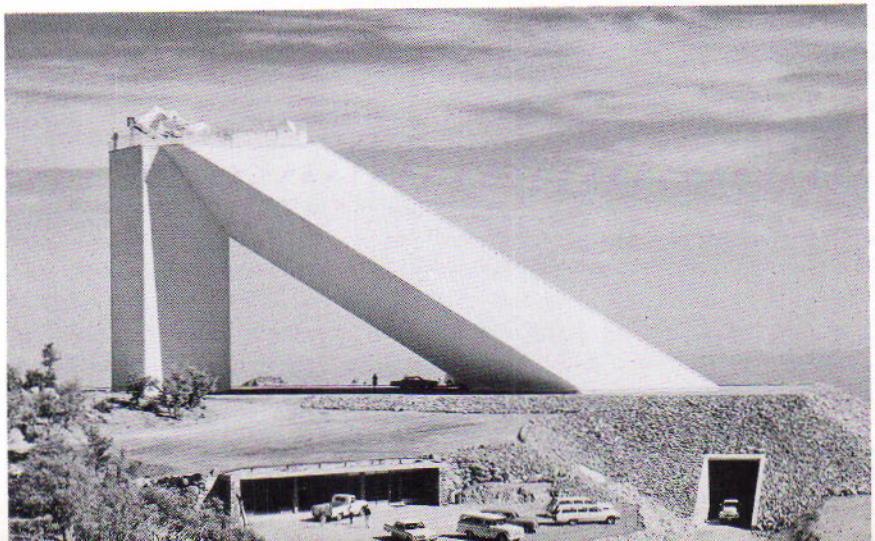
10.9 Solar Telescopes

Fixed telescopes are especially valuable in optical investigations of the sun. They permit the use of long-focus objectives and high-dispersion gratings with a minimum of mechanical construction. Among the larger examples of fixed telescopes is the 30 meter tower telescope of the Mount Wilson Observatory, which has a 20 meter tower and a horizontal solar telescope as well. There are also the 17 meter and 23 meter tower telescopes of the McMath-Hulbert Observatory.

A tower telescope has a coelostat at its summit. This consists of an equatorially mounted plane mirror, which is rotated around the polar axis at the rate of once in 48 hours, and a second fixed plane mirror beside it, which directs the sunlight to a fixed objective, either a lens or a mirror. The image of the sun formed at the base of the tower may be photographed directly, or else the light from a selected part of the disk may be passed on through a narrow slit to a grating in the well below the observing room, from which it is returned dispersed into spectra.

The new solar telescope of the Kitt Peak National Observatory has a sloping tube 152 meters long that is parallel to the earth's axis and three fifths below the ground. A *heliostat*, a rotating plane mirror, 203 cm in

Figure 10.9
The 152-meter R. R. McMath solar telescope at Kitt Peak. *(Kitt Peak National Observatory photograph)*



diameter, reflects the sunlight down the tube to the 152 cm objective, a paraboloidal mirror near the bottom. This mirror, having a focal length of 91 meters, forms an image of the sun's disk, averaging 85 cm in diameter through the year, in the observation room at the ground level. Here the spectrum of any part of the disk may also be observed.

10.10 The Photosphere

The photosphere is the visible surface of the sun. It is a layer 400 km thick, the direct source of practically all the sun's radiation. Here the sun becomes opaque, so that we cannot see into it any farther. Yet the gas is still very tenuous in this layer, producing a pressure not exceeding a hundredth of our atmospheric pressure at sea level. Its opacity is ascribed mainly to the abundance of negative hydrogen ions, hydrogen atoms which have acquired second electrons.

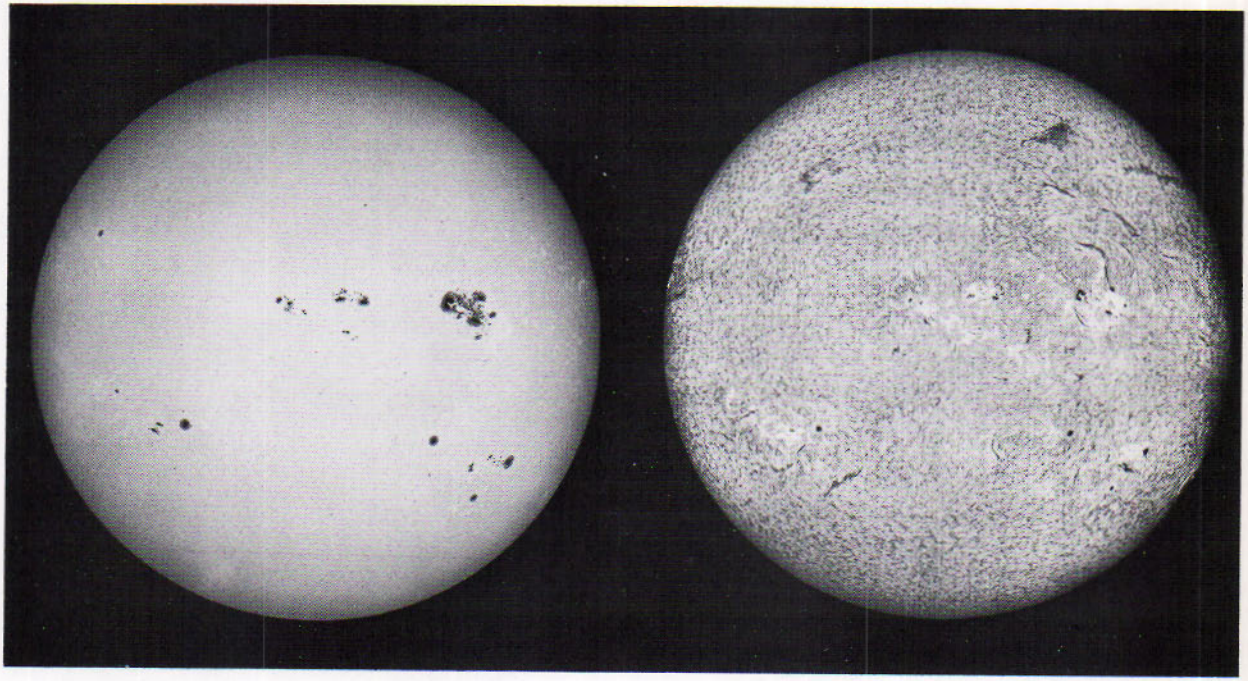


Figure 10.10
The sun in visible light (left) and in Hα light (right). Notice the limb darkening effect in both photographs. *(Photographs from the Hale Observatories)*

The sun's disk is less luminous and redder near its edge than near its center. The fading of the light toward the edge, or *limb darkening*, is especially conspicuous in photographs with blue-sensitive plates, an effect often partly compensated in the prints for the sake of clearer illustration. The explanation of the effect is that, because of the opacity of the photosphere, the light from the edge emerges from a higher and cooler level,

so that it is less bright and redder than the light from the lower levels we can see near the center of the disk. It is the custom to denote the heights of features of the sun from that of the visible edge.

10.11 Granulations of the Photosphere

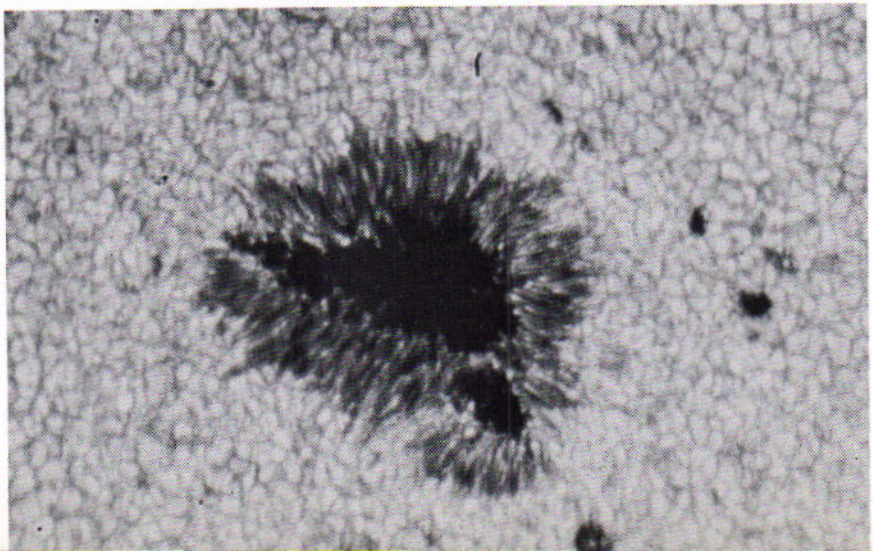
As observed with the telescope the photosphere presents a mottled appearance, which in good conditions is resolved into bright *granules.* These are scattered profusely on the less luminous and by contrast grayish background. The granules vary from 240 to 1400 km in diameter and are separated by narrow dark spaces. They are hot spots, hotter by 100°C or more than the normal surface, and each one lasts only a few minutes. What we see is a seething surface where hotter gases come up from below and quickly cool.

The granules are not distinct near the edge of the disk. Here we see instead the larger bright areas of the *faculae* (little torches) at somewhat higher levels. The faculae are inseparable companions of sunspots, and they often antedate and survive the associated spots.

10.12 Sunspots

The sunspots are dark spots on the sun's disk, ranging in size from spots scarcely distinguishable from the intergranular spaces to great spots that are visible to the unaided eye. They appear dark by contrast with their surroundings, although they are brighter than most artificial sources of light. A sunspot generally consists of two parts; the *umbra,* the central darker region, and the *penumbra,* the lighter surrounding region which is three fourths as bright as the photosphere. The penumbra, as it appears in the Princeton photographs, is described by R. E. Danielson as a complex array of predominantly radial bright filaments. The filaments, having

Figure 10.12
Solar granulation around a sunspot. This remarkably detailed photograph was taken by Stratoscope I from above 80,000 feet (24,000 m). *(Princeton University Observatory, NASA, ONR and NSF in Project Stratoscope)*



lifetimes roughly five times those of the granules, are interpreted by him as convection rolls in the magnetic field of the sunspot.

Sunspots occur in groups. Where a solitary spot is seen, it is likely to be the survivor of a group. The unusually large and long-lived group shown in Fig. 10.13 began coming into view at the sun's eastern edge on March 4, 1947. Its largest spot measured 145,000 by 95,000 km. The group attained a length of 320,000 km and a maximum area of 14,760 million square km which, however, was less than half of 1 per cent of the visible hemisphere of the sun. The group was brought around to the eastern edge three times by the sun's rotation and was last seen disappearing at the western edge on April 13. A somewhat larger group, the largest so far recorded, was observed from January 29, 1946 to May 8, having survived more than 99 days.

10.13 The Life of a Sunspot Group

Rapid development and slower decline characterize the life history of the normal group. In its very early stages the group consists of two clusters of small spots in about the same latitudes and 3° or 4° apart in longitude. As the group develops, a large compact *preceding spot* dominates the part that is going ahead in the direction of the sun's rotation, while a somewhat less large and compact *following spot* is conspicuous in the part that is moving behind. The following spot is usually a little farther from the equator than the preceding spot. Meanwhile, these *principal spots* draw apart to a difference of 10° or more in longitude. By the end of a week the group generally attains its greatest area and the decline sets in.

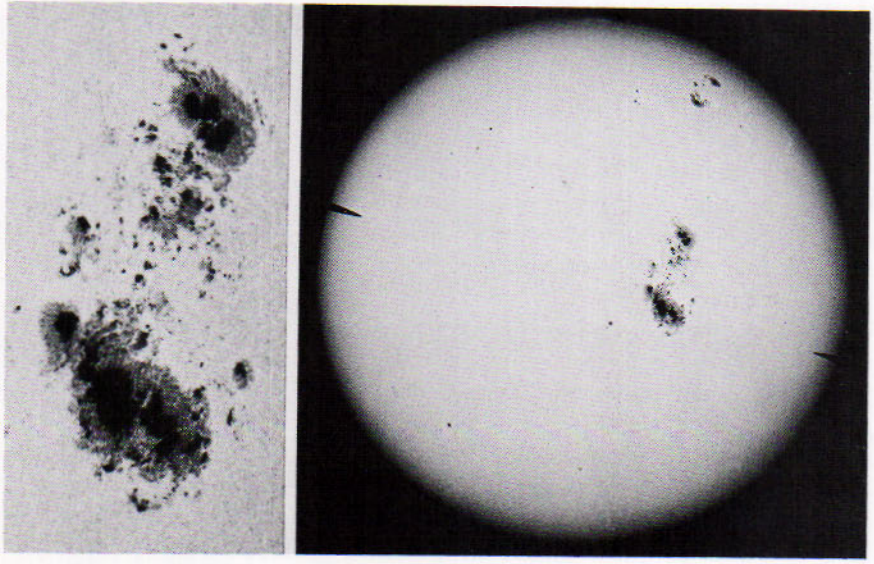


Figure 10.13
The great sunspot group of April 7, 1947, in overall view of the sun and in closeup. Notice the limb darkening in the sun's disk. *(Photographs from the Hale Observatories)*

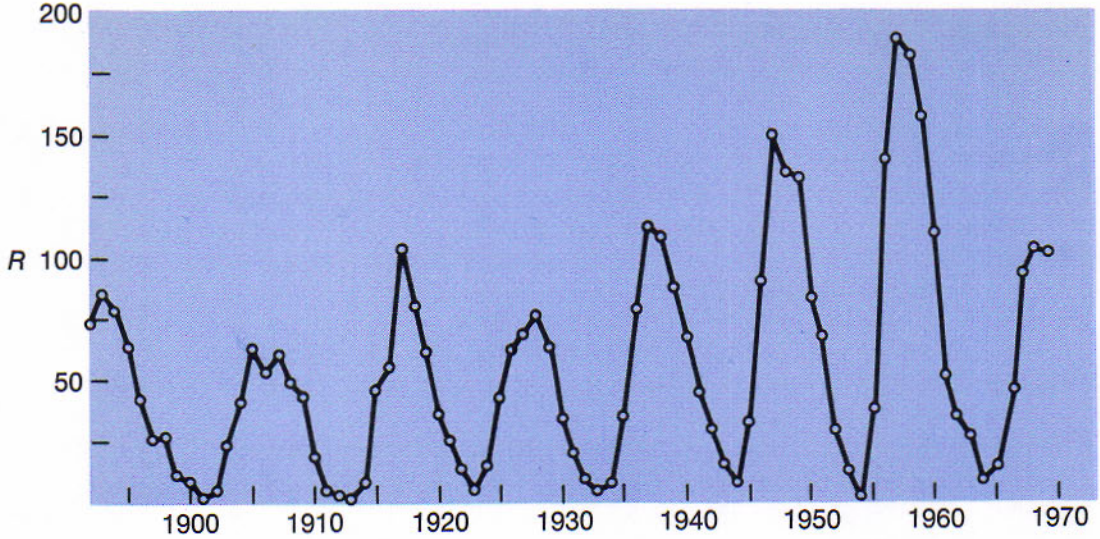


Figure 10.14
Average daily number of sunspots (R), plotted by the year.

The following spot breaks up into smaller spots, which diminish in size and vanish along with other small spots. Finally the preceding spot remains alone, a single umbra at the center of a nearly circular penumbra, which may persist for several weeks or even months. This spot usually disappears after becoming progressively smaller, but not by breaking up as the following spot generally does. Sometimes, as in the case of the great group of early 1946, the following spot is the larger and surviving member.

10.14 The Sunspot Cycle

In some years the sun's disk is seldom free from spots, whereas in other years it may remain unspotted for several days in succession. Sunspots vary in number in a roughly periodic manner. The variation was first announced, in 1843, by H. Schwabe, an amateur astronomer in Germany, after he had observed the sun systematically with a small telescope for nearly 20 years.

The overall average interval between the times of the greatest numbers of spot groups has been 11.1 years, but for the past half-century it has been more nearly 10 years. The rise to maximum spottedness is generally faster than the decline. Not only is the period of the sunspot cycle variable, but the number of groups at the maximum also varies from cycle to cycle and is likely to be greater in the short-period cycles. It is accordingly possible to predict only approximately the date of a future maximum and how plentiful the groups will then become.

A maximum of 1060 groups was reported for 1957, the highest since the year 1778. The following minimum occurred in 1964 when only 116 groups were reported. The sun was spotless on 111 days and not a single day throughout that year had as many as five spots. In 1954 the sun was spotless on 213 days. In contrast there was no day with less than 100 spots during the year of 1957. The present maximum was reached in 1969. In Figure 10.14 we have plotted the relative sunspot index (R) from 1892 through 1968. We note that the maximum in 1917 was only one half that of 1958 and that the current maximum is about equal to that of 1917. A more extensive figure would show that there is a very long term cycle with the highest maxima occurring about every 100 years. We have reached the peak of this long cycle and expect the next maximum in 1980 to have fewer spots than the current one.

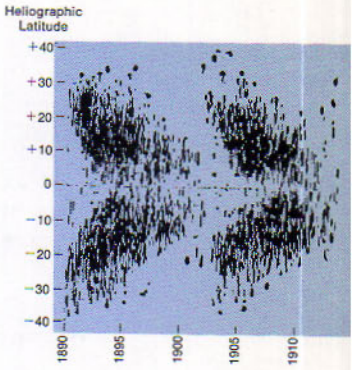


Figure 10.16
Shifting of sunspot zones. The disturbed areas gradually draw in toward the sun's equator. *(From a diagram by E. W. Maunder)*

10.15 Shifting Sunspot Zones

Sunspots are confined mainly between heliographic latitudes 5° and 30°. Few are seen at the equator or beyond latitudes 45°. At a particular time they occur in two rather narrow zones equidistant from the equator. The zones shift toward the equator in cycles paralleling the variation in the spot numbers.

About a year before sunspot minimum the new cycle is announced by the appearance of spots in the higher latitudes. As these spots vanish and others appear, the disturbance gradually closes in toward the equator. When the minimum of the numbers is reached, the fading cycle is marked by spots near the equator. Meanwhile, the early members of the next cycle have appeared in the higher latitudes. Thus in the year near the minimum of 1954 the Mount Wilson observers recorded 15 groups of the old cycle and 31 groups of the new one.

10.16 The Sunspot Spectrum

This spectrum gives decisive evidence as to two characteristics of the spots:

1. *Lower temperature of sunspots.* As compared with the normal solar spectrum (10.20): (a) The continuous background of the spot spectrum is weakened progressively from red to violet. (b) Certain dark lines are strengthened; they are lines that are more conspicuous in laboratory spectra of sources at lower temperatures. (c) Other dark lines are weakened; these are lines of ionized atoms which are weaker at lower temperatures when other conditions are unaltered. (d) Dark bands produced by chemical compounds appear in the spot spectrum, compounds which cannot often form at the higher temperatures above the undisturbed surface of the sun. The temperature of the ordinary spot umbra is 1200°C below that of the photosphere, according to Pettit.

2. *Magnetic fields of sunspots.* Many lines in the sunspot spectrum are widened, and some are plainly split. This effect had been known for some time before G. E. Hale, in 1908, demonstrated its association with the magnetism of sunspots.

10.17 Magnetism of Sunspots

The Zeeman effect ($\delta\lambda$), known by the name of the physicist who discovered it in the laboratory in 1896, is the splitting of the lines in the spectrum when the source of the light is in a magnetic field. We are concerned here with the effect when the light emerges along the line joining the poles of the magnet. The lines in the spectrum are then divided into pairs, the components of which are circularly polarized in opposite directions.

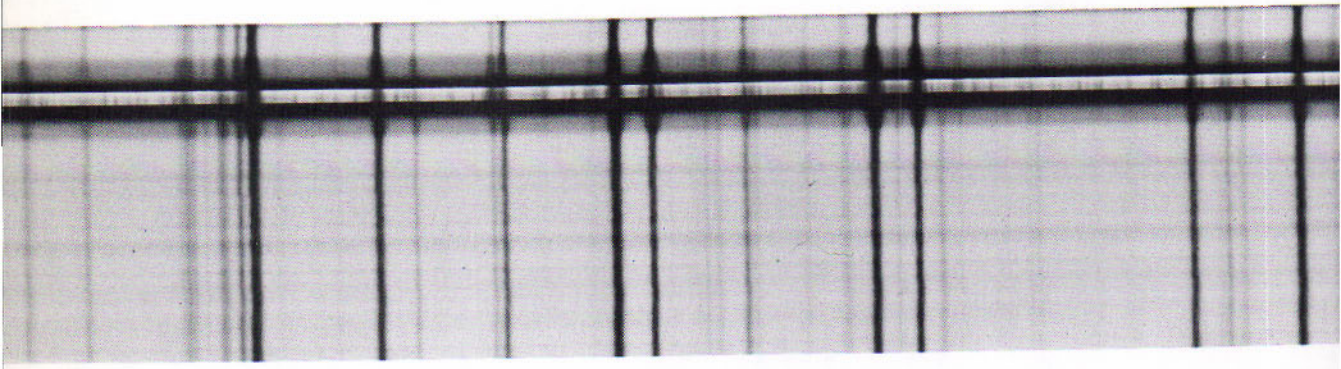
$$\delta\lambda = \frac{e\lambda^2 B}{4\pi mc^2}$$

e charge on electron in esu
m mass of electron in grams

In the spectrum of a sunspot near the center of the sun many lines are so divided, showing that sunspots have magnetic fields. The field strength, indicated by the amount the lines are separated, ranges from 100 gauss or less for small spots to as much as 3700 gauss in the case of the great sunspot of 1946. The *gauss* is the unit of magnetic field intensity.

Strips of mica or other quarter-wave device along with a nicol prism in front of the slit of the spectroscope alternately suppress the components of the lines to facilitate the measurements and to show the *polarity* of the spot, whether its positive or negative pole is toward the observer. Daily polarity records of sunspots are kept at Mount Wilson and elsewhere.

Figure 10.17
Zeeman splitting in the sunspot spectrum at 5250 Å. *(Photograph from the Hale Observatories)*



10.18
Polarities of Sunspots

Most sunspot groups are bipolar. Where the leading members of a group have positive polarity, the following members have negative polarity. If the group we are considering is in the sun's northern hemisphere, the statement applies to all other bipolar groups in this hemisphere at the time. For all such groups in the southern hemisphere the situation is reversed; the leading members have negative and the following members positive polarity.

A remarkable feature of sunspot magnetism is the reversal of polarity with the beginning of each new cycle. When the groups of the new cycle appear in the higher latitudes, the parts which had positive polarity by the rule of the old cycle now have negative polarity, and vice versa. First recorded at Mount Wilson around the sunspot minimum of 1913, this reversal of polarities has been observed at each succeeding minimum.

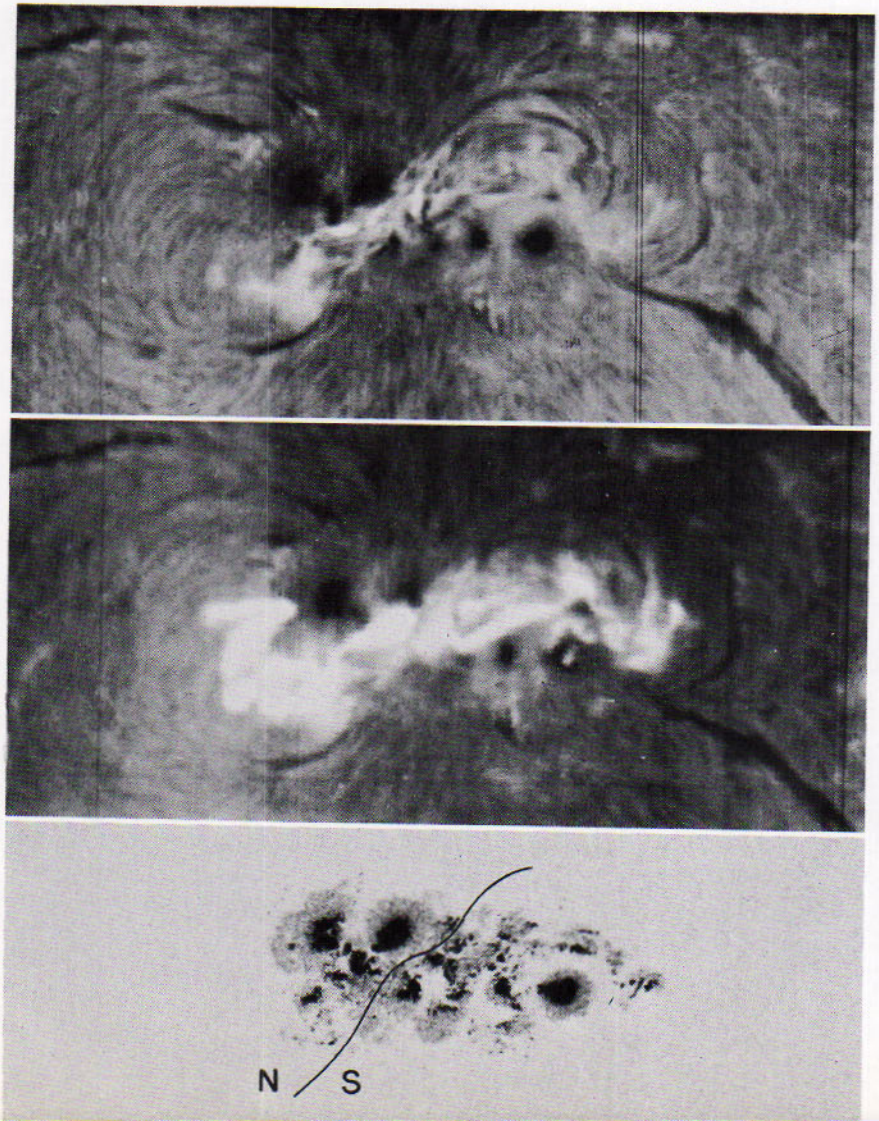


Figure 10.18
Sunspot and flare of July 25, 1946. Bottom print shows spot group and polarity in visible light. *(Photographs from the Hale Observatories)*

10.19 Magnetic Fields of the Sun

A scanning process is employed by H. W. and H. D. Babcock at Mount Wilson Observatory for mapping the magnetic fields over the sun's disk. The image of the sun formed by a solar telescope is allowed to drift repeatedly over the slit of a spectrograph, giving a succession of traces parallel to the sun's equator. The Zeeman effect at each point of the disk is recorded at a corresponding point in the tracing by a vertical deflection proportional to the intensity of the field—to the north for positive polarity and to the south for negative polarity. The magnetogram so produced is illustrated in Fig. 10.19. By a supplementary process developed by R. B. Leighton, positive fields can also be shown bright and negative fields dark.

Localized magnetic fields generally within 40° or 50° from the sun's equator are likely to be most intense around sunspot groups or where only conspicuous plages appear, or sometimes even where no visible evidence of special disturbance is observed. Beginning with the 1953 magnetograms, weak fields at higher latitudes were also recorded; they had positive polarity near the sun's north pole and negative polarity near the south pole. The polar fields reversed sign around the time of sunspot number maximum, in 1958.

Thus the sun's general magnetic field could be regarded as having two components: (1) *Polar fields* have opposite polarity around the two poles, and these have reversed sign at sunspot maximum. (2) *Ring-shaped fields* parallel to the equator have opposite field direction in the two hemispheres; they migrate toward the equator in the sunspot cycle and are replaced at sunspot minimum by new fields in the higher mid-latitudes, where the directions are reversed. These fields provide the basis for H. W. Babcock's magnetohydrodynamic interpretation of observed features at and above the solar surface, a theory in which he combines the results of his own studies with conclusions of other investigators.

The ring-shaped magnetic fields are mainly submerged below the surface, as Babcock pictures them. Because of the increasing period of the sun's rotation with distance from its equator, the lines of force are twisted into "magnetic ropes" where, especially in their loops formed by excessive twisting, they can exert pressure comparable with the gas pressure. These equatorial fields undulate like sea serpents. Many loops are brought up to the surface by the convection currents. As they break through the surface, the loops produce pairs of oppositely polarized regions where sunspot groups may form. Anchored to these regions the loops arch above them to support the material of the solar prominences. The author of the theory points out that its adequate development will require a body of data which can be obtained only through long-continued observation of the sun's magnetism.

The field of *magnetohydrodynamics*, or hydromagnetics, combines the pre-

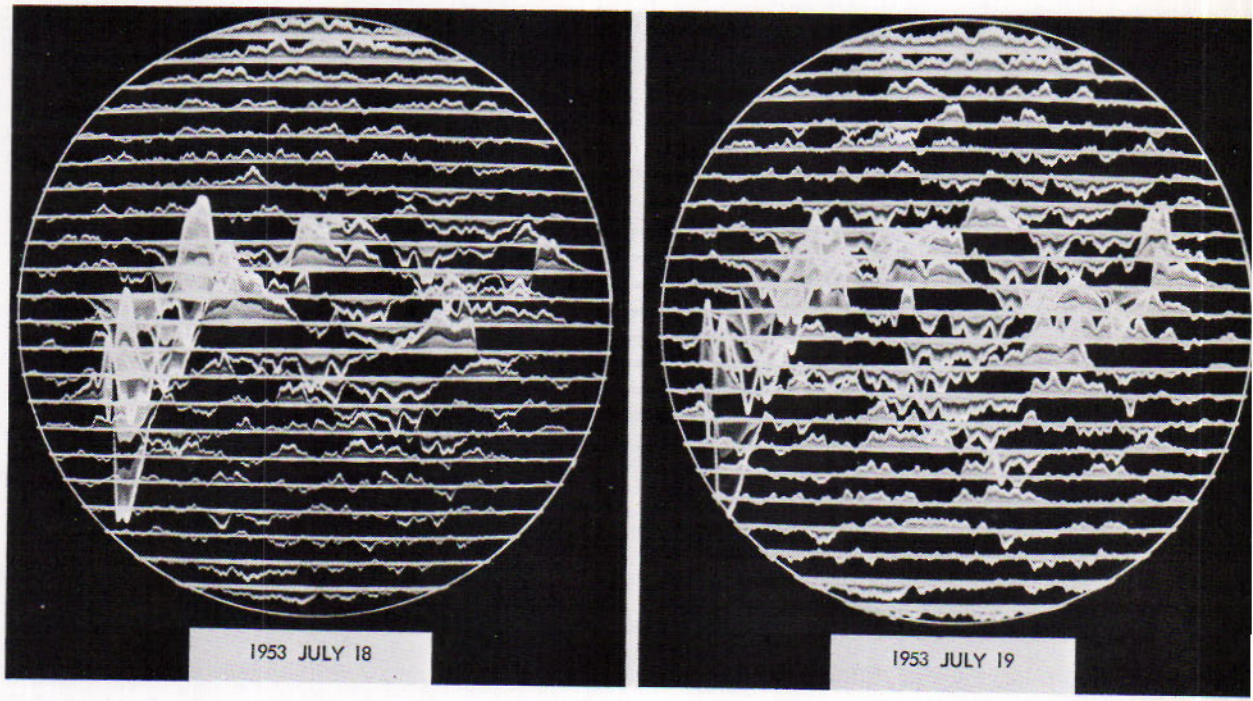


Figure 10.19
Solar magnetograms. These magnetic maps of the sun show the location, field intensity and polarity of weak magnetic fields in the photosphere of the sun, apart from sunspots. A deflection of one trace interval corresponds to a field of about one gaus. The small deflections of opposite polarity near the north and south poles are indicative of the sun's "general magnetic field," while the extended fields near the equator arise from characteristic "bipolar magnetic" regions that sometimes produce spots. North is at the top; east is at the right. *(Photograph from the Hale Observatories)*

viously independent procedures of hydrodynamics and electrodynamics; it deals with the large-scale behavior of an electrically conducting fluid in a strong magnetic field. In most astrophysical applications the fluid consists of ionized gases. The main subjects of the applications, as listed by the Swedish scientist H. Alfvén, are the following. (1) The origin of the earth's magnetic field, and the fields of other planets and of stars. (2) The electromagnetic states of the upper atmosphere and of interplanetary space; this includes the radiation belts around planets, the magnetic storms, and auroras. (3) Solar activities. (4) The magnetism of stars. (5) The physics of interstellar matter, including the dynamics of cosmic clouds.

THE CHROMOSPHERE AND CORONA

It is the custom to speak of the photosphere as the surface of the sun, and of the more nearly transparent gases above it as constituting the sun's atmosphere. The photosphere, as we have noted, is the region from which most of the sunlight emerges. The lowest stratum of the chromosphere immediately above the photosphere is the most effective in producing the dark lines in the solar spectrum.

10.20 The Visible Solar Spectrum

This is an array of colors from violet to red interrupted by thousands of dark lines, which are known as *Fraunhofer lines* in honor of their discoverer, the German optician Joseph Fraunhofer, who was the first to distinguish them clearly in 1814. He mapped 574 and labeled the most prominent ones with letters of the alphabet, A through H, starting at the red end of the spectrum. Thus the pair close together in the yellow are the D lines. The significance of the lines remained unknown until 1859, when it was understood that they are wavelengths abstracted from sunlight chiefly by gases above the photosphere.

These lines are designated by their wavelengths expressed in *angstroms*, abbreviation Å or sometimes just A, and sometimes with the prefix λ. One angstrom is 10^{-8} cm. Thus 4000 Å denotes the wavelength of 4×10^{-5} cm. The visible region of the spectrum is between 3900 and 7500 Å. Some conspicuous lines and bands in the visible part of the solar spectrum are the following:

Fraunhofer Letter	*Wavelength*	*Identification*
A	7594 Å	oxygen (telluric)
B	6867	oxygen (telluric)
C	6563	hydrogen
D	5893	sodium (double)
E	5270	iron
F	4861	hydrogen
G	4310	composite blend
H	3968	calcium
K	3934	calcium

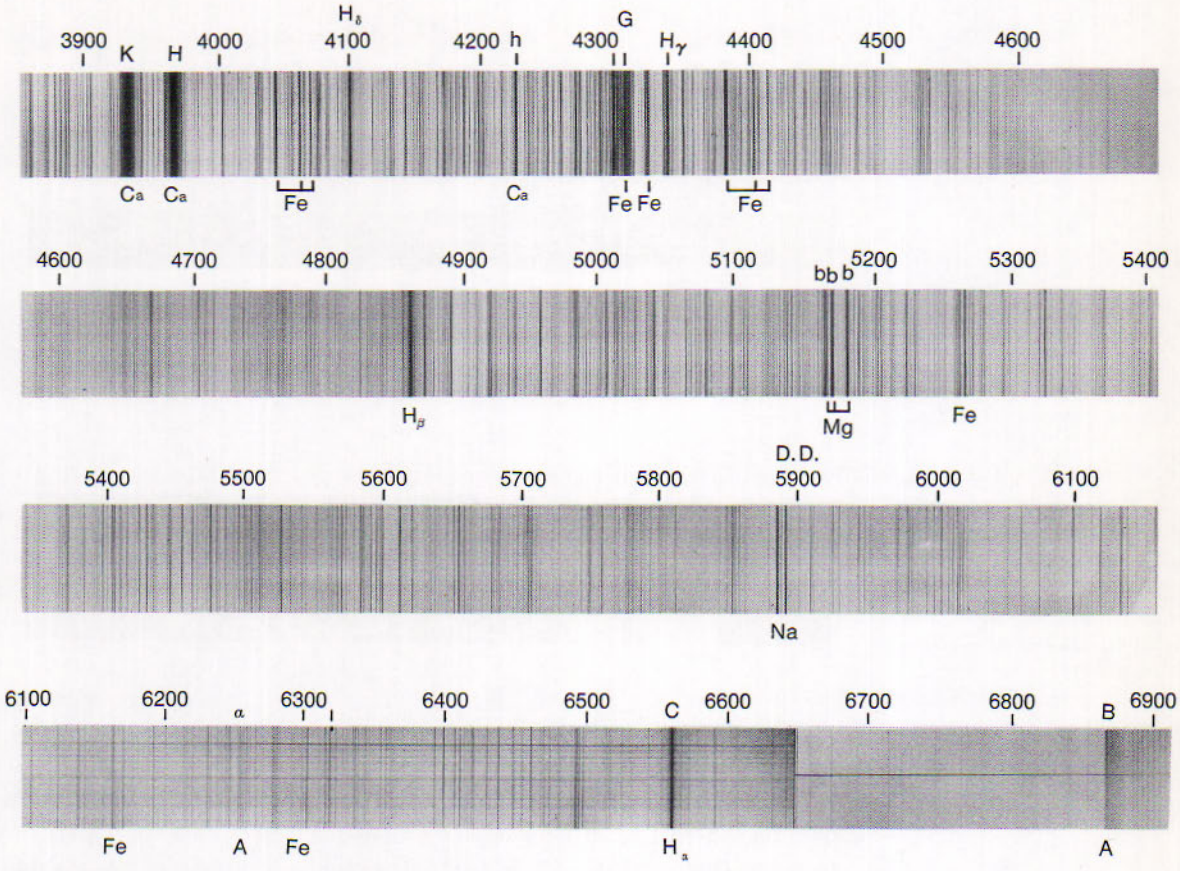


Figure 10.20
Parts of the visible solar spectrum. *(Photograph from the Hale Observatories)*

By far the strongest lines are the Fraunhofer H and K of calcium near the termination of the visible spectrum in the violet. Although Fraunhofer himself apparently did not observe the strong calcium K line, it has been designated by a letter in respect to Fraunhofer. Not all the lines are of solar origin. There are also *telluric bands* produced by absorption of sunlight in the earth's atmosphere. The Fraunhofer A and B bands are identified with terrestrial oxygen.

10.21 The Ultraviolet Solar Spectrum

Beyond the shortest wavelength visible to the eye, the solar spectrum can be photographed ordinarily to about 2900 Å. Beginning there, the sun's radiations are absorbed by ozone molecules and other constituents of our atmosphere. The extreme ultraviolet rays, X rays and gamma rays,

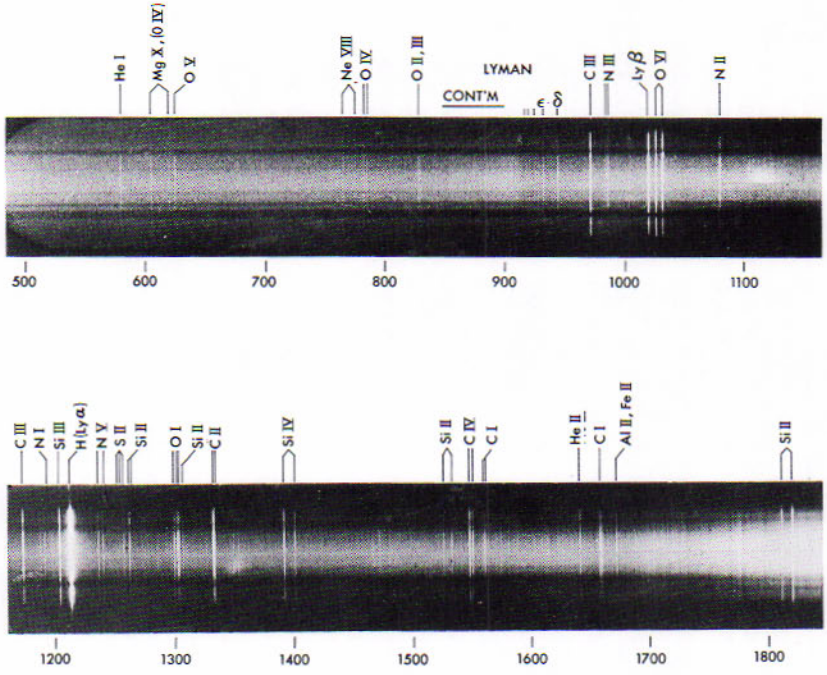


Figure 10.21
The solar spectrum from λ500 Å to λ1800 Å. Note the emission lines of ionized carbon, silicon, neon and oxygen. *(Official U.S. Navy Photograph)*

which would preclude the existence of large molecules of living matter, are thereby prevented from reaching the ground. Studies of the extreme ultraviolet solar spectrum are being made in photographs from rockets above the obstructing atmosphere.

In a spectrogram from an Aerobee rocket at a height of about 200 km on March 13, 1959 (Fig. 10.21), the Naval Research Laboratory group of J. D. Purcell, D. M. Packer, and Richard Tousey observed the major features of the ultraviolet solar spectrum as far as 584 Å. Absorption lines of the photosphere crowd together and finally disappear for this dispersion at about 1700 Å. Shortward from there nearly 100 emission lines of the chromosphere, about 60 of which had not been previously observed, were identified in the spectrogram. Among these are lines of neutral neon, silicon, and carbon, and of ionized magnesium, oxygen, neon, and silicon. The very strong Lyman-alpha line at 1216.7 Å, other lines of that hydrogen series, and part of the Lyman continuum beginning at 910 Å were well observed. In later photographs, where stray light was suppressed, the NRL scientists extended their investigations farther into the ultraviolet.

W. A. Rense and T. Violett of the University of Colorado have listed solar spectrum lines as far as 83.9 Å, which they have observed in other rocket spectrograms.

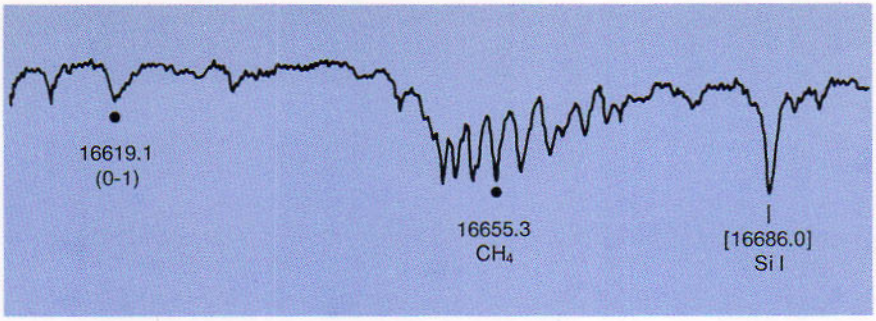


Figure 10.22
A section of the solar spectrum in the infrared. The methane band at 16655 Å is produced in the earth's atmosphere. The silicon line at 16686 Å originates in the sun. *(Tracing by McMath, Mohler and Goldberg with lead sulfide cell, McMath Hubert Observatory)*

10.22 The Infrared Solar Spectrum

This spectrum has been studied photographically to 13,500 Å, and the line patterns are clearly resolved to 24,500 Å in tracings with the lead sulfide photoconductive cell. The gross details have been shown by heat-detecting apparatus as far as 200,000 Å. Heavy bands, absorbed particularly by oxygen, carbon dioxide, and water vapor molecules of our atmosphere, conceal much of the infrared region of the sun's spectrum itself. Bands of atmospheric methane are also recognized; one of these is represented in the tracing in Fig. 10.22. Knowledge of the infrared solar spectrum has been greatly improved in recordings from above our atmosphere.

10.23 The Spectrum of the Chromosphere

The gases above the photosphere, which produce the dark lines of the solar spectrum, give a bright-line spectrum when they are observed alone, as at the time of total solar eclipse. This was called the *flash spectrum* originally, because it flashes into view near the beginning of total eclipse and disappears soon after the end of totality. When the slitless spectroscope is employed (Fig. 10.23), the bright lines are images of the thin crescent left uncovered by the moon. The thickest crescent images are produced by calcium and hydrogen. The strong red line of hydrogen is responsible for the characteristic hue of the chromosphere.

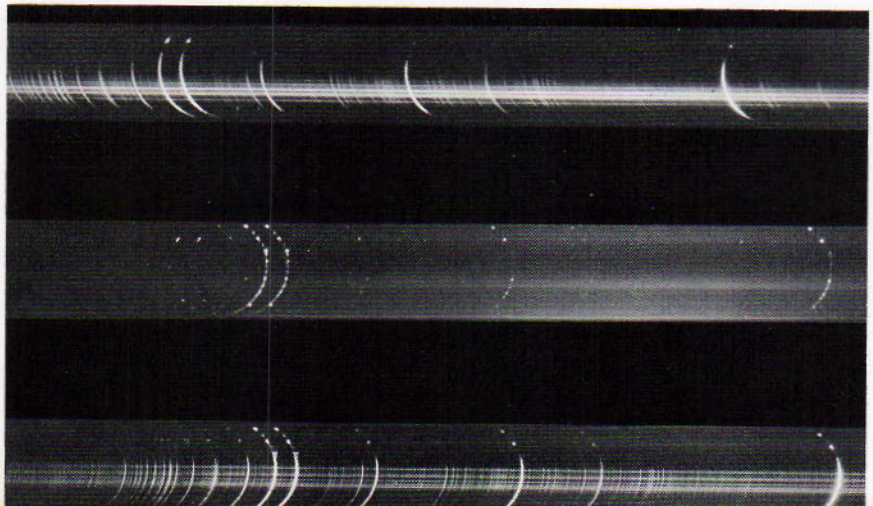


Figure 10.23
"Flash" spectrum of the solar chromosphere, obtained with a slitless spectrograph which used the edge and chromosphere of the sun as its slit. The bright pair of lines are the H and K lines of calcium. The bright line on the right is the F line of hydrogen. Photographed during the total eclipse of January 24, 1925 at Middletown, Connecticut. *(Photograph from the Hale Observatories)*

With some exceptions the bright crescents match the dark lines of the normal solar spectrum. The most conspicuous differences are found in the hydrogen and helium lines. All the hydrogen lines of the Balmer series are present in the flash spectrum, but only the first four are noticed in the dark-line spectrum. Helium lines, which are prominent in the spectra of the chromosphere and prominences, are almost entirely absent in the visible dark-line spectrum; a dark triplet of helium, discovered by the Babcocks in 1934, appears in the infrared at about 10,830 Å.

Studies of the chromospheric spectrum are now extended to the extreme ultraviolet, as we have seen (10.21), in spectrograms from rockets.

10.24 Chemical Elements in the Sun

More than 60 chemical elements are definitely recognized in the sun, according to a list prepared by H. D. Babcock and Charlotte Moore. These elements have been identified by comparing their laboratory spectra with lines in the solar spectrum. At the high temperatures of the sun the elements exist generally as dissociated atoms. The molecules of 18 compounds are recognized by their characteristic spectrum bands; they are found mainly in the cooler regions of sunspots, but a few hold together above the undisturbed photosphere as well. All the natural elements are presumed to be present in the sun.

Hydrogen is the most abundant element in the sun's atmosphere, and helium is second (Table IV, Chapter 12). These two elements also predominate in the sun's interior, in the stars, and in the universe generally. Hydrogen contributes 55 per cent of the mass of the cosmic material, helium about 44 per cent, and the heavier elements the remainder. Exceptions to these proportions occur in the earth and other smaller members of the planetary system from which most of the lighter gases have escaped.

10.25 Spectroheliograms

Spectroheliograms are photographs of the chromosphere and prominences taken outside eclipse in the light of a single spectral line. These revealing photographs show how the gases of the element are distributed above the sun's surface. They are taken with the *spectroheliograph*, a special adaptation of the spectroscope. The operation of this instrument, which employs two slits, is as follows.

The image of the sun is focused by the telescope objective onto the first slit, which admits the light from a narrow strip of the sun's disk to the diffraction grating. The spectrum produced by the grating falls on a screen containing a second slit parallel to the first. This slit allows the light from only a limited region of the spectrum to pass through to the photographic plate. By a slight rotation of the grating any part of the spectrum can be brought upon the second slit, for example, the "dark" K line of calcium. It will be understood that the dark lines of the solar

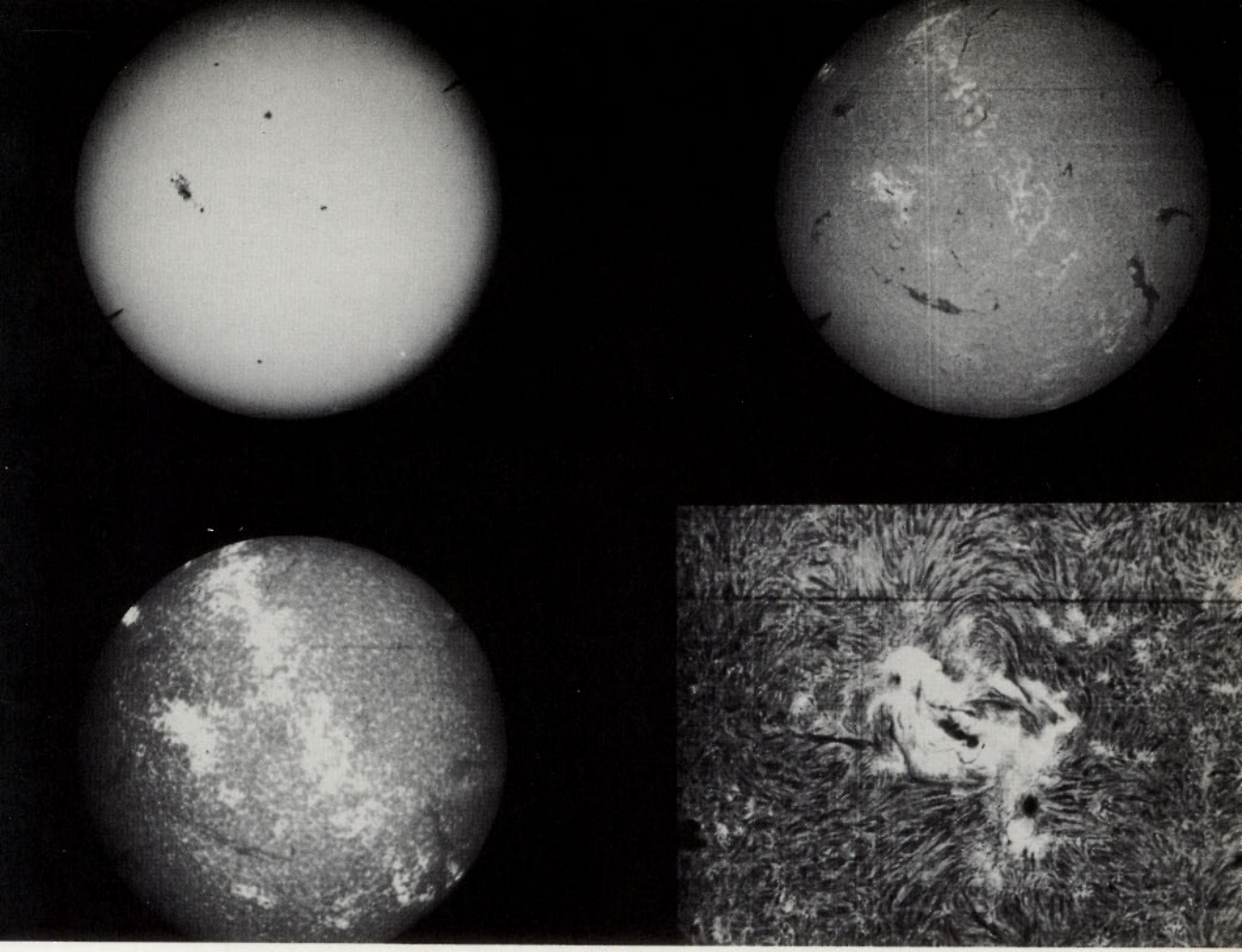


Figure 10.25
Four photographs of the sun: By ordinary light (upper left); hydrogen (Hα) light (upper right); calcium spectroheliogram (lower left) and enlargement of hydrogen spectroheliogram (lower right). *(Photograph from the Hale Observatories)*

spectrum are not devoid of light; they appear dark by contrast with the brighter continuous spectrum.

The operation so far described would give a photograph of only a narrow strip of the sun's disk in calcium light. When the first slit is moved across the disk and the second slit correspondingly over the photographic plate, the result is a spectroheliogram of as much of the disk as is desired. The spectrohelioscope accomplishes the same result for visual observations. Here the two slits are made to oscillate rapidly enough to give a persistent image of a part of the sun's disk in the light of a spectrum line.

An alternate device is the bifringent filter, a sharp-band monochromator which transmits the whole image of an extended source instead of assembling it by scanning; such a filter may consist of a succession of quartz crystal and polaroid sheets. This filter is useful for both photographic and visual purposes. It is also employed with the coronagraph (10.31) for observing the corona outside eclipse.

10.26 Features of the Chromosphere

The chromosphere rises to a height of 8000 km or more. Photographed beyond the edge of the sun, it appears as a fairly uniform layer from which many narrow bright *spicules* keep emerging, forming a grass-like upper surface. Each spicule reaches a height of about 10″ (about 14,500 km) above the edge in a few minutes and then disappears. The association of the spicules with the granules of the photosphere has been suggested.

Spectroheliograms and other monochromatic photographs of the sun are commonly made with either the K line of calcium or the C line of hydrogen. The former show the disk mottled with bright calcium *flocculi;* the word flocculus means a tuft of wool. Where these are bunched around sunspot groups and other active centers, they are known as *plages.* Photographs in hydrogen light also show flocculi and plages. Photographed from rockets in the light of the Lyman-alpha line of hydrogen, the chromosphere shows a grosser plage structure and much greater contrast than in the lower-level spectroheliograms. The solar image in X-ray wavelengths (20 to 60 Å) has also been photographed.

The hydrogen photographs often show masses of this gas dark against the brighter background and drawn out in filaments which may extend as far as 60° in longitude; they are huge curtain-like structures hanging above the chromosphere. Whenever the dark filaments are carried beyond the edge of the disk by the sun's rotation, they appear bright against the sky. These are the prominences.

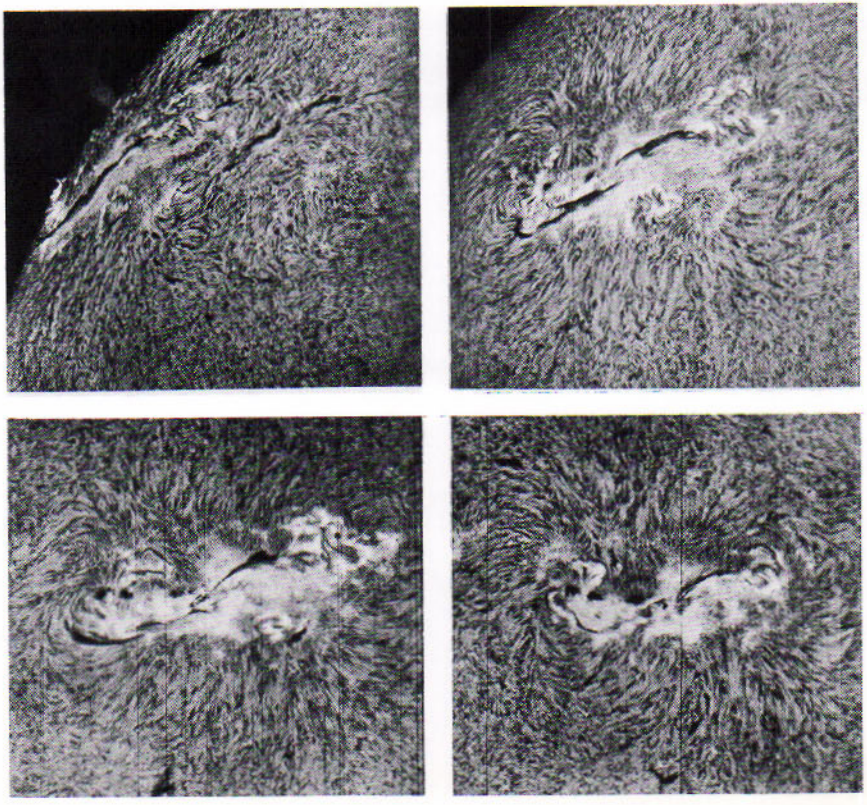
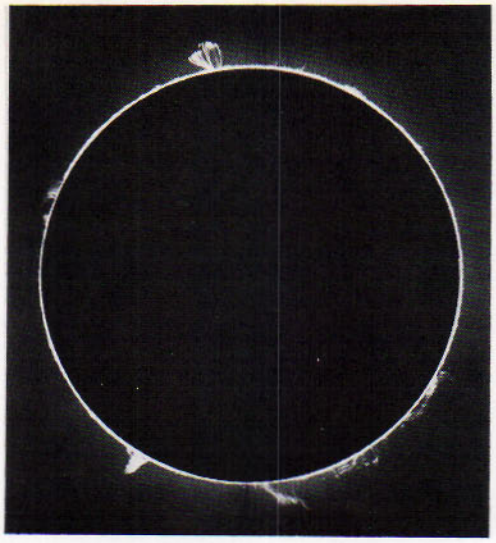


Figure 10.26
A section of the sun photographed in red hydrogen light (Hα) on August 3, 5, 7 and 9, 1915. *(Photograph from the Hale Observatories)*



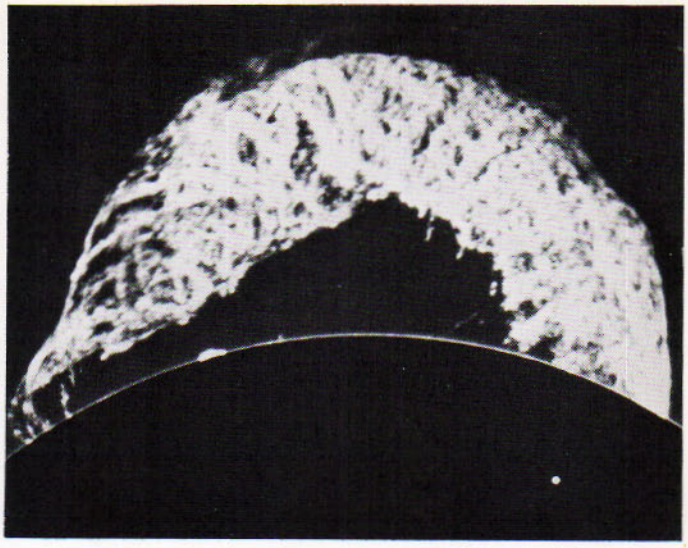


Figure 10.27A (left)
Solar prominences. Whole edge of the sun photographed with calcium K line on December 9, 1929. *(Photograph from the Hale Observatories)*

Figure 10.27B (right)
The magnificent solar prominence of June 4, 1946. Notice size of the earth indicated by white dot for comparison. *(Photograph from the High Altitude Observatory, Boulder, Colorado)*

10.27 Solar Prominences

The solar prominences are best observed beyond the edge of the sun. They show wide diversity in their behavior. Some are like sheets of flame; some rise to great heights; many move downward. Their red color, contrasting with the white glow of the corona, contributes to the splendor of the total solar eclipse. The prominences are most often observed, however, in monochromatic light at other times than during eclipses.

Our appreciation of how prominences behave began with their photography on motion-picture films, initiated by R. R. McMath and now in frequent use. Continuous records of prominence and chromospheric activity are made automatically on 35-mm film in monochromatic light at a number of solar observatories in America and elsewhere.

Most prominences are of the *active* type; they originate high above the chromosphere and pour their streamers down into it. *Quiescent* prominences are the least active and have the longest lives; their most common form is the "haystack." *Eruptive* prominences are among the rarer types. These rise from material above the chromosphere, attaining high speeds and great altitudes before they vanish. A prominence of 1938 attained a speed of 725 km a second. A prominence of June 4, 1946, rose to a distance of more than 2 million km above the sun's surface.

Cooled gases that supply the prominences may collect near the tops of magnetic arches above bipolar regions, as H. W. Babcock suggests. Here they are precariously supported for a time, perhaps eventually to be propelled outward at high speed by an increase of the strength of the magnetic field.

10.28 Solar Flares

These phenomena appear as sudden increases in brightness of areas near sunspot groups and persist from tens of minutes to as long as several hours; the larger ones are likely to be associated with large and active groups. The flares are most easily observed in the light of the red line of hydrogen and are rarely seen at all in the direct view of the sun. These outbursts are often attended by active dark filaments, the projections on the disk of surge-type prominences. They are sources of intense ultraviolet radiations, high-speed particles, and bursts of noise in the radio frequencies. X-ray emission is one of the major features of the flare outbursts.

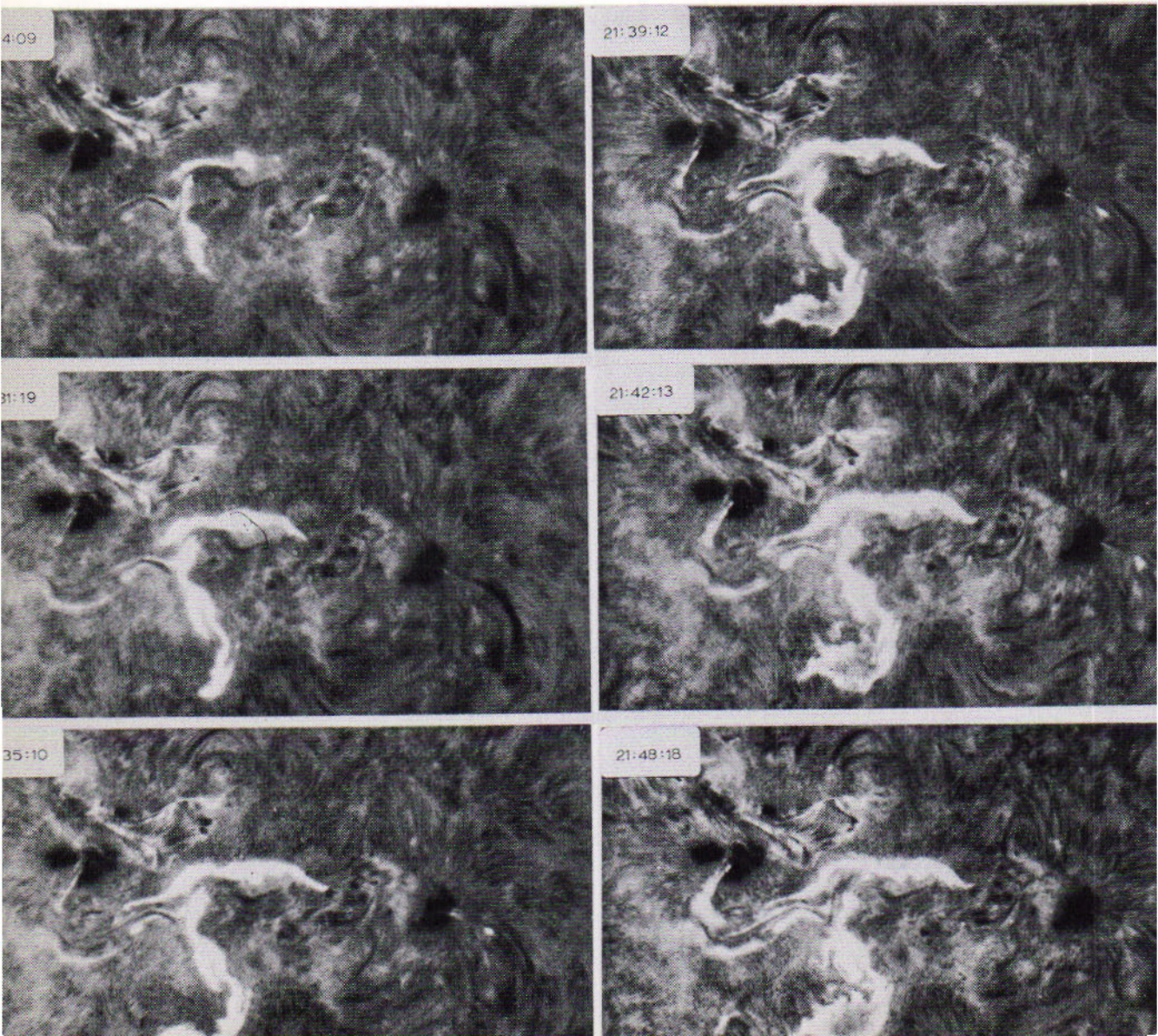


Figure 10.28A Development of a solar flare on November 21, 1969 seen on a series of spectroheliograms. *(Big Bear Solar Observatory Photograph)*

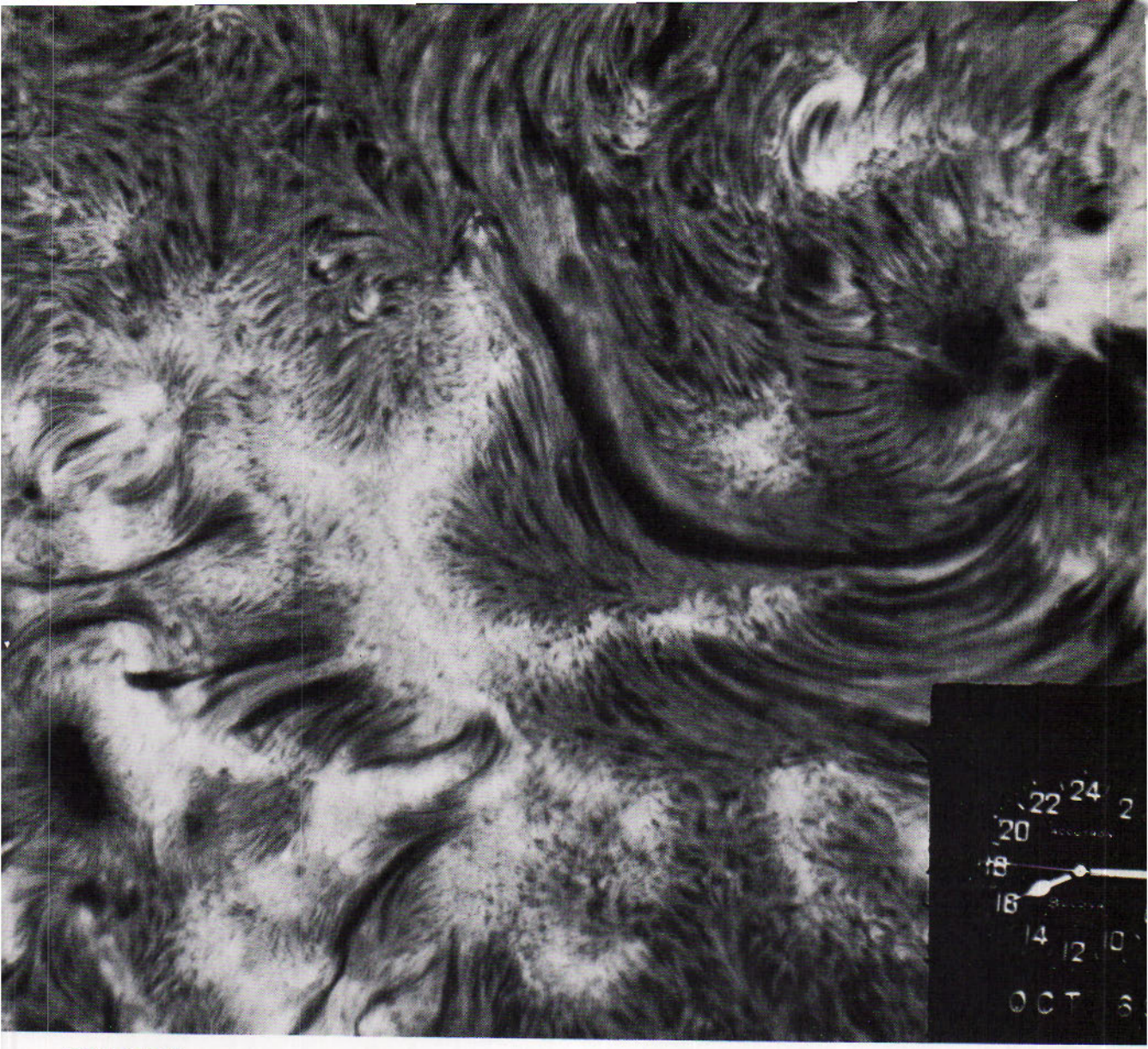


Figure 10.28B
High resolution photograph of the solar surface in Hα light. *(Big Bear Solar Observatory Photograph)*

Direct evidence of high velocity of the bright flare was first recorded by Helen Dodson at the McMath-Hulbert Observatory in photographs with a motion-picture camera. An eruption associated with a flare of May 8, 1951, rose at the edge of the sun's disk in the first minute of its life at the rate of 725 kilometers a second; it reached the height of 50,000 kilometers.

Thus the sun is a flare star, perhaps resembling in this respect the red stars that brighten suddenly and repeatedly (13.15).

10.29 The Corona

The tenuous outer envelope of the sun, seen ordinarily only during total solar eclipse, is called the *corona.* Its feeble light, which averages only half as bright as the light of the full moon, is generally lost in the glare of the sky around the uneclipsed sun. Half of this light comes from the inner corona which is within 3′, or less than 160,000 kilometers, from the sun's surface. The outer corona is characterized by delicate streamers which may extend more than two million kilometers from the edge of the sun and which contribute to the beauty of the total solar eclipse.

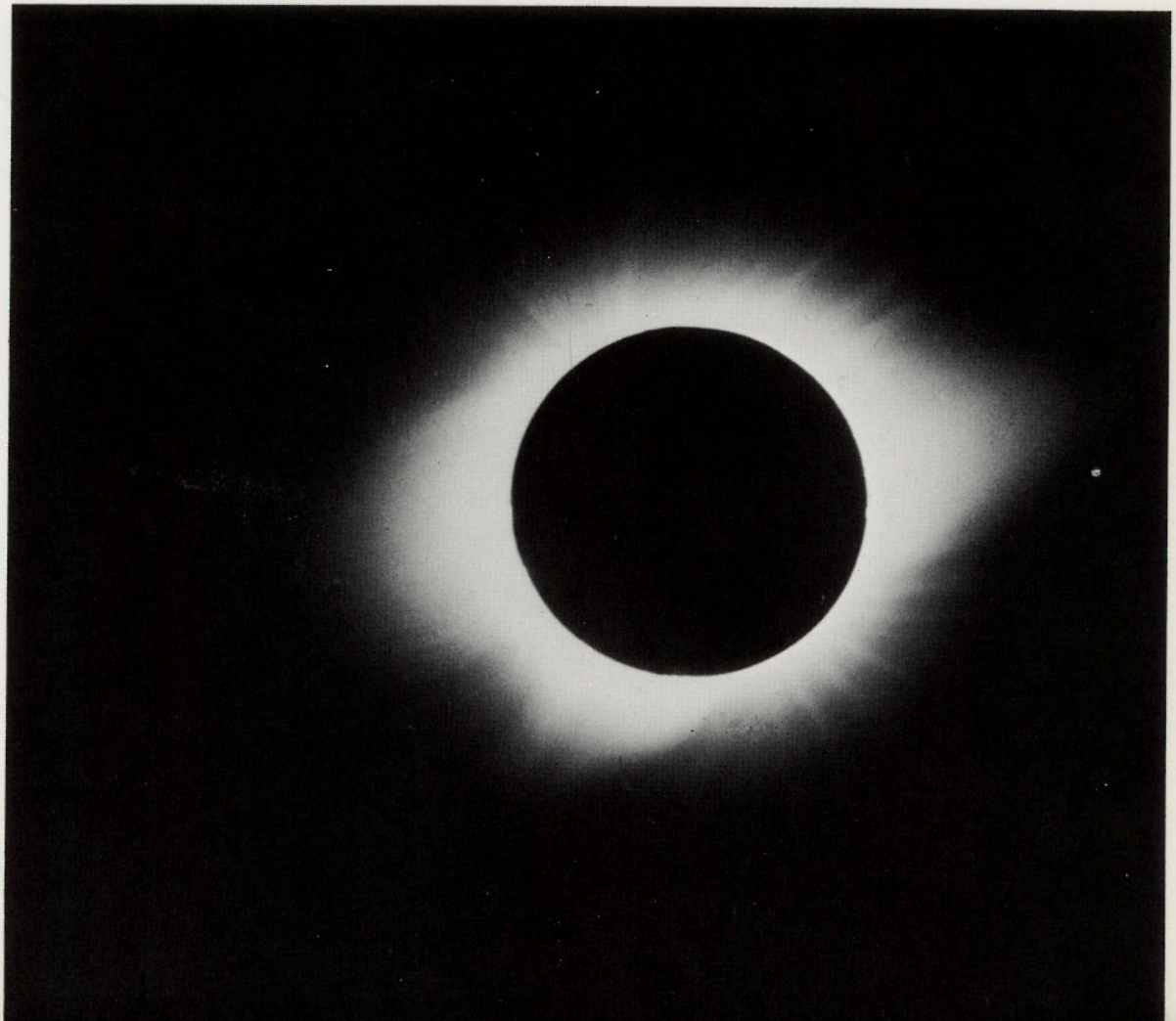
The details of the outer corona exhibit a cycle of changes in the 11-year period of sunspot frequency. Near sunspot maximum the petal-like stream-

Figure 10.29A
The corona of the sun during the total eclipse of March 7, 1970, showing the uniformly rounded shape typical of periods of sunspot maximum.

ers extend to about the same distance all around the disk, so that the corona resembles a dahlia. Near sunspot minimum short, curved streamers appear around the poles, remindful of the lines of force around the poles of a magnet, and long streamers stretch far out from the equatorial regions.

The corona proper is a highly ionized gas, glowing partly with sunlight scattered by its free electrons and partly with its own light which is emitted as shattered ions keep recombining with electrons. This light of the true corona is confused with the inner zodiacal light (9.31), which is sunlight scattered by dust particles, and with the blue light of the sky.

Figure 10.29B
The sun's corona during the total eclipse of August 31, 1932 at Fryburg, Maine. Note the equatorial extension of the corona which is typical during sunspot minimum. *(Georgetown University Observatory Photograph).*



10.30 The Spectrum of the Corona

The spectrum of the corona shows a continuous background containing dark Fraunhofer lines and a superposed array of bright lines called the E corona or emission corona. The background consists of two components, which are designated as K and F.

The K-component of the spectrum is the brighter within a solar radius from the sun's edge. It resembles the solar spectrum except that the dark lines are almost completely washed out by Doppler effects of the swiftly moving scattering electrons; only the H and K lines are clearly seen, combined as a darkening 300 Å wide. The F-component, often called false corona, caused by the zodiacal light, is a faint replica of the solar spectrum, and the strength of its dark lines shows the fraction of light it contributes to the total. This component is equally bright at equator and poles; it decreases more gradually outward than does the other.

The E component was for long an enigma since its lines did not agree with any known terrestrial elements. As a result it was believed that some new element was involved and it was given the name *coronium.* Actually the lines arose from known elements in abnormal conditions. At the high kinetic energies encountered in the corona, collisions result in *highly ionized atoms* which give rise to so-called "forbidden" transitions. These *transitions* require high temperatures and very low densities.

Twenty-seven bright lines are well observed in the region of the spectrum which has been photographed. The most conspicuous ones are at 3388 Å in the ultraviolet, 5303 Å (the brightest line) in the green, 6375 Å in the red, and 10,747 and 10,798 Å in the infrared. The stronger lines were first detected in an extrasolar source by Adams and Joy in the spectrum of the recurrent nova RS Ophiuchi. They have since been recognized in the spectra of several novae, including Nova Herculis 1960. The identifications of the emission lines were announced in 1942 by the Swedish scientist B. Edlén. They are unusual lines of from 9 to 15 times ionized atoms of iron, nickel, calcium, and argon. The brightest ones are ascribed to iron atoms which have lost from 9 to 13 electrons. Hydrogen and helium are not represented in the spectrum because these less complex atoms would be kept permanently stripped of electrons in these conditions.

Such high degree of ionization and the considerable widths of the emission lines themselves indicate a kinetic temperature of the corona of the order of a million degrees.

10.31 The Corona Outside Eclipse

Until recent years, the sun's corona was not observed at all except during the infrequent moments of total solar eclipse. Its faint glow seemed hopelessly concealed by the glare of the sunlit atmosphere, which at 1′ from the edge of the sun is from 500 to 1000 times as bright as the coronal light.

In 1930, B. Lyot succeeded in photographing features of the inner corona

with a special type of telescope, the *coronagraph,* at the Pic du Midi Observatory in the French Pyrenees. This instrument depends for its effectiveness mainly on the quality and spotlessness of its objective, so that it brings in a minimum of scattered sunlight, and also on the clearer air of the mountains.

In addition to the original instrument, there are a dozen or more coronagraphs now in use at other mountain observatories, including the High Altitude Observatory at Climax, Colorado, and the Sacramento Peak Observatory at Sunspot, New Mexico. The coronagraph is also useful for observing the chromosphere and prominences outside eclipse.

Wider extension of the corona has been recorded with radio telescopes. The evidence was first obtained while the sun in its annual circuit of the heavens was passing the Crab nebula, an intense radio source. The radiation from this source was found to be scattered and dimmed by the solar corona out to a distance of 30 or 40 solar radii from the sun. In 1960, O. B. Slee of the Radiophysics Laboratory, Sydney, Australia, observed this effect on a dozen fainter radio sources as well, and was able to determine the amount of the scattering at 350 points in the corona. He reported that the extended corona was an ellipse about 110 by 80 solar radii in size; it was similar in shape to the corona in visual light at that stage of the sunspot cycle.

10.32 Radio Reception from the Sun

The reception of radiation from the sun at radio wavelengths was first recorded in 1942 at radar defense stations in Great Britain, but was not then published. The radiation was in excess of what was expected from thermal radiation at the sun's surface; and the very high kinetic temperature of the corona implied by Edlén's identification of its emission lines in that year had not yet become well known. The radiation is now being studied with radio telescopes at wavelengths from about 1 cm to 15 meters. The shortest of these are emitted from the lower chromosphere and the longer lengths from the corona. Drift curves at these longer wavelengths show a limb brightening for the sun.

The distinction is drawn between the quiet sun and the active sun. From the *quiet sun,* around sunspot minimum, the strength of the radio reception is near the thermal values. From the *active sun,* when sunspots, plages, and flares occur in great centers of activity, increased radiation and irregular *bursts* of much greater strength are superposed on the quiet sun emission. *Outbursts* of very great strength and lasting for minutes occur when large solar flares are observed. These have been attributed to streams of protons and other ions sucked out by the flares and propelled outward at speeds of the order of 1600 km/sec. They begin to disturb the ionized gas of the corona in seconds, causing the radio outbursts, and reach the earth a day later to produce our magnetic storms.

The background radio spectrum of the quiet sun is quite different from

the active sun and deviates from the expected black body (thermal) curve. These deviations indicate the temperature of the corona to be above (10^6) °K when the sun is quiet and temperatures in excess of (10^{10}) °K when the sun is active. The radiations arise from electrons moving at relativistic velocities and from electrons spiraling in strong magnetic fields.

ASSOCIATED IONOSPHERIC DISTURBANCES

When emissions from a flare finally reach the earth, magnetic storms occur when they interact with our magnetic field. During these periods the most spectacular of the auroras occur. We include the earth-sun interaction to emphasize that even a normal star is not quiet and to demonstrate what we should expect from other stars, but should never be able to see.

10.33 Magnetic Storms

The appearance of a large solar flare is likely to be the signal for the deterioration of our radio communications in the higher frequencies. Strong ultraviolet radiations arrive from the sun with the visible evidence of the flare. These radiations disrupt ionized layers of the upper atmosphere, which normally reflect our radio beams back to the ground. This effect and the magnetic storm which follows are most pronounced when the flare appears in a large sunspot group near the central meridian of the sun's disk.

Geomagnetic storms are unusual agitations of the earth's magnetic field; they are indicated by erratic gyrations of the compass needle. They occur when streams of ions from the solar disturbance arrive here, a day or more after the flare is seen. These effects are sometimes accompanied by strong induced earth-currents, which can interfere with our communications by wire. The incoming protons also cause the primary glow of the aurora in the upper atmosphere.

A good example of the association between solar and ionospheric disturbances was inaugurated by a brilliant solar flare on September 28, 1961. Two days later, a fine auroral display was observed in the north-

eastern United States. Three photographs of the flare from a series taken by G. E. Moreton at the Lockheed Solar Observatory and some pictures of the aurora are shown in *Sky and Telescope* for November, 1961.

10.34 The Aurora

Characteristic of many displays of the "northern lights" of our hemisphere is a luminous arch across the northern sky, having its apex in the direction of the geomagnetic pole. Rays like searchlight beams reach upward from the arch, while bright draperies may spread to other parts of the sky, altogether often increasing its brightness from 10 to 100 times that of the ordinary night sky. These displays, which also occur in the southern hemisphere, are likely to be especially spectacular when large spot groups appear on the sun.

Auroral displays are most frequently observed in two zones centered about 23° from the two magnetic poles. The zone in North America extends from Alaska across Hudson Bay to northern Labrador. The southern boundary of auroral visibility is normally a parallel through San Francisco, Memphis, and Atlanta. Auroras have appeared on rare occasions as far south as Mexico City and Cuba. The southward shifting of this zone at the maximum of the sunspot number cycle strengthens the correlation that is found between the frequencies of auroras and sunspots.

The light of the aurora is believed to be produced by streams of protons and electrons, which emerge from solar upheavals and are trapped by the earth's magnetic field. The arches, as A. B. Meinel has explained, are caused mainly by the focusing of protons by the magnetic field in narrow bands, where they combine with electrons already in the ionosphere to form normal hydrogen atoms with the emission of light. The arches are subsequently broken into rays and draperies by impacts of streams of electrons, which combine with ionized atoms of oxygen and molecules of nitrogen. These effects are greatest at altitudes of 80 to 160 kilometers but may extend to heights of several hundred kilometers.

Most of the light of an auroral display is produced in the colors green, red, and blue by the combining of electrons with oxygen atoms and nitrogen molecules. Hydrogen combinations contribute red and blue-green light. The great display of February 10–11, 1958, contained a remarkable intensity of red light from oxygen at high levels in our atmosphere.

10.35 The Airglow

The airglow is permanently suffused over the sky day and night. It is caused by excitation of air molecules and atoms by energy coming from outside and presumably from the sun. The glow appears between altitudes of 100 to 190 kilometers. When it was first detected, it was called the "permanent aurora."

The night airglow gives us twice as much light as do all the stars,

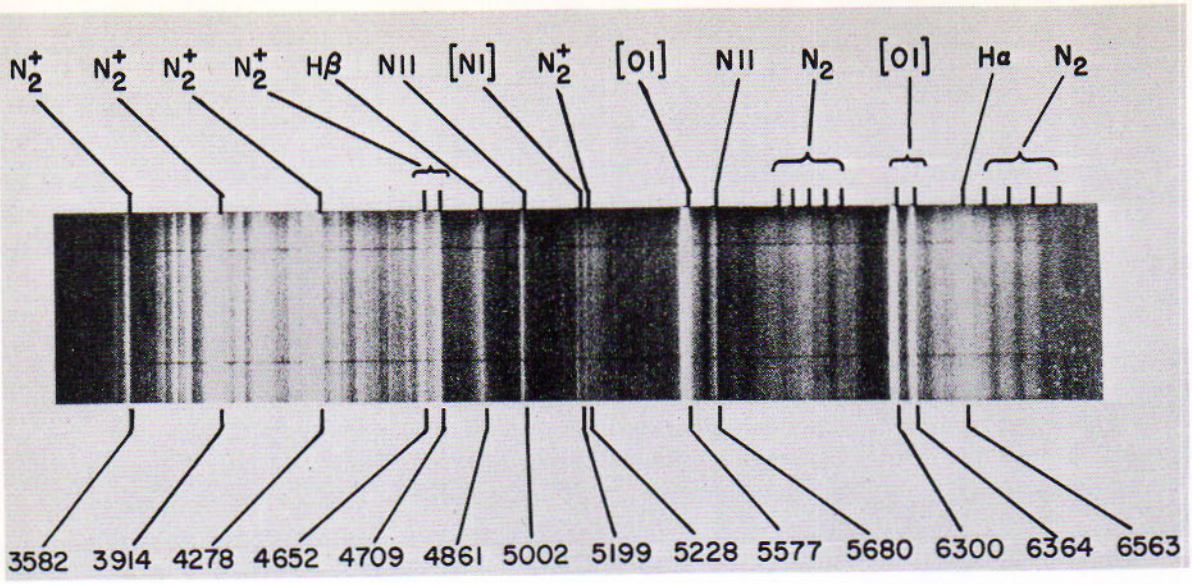


Figure 10.35
Spectrum of an aurora. Notice the preponderance of ionized molecular nitrogen in emission in the violet region of the spectrum. *(Courtesy of J. W. Chamberlain, Kitt Peak National Observatory)*

according to C. T. Elvey and F. E. Roach. It places a limit on the durations of exposures in direct celestial photography before the plates become hopelessly fogged. Almost invisible to the eye, it is studied effectively by use of the photoelectric cell and filters. The glow is faintest overhead and reaches its greatest intensity 10° above the horizon, where we look through a greater thickness of air. The night airglow is mainly at 4 wavelengths: (1) the green line of oxygen at 5577 Å; (2) the red line of oxygen at 6300 Å; (3) the yellow line of sodium at 5893 Å; (4) the very strong infrared line of hydroxyl (OH) at about 10,000 Å, which has been ascribed to impacts of incoming protons on ozone molecules.

The *twilight airglow* is 100 times as intense as the night airglow, but is not detected by the eye because of the brighter sky. This is believed to be caused by action of direct sunlight on the air molecules.

REVIEW QUESTIONS

1. Explain the solar limb darkening.
2. If the temperature of the sun were to increase by a factor of three, how much would its total energy output increase and where would the maximum be?
3. Define and describe: the photosphere, chromosphere, corona.
4. Explain the limb brightening at meter wavelengths.
5. What are the two methods for getting the rotation period of the sun?
6. Compute the new solar constant under the conditions given in question 2.
7. Why do the various values for the sun's effective temperature calculated from the basic formulae not agree?
8. Describe the typical sunspot history.

9. What does doubling the magnetic field strength do to the Zeeman splitting? What does working at twice the wavelength do to this splitting?
10. What causes the E corona? the K corona?
11. Lines similar to the E corona appeared in the spectrum of Nova Herculis. What can we say about the condition where these lines arose in the nova?

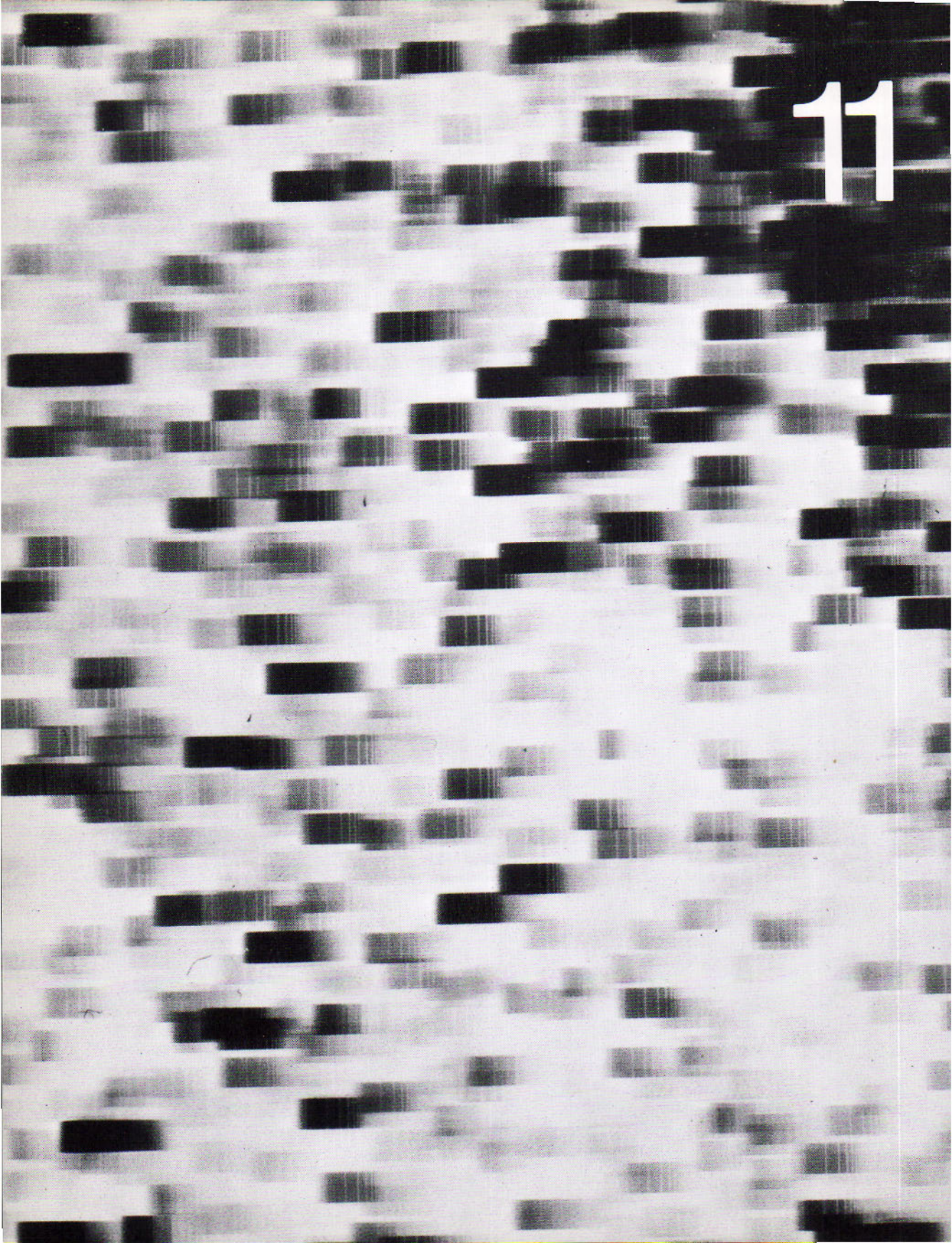
REFERENCES

Abetti, Giorgio, *Solar Research,* New York: The Macmillan Co., 1963.
Brandt, John C., *The Sun and Stars,* New York: McGraw-Hill Book Co., 1966.
Billings, Donald E., *A Guide to the Solar Corona,* New York: Academic Press, 1966.
Menzel, Donald H., *Our Sun,* Cambridge, Mass.: Harvard University Press, 1959.

FOR FURTHER STUDY

Aller, Lawrence H., *Astrophysics, the Atmospheres of the Sun and Stars,* New York: Ronald Press, 1953.
Brandt, John C. and Paul W. Hodge, *Solar System Astrophysics,* New York: McGraw-Hill Book Co., 1964.
Kuiper, Gerard P., ed. *The Sun,* Chicago: University of Chicago Press, 1954.
Tandberg-Haussen, Einar, *Solar Activity.* Waltham, Mass.: Blaisdell Co., 1967.

The 16-inch (40.6 cm) coronagraph at Sacramento Peak Observatory, New Mexico. *(Sacramento Peak Observatory photograph)*


11

THE STARS

DISTANCES OF THE STARS — MOTIONS OF THE STARS — STELLAR SPECTRA — MAGNITUDES OF THE STARS — LUMINOSITIES OF THE STARS

In our studies of the stars we first consider the basic data of observation, how they are determined and in what terms they are expressed. These are the distances of the stars, their motions relative to the sun, the character of their spectra, and their relative brightness.

Objective prism plate; NE of NGC 3532. *(A. G. Davis Philip)*

DISTANCES OF THE STARS

The obvious proof of the heliocentric theory rested upon the discovery of the annual parallax of a distant star. Only after the discovery of the aberration of starlight did astronomers begin to appreciate how far away the stars really are. The parallax of the nearest star is not even as large as 1 second of arc.

11.1 The Parallax of a Star

While the earth is revolving around the sun, a nearer star seems to be describing a little orbit with respect to more distant stars. This apparent orbit has almost the same form as the aberration orbit (2.21), but 90° out of phase; it varies from nearly a circle for a star at the ecliptic pole to a straight line for a star at the ecliptic. It is much smaller than the aberration orbit even for the nearer stars and shrinks to imperceptible size for the more distant ones.

The *heliocentric parallax* of a star is half the major axis of the parallax orbit, with slight correction for the eccentricity of the earth's orbit. It is otherwise the greatest difference between the directions of the star as seen from the earth and sun during the year. We refer to it simply as the *parallax* of the star.

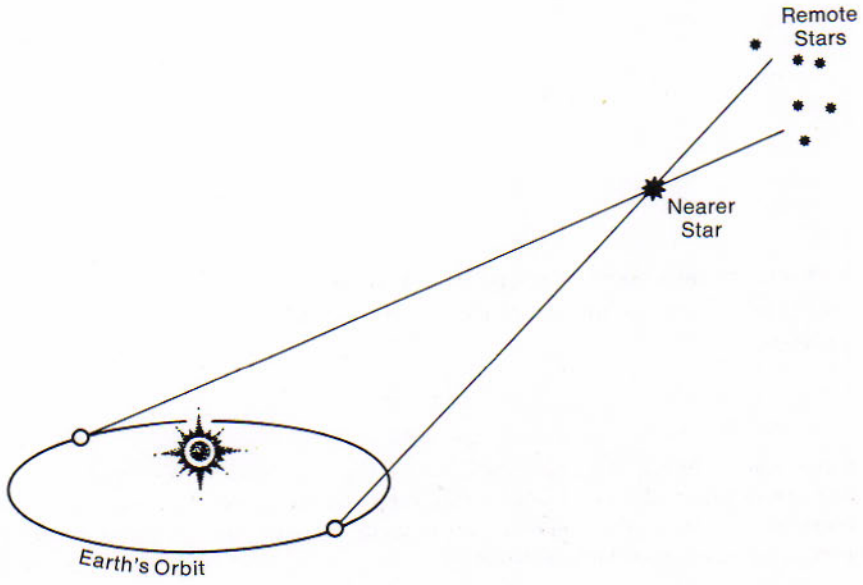


Figure 11.1
The parallax of a star. Because of the earth's revolution, the nearer star appears to oscillate annually relative to more remote stars.

After Copernicus had proposed the heliocentric theory of the planetary motions, the inability of astronomers to detect the parallax displacements of stars meant either that the earth was stationary after all or that the stars are enormously more remote than they were believed to be at that time. When the earth's revolution was decisively demonstrated by the discovery of the aberration of starlight, the search for perceptible parallaxes was renewed as a promising means of determining the distances of stars.

It was not until the years 1837–39 that the attempts to observe this effect finally met with success. The parallax of Vega was announced by F. G. W. Struve at Dorpat in 1837, of 61 Cygni by F. W. Bessel at Königsberg in 1838, and of Alpha Centauri by T. Henderson at the Cape of Good Hope in 1839. Up to 1904, fairly reliable parallaxes of 55 stars had been observed by comparing transit times for close reference stars over a period of several years. The early visual methods were adequate to deal only with the larger parallaxes.

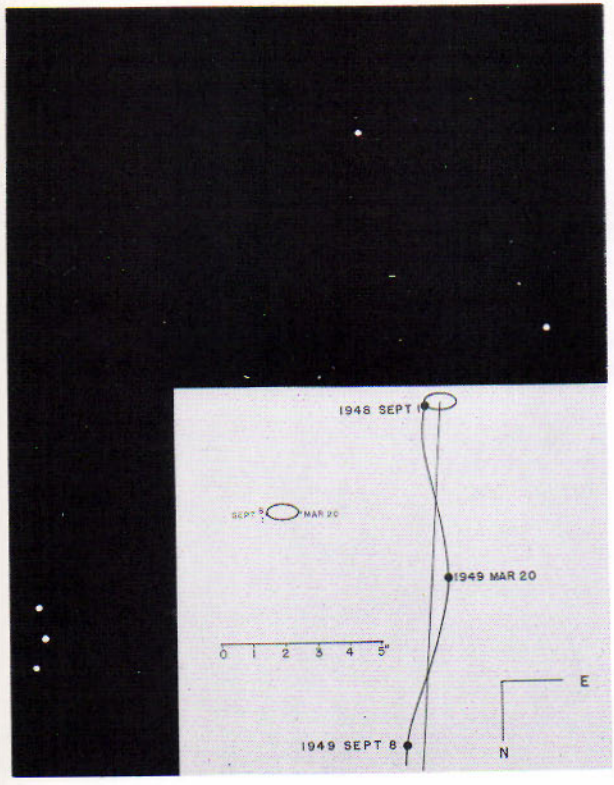
11.2 Measurements of Parallax

The present photographic method of measuring stellar parallaxes was developed, in 1903, by F. Schlesinger with the 40-inch Yerkes refractor. He attained such greatly increased accuracy that other astronomers with long-focus telescopes were encouraged to enter this exacting and important field. The latest *General Catalogue of Trigonometric Stellar Parallaxes* of the Yale Observatory (1952) lists 5822 stars with known direct parallaxes.

The parallax of a star is determined by observing its change of position relative to stars that are apparently close to it, but are really so much farther away from us that they are not greatly affected by the earth's revolution. Sets of photographs of the region are obtained at intervals of about 6 months, when the star under investigation appears near the extremities of its small parallax orbit. Two sets are not enough, because the star is also moving in a straight line with respect to the more distant stars and it accordingly advances among them in a series of loops. At least 18 sets of photographs are usually required to extricate the parallax accurately. Present-day astronomers generally require many more than 18 and in extreme cases have used as many as 700 photographs. The reason for this is that the parallax of a star is a vital piece of data. For example, to determine the masses of the components of a double star we must know the distance, and the parallax enters the calculation as the third power. The accuracy from a single plate is limited by the plate error (ε) and the only way to reduce this error is by taking a large number of plates since the combined error is smaller by the square root of the number of plates.

$$M_1 + M_2 = \frac{\alpha^3}{\pi^3 P^2}$$

The measuring of the images is now carried out by automatic or semi-automatic machines and the analysis is handled by electronic computers.



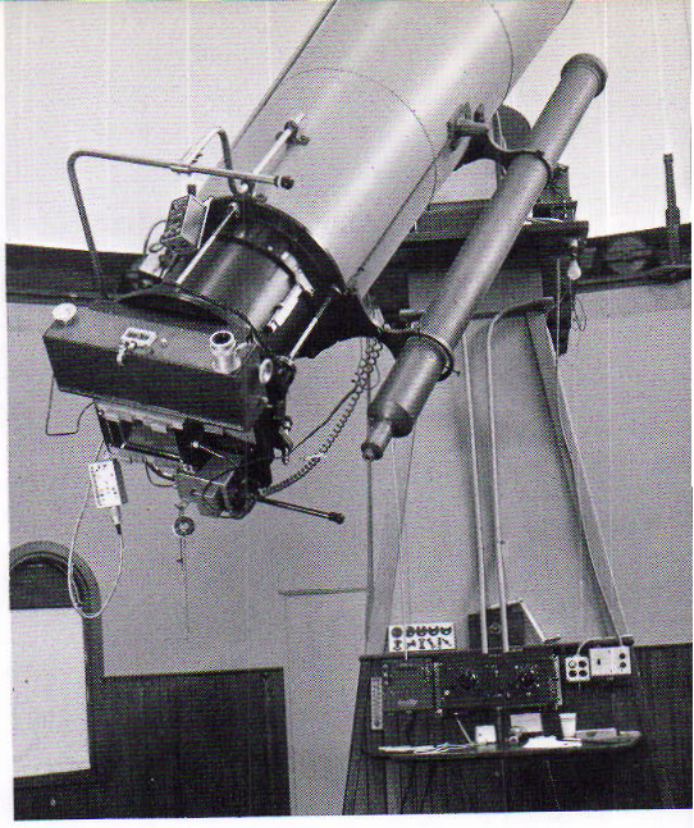
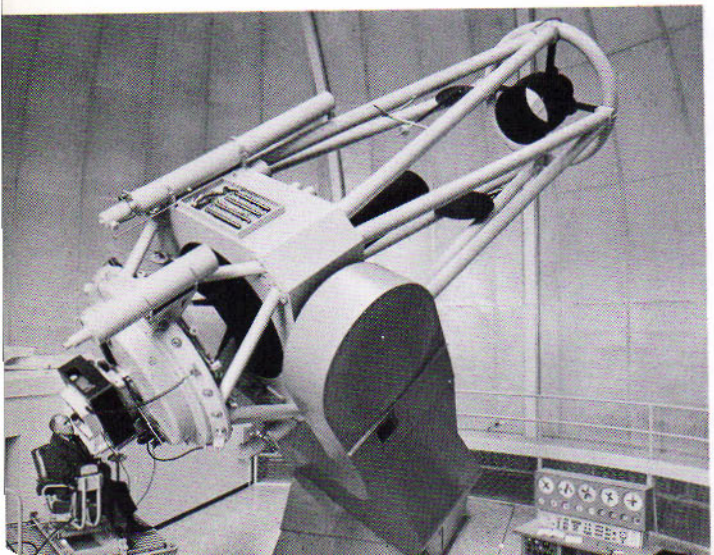


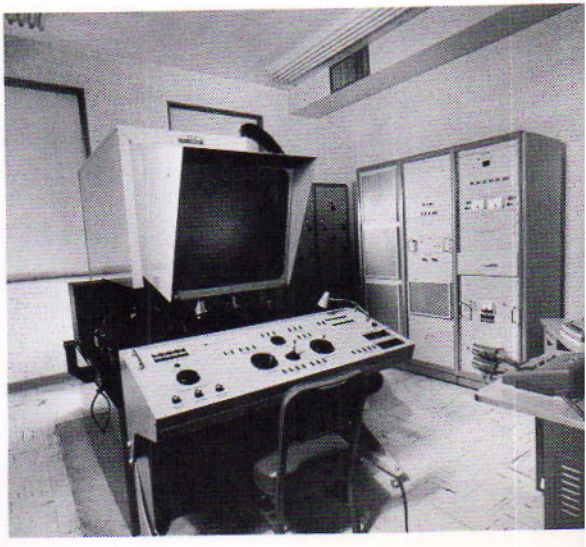
Figure 11.2A (Top left)
Parallax and Proper Mition of Barnard's Star. Three photographs have been superimposed and clearly demonstrate the parallactic shift due to the earth's motion as well as the proper motion of Barnard's Star. *(Sproul Observatory Photograph)*

Figure 11.2B (Top right)
The tailpiece of the Leander-McCormick refractor, consisting of a double-slide plate holder with automatic sequencing and timing of exposures. *(Photograph by H. Bluemel)*

Figure 11.2C (Bottom left)
New Astrometric reflector at Flagstaff, Arizona. *(Official U.S. Navy Photograph)*

Figure 11.2D (Bottom right)
Automatic measuring equipment at the Flagstaff Station. *(Official U.S. Navy Photograph)*





Special telescopes have been designed and built at the United States Naval Observatory and the Leander McCormick Observatory in an effort to achieve greater accuracy. Since stars being observed for parallax usually have large motions (indeed, they are often selected because of this) there is always the chance that patient observing will eventually reveal a periodic wobble in this motion. According to Newton's Laws, this reveals the presence of a second body as will be seen in Chapter 14.

It will be noted that the result obtained is the *relative parallax;* for the comparison stars themselves appear to be shifted slightly by the earth's revolution in the same directions as the parallax star. The *absolute parallax* is derived by making a correction not exceeding a few thousandths of a second of arc. This correction is based upon the brightness of the comparison stars and a statistical study of the motions of various classes and groups of distant stars (11.14). Although small, the correction applied often gives rise to uncertainties much to the discomfort of astronomers and a considerable effort is being made to improve these statistical parallaxes. One way around this problem is to observe parallaxes using reference points arising outside the galaxy or at least taking the intermediate step of basing the statistical parallaxes on these "external" reference points.

11.3 Units of Distance: the Parsec and the Light Year

When the star's parallax, p, has been measured, its distance D is found by the relation (Fig. 11.3):

$$\text{D (in kilometers)} = \frac{1.495 \times 10^8 \times 206{,}265}{p''}$$

$$\text{D (in miles)} = \frac{9.29 \times 10^7 \times 206{,}265}{p''}$$

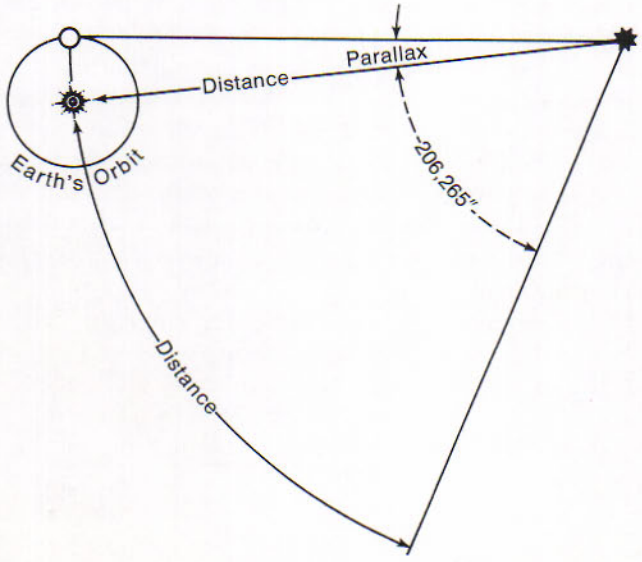


Figure 11.3
Relation Between Parallax and Distance. The radius of a circle laid off along the circumference subtends an angle, the radian, equal to 206,265". From the two sectors we have the proportion: distance of the star is to the radius of the earth's orbit as 206,265" is to the parallax.

The distance of a star expressed in miles or even astronomical units is an inconveniently large number. It is better to use larger units, either the parsec or the light year.

The *parsec* is the distance at which a star would have a *par*allax of 1 *sec*ond of arc. This distance by the above relations is 206,265 astronomical units, or 3.08×10^{13} km, or 1.92×10^{13} miles. The advantage of the parsec is its simple relation to the parallax.

$$\text{Distance (in parsecs)} = \frac{1''}{p''}.$$

The *light year* is the distance traversed by light in 1 year; it is equal to the speed of light, 2.998×10^5 km/sec, multiplied by 3.156×10^7, the number of seconds in a year. The light year is therefore 9.46×10^{12} km, or 5.88×10^{12} miles (nearly 6 million million miles). One parsec equals 3.26 light years.

$$\text{Distance (in light years)} = \frac{3''.26}{p''},$$

therefore the distance in light years is 3.26 × distance in parsecs.

As an example, consider the brightest star, Sirius, also one of the nearest, having a parallax of 0″.377. The distance of Sirius in astronomical units is 206,265″/0″.377, or somewhat more than half a million astronomical units, which amounts to about 50 million million miles. The distance in parsecs is 1″/0″.377, or 2.6 parsecs. The distance in light years is 3″.26/0″.377, or 8.6 light years.

11.4 The Nearest Stars

Data concerning the nearest stars, furnished by P. van de Kamp, are listed in Table 11.I. The magnitudes used by van de Kamp in Table 11.I are apparent visual magnitudes (m_v) discussed in section 11.20 and differ from apparent magnitudes measured on the U, B, V system which are generally used in this book. Although three of the very brightest stars in our skies are included, more than half of the nearest stars are invisible without the telescope, for their magnitudes are numbers greater than 6; and it is probable that other faint stars will eventually be added to the list. It is to be noted in passing that the annual proper motions (11.6) of the nearest stars exceed their parallaxes, and also that several of these stars are double.

The distinction of being the sun's nearest neighbor is held by the bright double star Alpha Centauri. A star of the 11th magnitude, sometimes called "Proxima," is a member of this system; it is situated a little more than 2° from the two bright stars and seems to be slightly nearer us than they are.

TABLE 11.I Stars Nearer than Five Parsecs (P. van de Kamp)

NO.	NAME	R.A. (1950.0)	DECL. (1950.0)	PROPER MOTION	PARALLAX	DISTANCE IN LIGHT YEARS	VISUAL APPARENT MAGNITUDE AND SPECTRUM A	B	C
1	Sun	....	....	....	....	....	−26.8 G2	....	....
2	Alpha Centauri†	$14^h36^m.2$	−60°38′	3″.68	0″.760	4.3	0.1 G2	1.5 K5	11 M5e
3	Barnard's star	17 55 .4	+ 4 33	10. 30	.552	5.9	9.5 M5	*	....
4	Wolf 359	10 54 .1	+ 7 19	4 .84	.431	7.6	13.5 M6e	....	....
5	Lalande 21185	11 00 .6	+36 18	4 .78	.402	8.1	7.5 M2	*	....
6	Sirius	6 42 .9	−16 39	1 .32	.377	8.6	− 1.5 A1	7.2 wd	....
7	Luyten 726-8	1 36 .4	−18 13	3 .35	.365	8.9	12.5 M6e	13.0 M6e	....
8	Ross 154	18 46 .7	−23 53	0 .74	.345	9.4	10.6 M5e	....	....
9	Ross 248	23 39 .4	+43 55	1 .82	.317	10.3	12.2 M6e	....	....
10	Epsilon Eridani	3 30 .6	− 9 38	0 .97	.305	10.7	3.7 K2	....	....
11	Luyten 789-6	22 35 .7	−15 36	3 .27	.302	10.8	12.2 M6	....	....
12	Ross 128	11 45 .1	+ 1 06	1 .40	.301	10.8	11.1 M5	....	....
13	61 Cygni	21 04 .7	+38 30	5 .22	.292	11.2	5.2 K5	6.0 K7	*
14	Epsilon Indi	21 59 .6	−57 00	4 .67	.291	11.2	4.7 K5	....	....
15	Procyon	7 36 .7	+ 5 21	1 .25	.287	11.4	0.3 F5	10.8 wd	....
16	Σ 2398	18 42 .2	+59 33	2 .29	.284	11.5	8.9 M3.5	9.7 M4	....
17	Groombridge 34	0 15 .5	+43 44	2 .91	.282	11.6	8.1 M1	11.0 M6	....
18	Lacaille 9352	23 02 .6	−36 09	6 .87	.279	11.7	7.4 M2	....	....
19	Tau Ceti	1 41 .7	−16 12	1 .92	.273	11.9	3.5 G8	....	....
20	BD +5°1668	7 24 .7	+ 5 23	3 .73	.266	12.2	9.8 M4	*	....
21	Lacaille 8760	21 14 .3	−39 04	3 .46	.260	12.5	6.7 M1	....	....
22	Kapteyn's star	5 09 .7	−45 00	8 .79	.256	12.7	8.8 M0	....	....
23	Krüger 60	22 26 .3	+57 27	0 .87	.254	12.8	9.7 M4	11.2 M6	....
24	Ross 614	6 26 .8	− 2 46	0 .97	.249	13.1	11.3 M5e	14.8	....
25	BD −12°4523	16 27 .5	−12 32	1 .18	.249	13.1	10.0 M5	....	....
26	van Maanen's star	0 46 .5	+ 5 09	2 .98	.234	13.9	12.4 wdF	....	....
27	Wolf 424	12 30 .9	+ 9 18	1 .87	.229	14.2	12.6 M6e	12.6 M6e	....
28	CD −37°15492	0 02 .5	−37 36	6 .09	.225	14.5	8.6 M3	....	....
29	Groombridge 1618	10 08 .3	+49 42	1 .45	.217	15.0	6.6 M0	....	....
30	CD −46°11540	17 24 .9	−46 51	1 .15	.216	15.1	9.4 M4	....	....
31	CD −49°13515	21 30 .2	−49 13	0 .78	.214	15.2	8.7 M3	....	....
32	CD −44°11909	17 33 .5	−44 17	1 .14	.213	15.3	11.2 M5	....	....
33	Luyten 1159-16	1 57 .4	+12 51	2 .08	.212	15.4	12.3 (M7)	....	....
34	Lalande 25372	13 43 .2	+15 10	2 .30	.208	15.7	8.5 M2	....	....
35	AOe 17415-6	17 36 .7	+68 23	1 .31	.207	15.7	9.1 M3.5	*	....
36	CC 658	11 43 .0	−64 33	2 .69	.206	15.8	11 wd	....	....
37	BD −15°6290	22 50 .6	−14 31	1 .17	.206	15.8	10.2 M5	....	....
38	Omicron² Eridani	4 13 .0	− 7 44	4 .08	.205	15.9	4.4 K0	9.9 wdA	11.2 M4e
39	BD +20°2465	10 16 .9	+20 07	0 .49	.202	16.1	9.4 M4 .5	*	....
40	Altair	19 48 .3	+ 8 44	0 .66	.196	16.6	0.8 A7	....	....
41	70 Ophiuchi	18 02 .9	+ 2 31	1 .13	.195	16.7	4.2 K1	6.0 K6	....
42	AC +79°3888	11 44 .6	+78 58	0 .87	.194	16.8	11.0 M4	....	....
43	BD +43°4305	22 44 .7	+44 05	0 .84	.193	16.9	10.1 M5e	*	....
44	Stein 2051	4 26 .8	+58 53	2 .37	0″.192	17.0	11.1 (M5)	12.4 wd	....

* Unseen components.
† The position of Alpha Centauri C ("Proxima") is $14^h26^m.3$, −62°28′; 2°11′ from the center of mass of Alpha Centauri A and B.

11.5 Limitations of the Direct Method

The direct, or trigonometric, method of determining stellar parallaxes diminishes in accuracy as more distant stars are observed. The probable error of the most reliable parallaxes, in which several independent measures are averaged, is of the order of 0".002. For the very nearest stars this error is less than 1 per cent of the parallax. The percentage of probable error increases as the parallax decreases; it is 10 per cent for a parallax of 0".02. If the observed parallax is as small as 0".01 $\pm$ 0".002, it follows from the definition of the probable error that the chance is one half that the true value lies between 0".012 and 0".008, or that the star's distance is between 83 and 125 parsecs. It is equally probable that the true value is outside these limits.

Thus the percentage of error in measuring the distances of stars by the direct method increases with the distance and becomes very large for distances exceeding 100 parsecs, or about 300 light years. Because the success of many investigations of the stars depends on the knowledge of their distances, astronomers have sought for and have discovered indirect ways of determining the parallaxes of more remote stars. These will be noted as we proceed, but we should always be aware that the indirect methods require calibration.

MOTIONS OF THE STARS

As a result of trying to determine parallaxes, Edmund Halley, in 1718, was the first to demonstrate that the stars are not "fixed." He showed that certain bright stars had moved from the places assigned them in Ptolemy's ancient catalog by about the moon's apparent diameter. The stars are moving in various directions relative to one another. Their movements through space are often swift, but seem very slow to us because of the great distances of the stars.

11.6 Two Projections of a Star's Motion

(1) *Proper motion* (μ) is the rate of change in the star's direction, or apparent place on the celestial sphere; this *angular* rate decreases, in general, as the star's distance is greater. (2) *Radial velocity* (r_v) is the star's speed of approach or recession; this *linear* rate is independent of the distance and is often the only projection of the motion that can be measured. The

observed motion may be referred to the sun by correcting for the effects of the earth's motions.

Proper motions, radial velocities, and distances of the stars constitute the basic data in the studies of stellar motions.

11.7 Proper Motions

The proper motions of stars are determined by comparing their right ascensions and declinations in catalogs of star positions that are separated by sufficient intervals of time. The rates of change in these coordinates are the required data, after allowance has been made for apparent displacements of the stars caused by precession and other motions of the earth. By this procedure the proper motions of all lucid stars have become known and of multitudes of telescopic stars as well.

However, the derived proper motions slowly deviate from those assigned. The reason for this is that the coordinate system in which they are determined is not an inertial system and slowly deteriorates due to galactic rotation. To overcome this, an effort is being made to set up a coordinate system using galaxies as the fundamental defining points. Galaxies are so far away that they have no sensible proper motions. One way of overcoming this difficulty is to use radio interferometers and radio point sources that are known or believed to be far distant. The method depends only upon a knowledge of the baseline, the rotation of the earth and an ability to identify the central fringe. Once this independent coordinate system is set up it can be transferred to the stars by optical techniques since the majority of the radio sources used now have optical identifications as well.

Many thousand stars with large proper motions have been detected and measured by comparison of star positions in two photographs of the same region of the heavens obtained at different times. These motions are called *relative proper motions.* The comparison is relatively simple and rapid because precession, aberration, and other apparent displacements of stars, with the exception of parallax, are nearly the same over a small area of the sky. Such proper motion surveys have been carried out since the application of the photographic plate to astronomy. Monumental and carefully planned surveys have been conducted by W. J. Luyten, first with small astrographic cameras and later using the Palomar Schmidt telescope. An interesting proper motion survey is being carried out by H. Giclas at Lowell Observatory by using the extensive Pluto search plates as his first epoch material. These plates reach a faint limiting magnitude, provide a forty-year interval and cover a very broad region centered on the ecliptic.

The method of "blinking" is effective for the detection of any differences between two photographs that are being compared. The two plates are arranged under a blink microscope so that corresponding star images

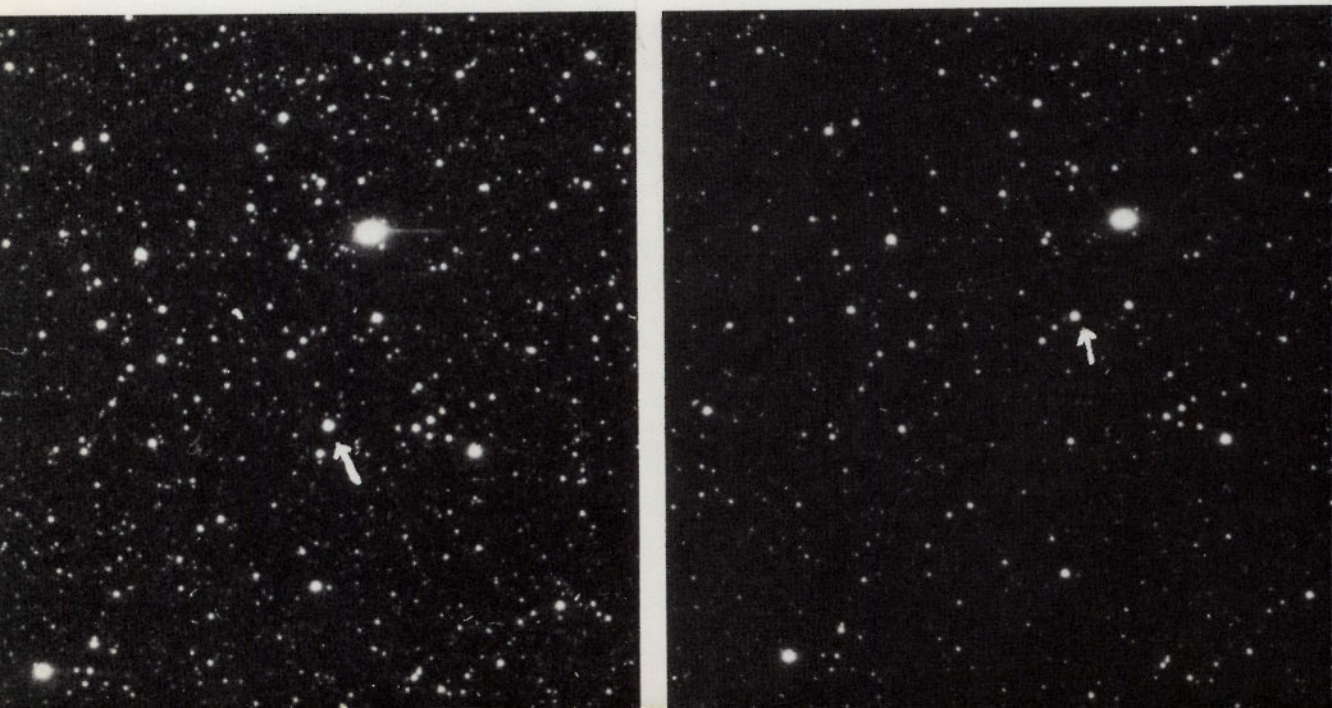
appear superposed. By mechanical means the plates are alternately hidden several times a second. If all the stars in the region have not moved appreciably between the exposure times, the appearance is the same as before the blinking began. If a star is displaced on one plate relative to the other, the result is a jumping effect, which at once attracts the observer's attention.

11.8 Stars with Large Proper Motions

The largest known proper motion is that of a telescopic star called "Barnard's star" after the name of the astronomer who first noticed its swift motion. This 10th-magnitude star in Ophiuchus is moving with respect to its neighbors at the rate of 10".3 a year, so that it advances in 175 years through an angle equal to the moon's apparent diameter. If all the stars were moving as fast as this and at random, the forms of the constellations would be altered appreciably in the course of a lifetime. As it is, the known proper motions of only about 330 stars exceed 1" a year, and the average for all the naked-eye stars is not greater than 0".1 a year.

We have noted that the proper motion is angular. A star having a large proper motion may actually be in rapid motion, or it may be nearer us than most stars, or both, as is true of Barnard's star. We have also noted that the proper motion relates only to that part of the star's motion that is transverse to the line of sight.

Figure 11.8
Barnard's star (marked with an arrow in both photographs) in June, 1937 (left) and in August, 1962 (right). Notice its motion against the background stars in the course of 25 years. Proper motion: 10".30 per year. *(Lowell Observatory Photograph)*



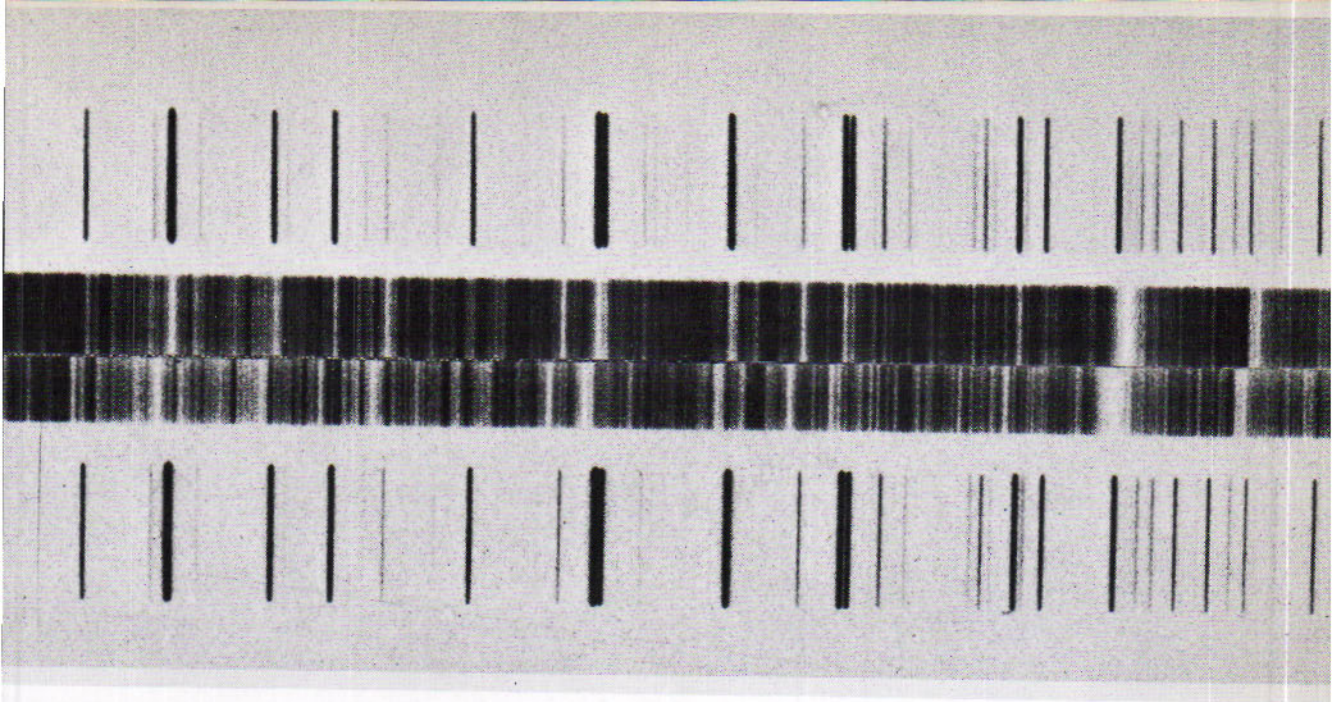


Figure 11.9A
Doppler displacements in stellar spectra, on stars HD 161096 (top) and HD 66141 (bottom). The different radial velocities are immediately noticeable. *(David Dunlap Observatory Photograph, University of Toronto)*

11.9 Radial Velocities

A star's rate of motion directly toward or away from us, or its *radial velocity*, is shown by the displacement of the lines of its spectrum from their normal positions. By the Doppler effect (4.9), the wavelengths are shortened if the star is approaching us and are lengthened if the star is receding; and the amount of the change in their lengths is directly proportional to the star's radial velocity. For the visual region of the spectrum, the lines are displaced respectively toward the violet or the red end. The more general terms *shortward* and *longward* denote without ambiguity the direction of the displacement of the lines in any region of the spectrum. If, for example, a line at 4000 angstroms is displaced 1 angstrom toward the violet, the star is approaching us with a speed of $\frac{1}{4000} \times 300{,}000$ km a second; its radial velocity is -74.9 km a second. Approach is indicated by the minus sign, since the distance is decreasing recession by the plus sign since distance is increasing.

$$v_r = \frac{\Delta\lambda}{\lambda} C$$

The spectroscope is attached at the eye end of the telescope, with the slit at the focus of the objective. With it the bright-line spectrum of a laboratory source, often iron or titanium, may also be photographed above and below the star's spectrum. The photograph, or *spectrogram* (Fig. 11.9A), is observed under a microscope, and the positions of the star-lines are measured micrometrically with respect to the *comparison lines* of the laboratory source, which have no Doppler displacements. The measurements are then reduced to the sun's position by correcting for the earth's orbital motion.

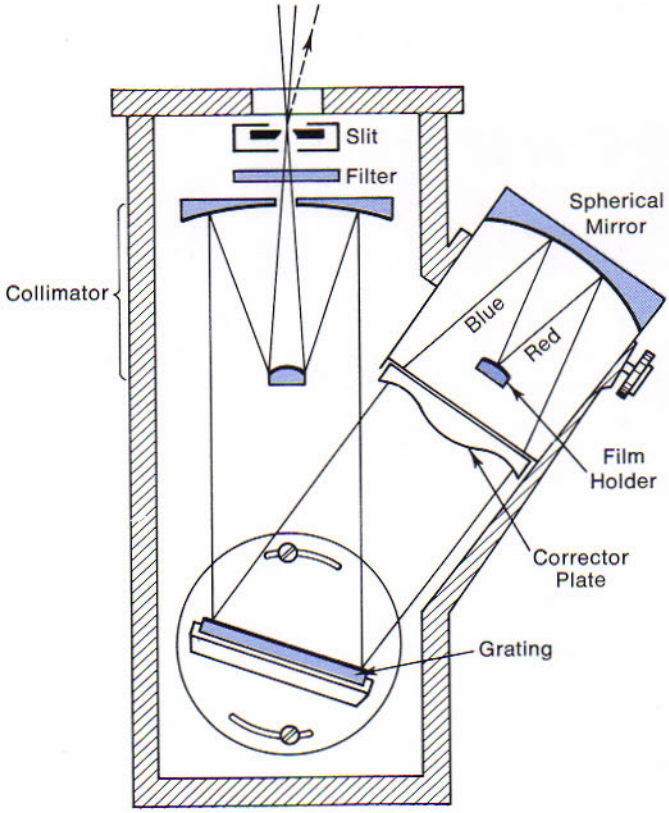


Figure 11.9B
Diagram of a simple grating spectrograph.

A catalog by R. E. Wilson lists the radial velocities of more than 15,000 stars. Velocities up to 30 km/sec, or about 20 miles a second, are common; those exceeding 100 km/sec are rare.

11.10 Annual Variation in the Radial Velocities

The earth's revolution around the sun causes a variation in the observed radial velocities of the stars during the year. When the earth is moving toward a star, the lines in the star's spectrum are displaced to shorter wavelengths; when the earth is moving away from the star, the lines are displaced to longer wavelengths than they would otherwise have. The observed radial velocities may be corrected for this and some other effects of the earth's motions.

The annual variation in the radial velocities provides another means (7.16) of determining the earth's distance from the sun. H. Spencer Jones derived in this way the value of 149,427,830 km, which is not far from the mean distance currently adopted.

As a simplified example, consider a star at the ecliptic and at rest. Once in a year the earth is moving directly toward the star. Six months

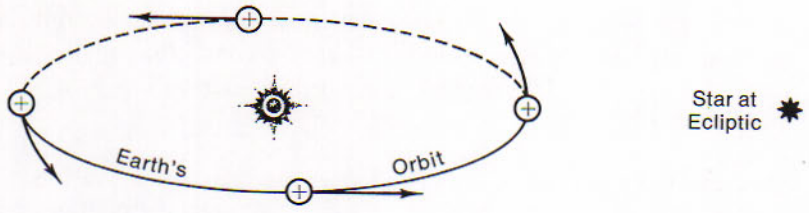


Figure 11.10
Annual Variation in the Radial Velocity of a Star. Because of the earth's revolution, the lines in the spectrum of a star oscillate in a period of a year. The effect is greatest for a star on the ecliptic.

later it is moving directly away from the star. On each occasion the radial velocity of the star as determined from the displacement of its spectrum lines is numerically equal to the speed of the earth's revolution. Suppose that this value is 29.6 km a second and that the earth's orbit is a circle. Multiplying 29.6 km per second by 31,558,150, about the number of seconds in the sidereal year, we find for the circumference of the orbit nearly 940 million km. Dividing this value by 2×3.1416 we have for the mean radius of the orbit about 149,500,000 km.

11.11 Space Velocities

When the annual proper motion, μ, of a star and its parallax, p, are known, the tangential velocity, v_T, can be calculated. The *tangential velocity* is the star's velocity in kilometers per second with respect to the sun at right angles to the line of sight.

$$v_T = \frac{4.74\mu}{p}$$

When the star's radial velocity, v_r, is known as well, the *space velocity*, v, which is the star's velocity with respect to the sun, can be derived. The space velocities of the stars in the sun's vicinity are generally of the same

$$v^2 = {v_r}^2 + {v_T}^2$$

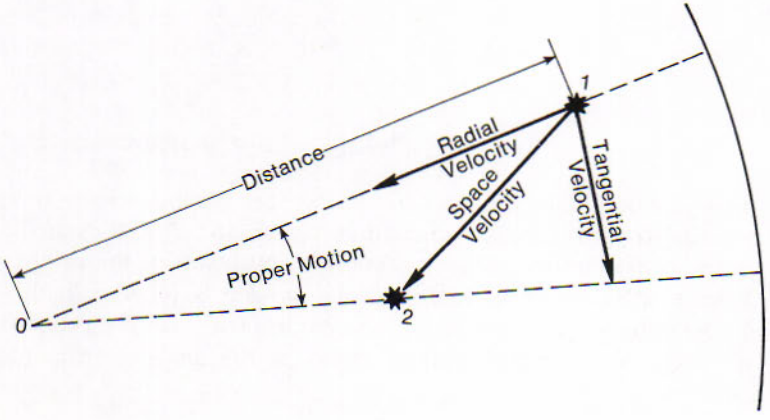


Figure 11.11
Relation between space velocity, tangential velocity, radial velocity, proper motion, and distance of a star.

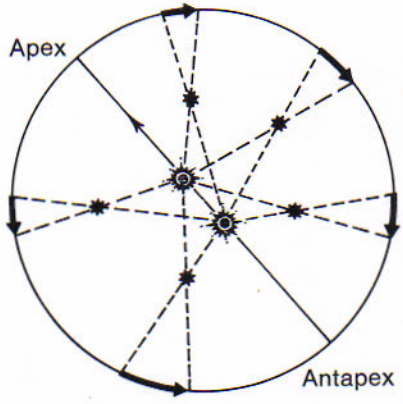
order as the velocities of the planets in their revolutions around the sun; the majority are between 8 and 30 km a second. Among the brightest stars, Arcturus has the highest space velocity; it is 135 km/sec, or 84 miles a second.

11.12 The Sun's Motion

Although the space velocities of the stars are referred to the sun, the sun itself is one of the stars and is in motion among them. It is therefore important to determine how the sun is moving and to be able to correct the space velocities for the effects of this motion.

If the sun with its planetary system is moving in a certain direction among the stars around it and if the stars have random motions, these stars should seem to be passing by in the opposite direction. The stars ahead of us should seem to be opening out from the *apex* of the sun's way, the point of the celestial sphere toward which the sun's motion is directed. The stars behind us should seem to be closing in toward the opposite *antapex*. So reasoned W. Herschel, the pioneer in the study of sidereal astronomy. Although the proper motions of only 13 stars were then available, he determined, in 1783, the position of the apex within 10° of the place now assigned to the "standard apex."

Figure 11.12
Apparent Motions of Stars Caused by the Solar Motion. The stars seem to be drifting away from the point on the celestial sphere toward which the sun is moving.



11.13 The Standard Apex of the Sun's Way

This is referred to the average of the stars visible to the naked eye. It is situated approximately in right ascension $18^h\,0^m$ and declination $+30°$, in the constellation Hercules about 10° southwest of the bright star Vega. With respect to these stars the solar system is moving in this direction at the rate of 20 km/sec. In the course of a year we progress through the local field of stars 4 times as far as the distance from the earth to

the sun. The corresponding antapex is in the constellation Columba, about 30° south of Orion's belt.

The more recent determinations are based on the proper motions and also the radial velocities of thousands of stars. When the radial velocities are employed, the apex is evidently the point around which the stars have the greatest average velocity of approach, while around the antapex they have the greatest average velocity of recession; and this average is the speed of the sun's motion relative to these particular stars. The positions from the proper motions and radial velocities nearly agree, as they should, because the reference stars are in both cases in the sun's neighborhood.

When the motions of more distant stars are studied, the apex of the sun's way is displaced progressively toward the northeast. Thus, from their analysis of the proper motions of 18,000 of the brighter telescopic stars, P. van de Kamp and A. N. Vyssotsky located the apex in right ascension $19^h.0$ and declination $+36°$, in the constellation Lyra. As we will comment on later, this is because the sun and the stars in its vicinity are moving swiftly toward Cygnus in the rotation of our galaxy.

It is very important to notice that the position of the apex differs depending upon our selection of objects, for example: early-type stars or late-type stars, planetary nebulae or interstellar hydrogen, etc. Apparently the dynamics of the Galaxy plays a determining role and by studying these differences we can learn something about it. When using radial velocities to determine the apex, we obtain the reflected solar motion as well. The solar motion differs for various types of stars and objects as we might expect: the largest difference from that given by stars in the solar neighborhood being for the late F stars.

11.14 Stellar Distances from the Solar Motion

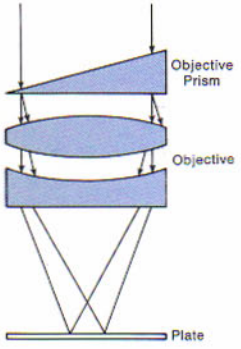
We have seen that the direct parallax method is limited to the nearer stars. For the more distant ones the diameter of the earth's orbit is too short to serve as an adequate base line. At first sight it might seem that the sun's motion could provide an ideal base for measuring stellar parallaxes. In only one year it takes us a distance twice as great as the diameter of the earth's orbit. If this longer base line is still too short to afford appreciable parallax of the distant stars, we could wait two years or perhaps a hundred, until it would become long enough. Such a parallax is called a *secular parallax.*

Because the stars have motions of their own, it is generally impossible to say what part of a star's proper motion is a parallax effect and what part is peculiar to the star itself. Thus the longer base line provided by the sun's motion cannot be used to determine the distances of individual stars, although it has given reliable average parallaxes of groups of stars. This is referred to as a *statistical parallax.*

STELLAR SPECTRA

The light from a star is the sum total of energy of many wavelengths. This light when spread out to display the spectrum conveys to us vital astrophysical information depending upon the presence or absence of certain wavelengths.

11.15 Photographs of Stellar Spectra



These are made in two different ways. One method employs a complete spectroscope with its slit at the focus of the telescope objective (Fig. 11.9A). Reflecting prisms over the ends of the slit bring in the light of a laboratory source on either side of the beam of starlight. The bright comparison lines of the source appear in the photograph adjacent to the star's spectrum; these can serve as standards for measuring wavelengths and Doppler shifts in the star's spectrum. This method gives the spectrum of only one star at a time and is rather wasteful of light. The present limit for the 200-inch telescope with a 1-night exposure is a low-dispersion spectrum of an 18th-magnitude star. This improves to about 20th magnitude using electronic image intensifiers.

The objective prism method is preferred when the spectra of many stars are to be examined, as in the classification of stellar spectra. A large prism of small angle is placed in front of the telescope objective,

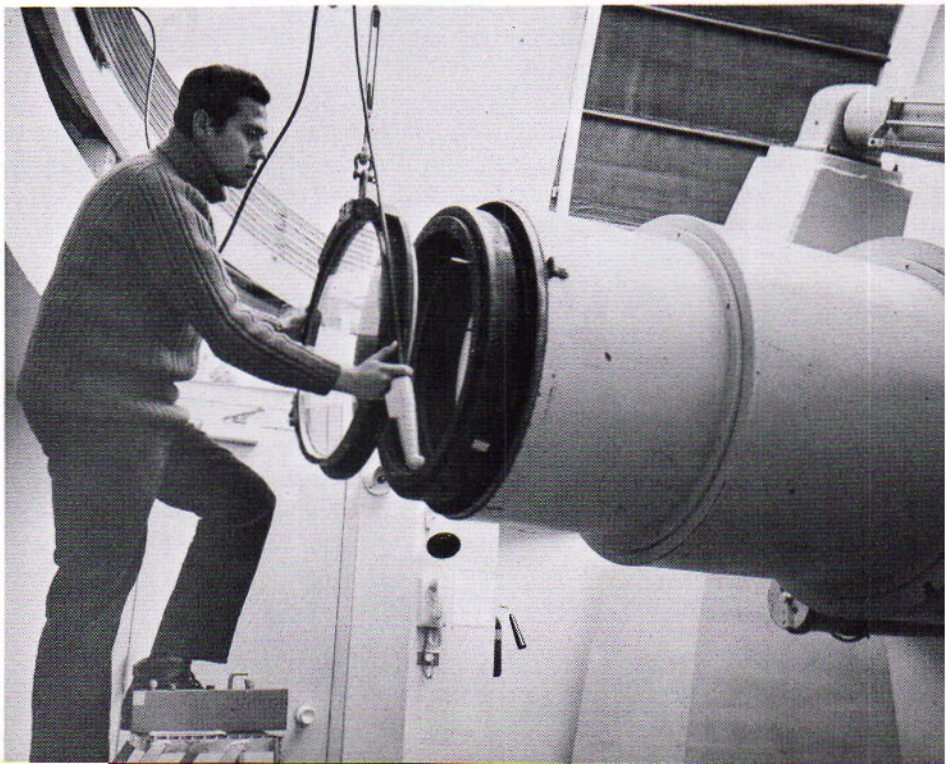


Figure 11.15 Objective prism for the Michigan Schmidt telescope at the Cerro Tololo Interamerican Observatory in Chile. *(Photograph courtesy of the Cerro Tololo Interamerican Observatory)*

so that the whole apparatus becomes a spectroscope without slit or collimator. The photograph shows appropriately widened spectra of all stars of sufficient brightness in the field of view. Exposure times are restricted by the light of the sky, as in direct photography. The method is now limited to stars brighter than photographic magnitude 12 or 13.

(See Plate I)

11.16 The Classification of Stellar Spectra

The photographic study of stellar spectra with the objective prism was inaugurated, in 1885, by E. Pickering at Harvard Observatory. This work was carried on for many years under the immediate direction of Annie J. Cannon. *The Henry Draper Catalogue,* in 9 volumes completed in 1924, gives the approximate positions, magnitudes, and spectral types of 225,300 stars in all parts of the heavens. Its extensions, particularly to fainter stars in the Milky Way, contain almost as many more stars.

One of the outstanding results of the program was the discovery that the great majority of stellar spectra can be arranged in a single continuous sequence. This gradation is the basis of the *Draper Classification,* for which Miss Cannon was chiefly responsible. Various stages in the sequence are denoted by the principal types O, B, A, F, G, K, and M, which are subdivided on the decimal system. Thus a G5 star is halfway between G0 and K0; B2 is nearer to B0 than to A0. Fully 99 per cent of the stars are included in the types from B to M.

Four additional types complete the Harvard sequence. These are W now paired with O at the blue end, and R-N and S, which form two side branches near the red end; the first branches off between G and K, and the second between K and M. Generally it is a sequence of diminishing surface temperature and increasing redness of the stars. We note here the main features of the changing patterns along the sequence in the visible region of the spectrum, leaving further description and explanation for the following chapter, especially the refinements that identify various luminosity classes.

11.17 The Sequence of Stellar Spectra

The Harvard classification was based on gradations in the patterns of the spectral lines and was independent of theoretical considerations. The sequence is characterized particularly by the rise and decline in the strength of the hydrogen lines throughout its extent. Lines of other elements become prominent at different stages of the sequence, and bands of chemical compounds appear toward the end of the sequence. The main features in the visual spectra of the different types are as follows:

Type O. Lines of ionized helium, oxygen, and nitrogen appear along with those of hydrogen. *Type W* contains the bright-line Wolf-Rayet stars.

Type B. Lines of neutral helium are most intense at B2 and then fade,

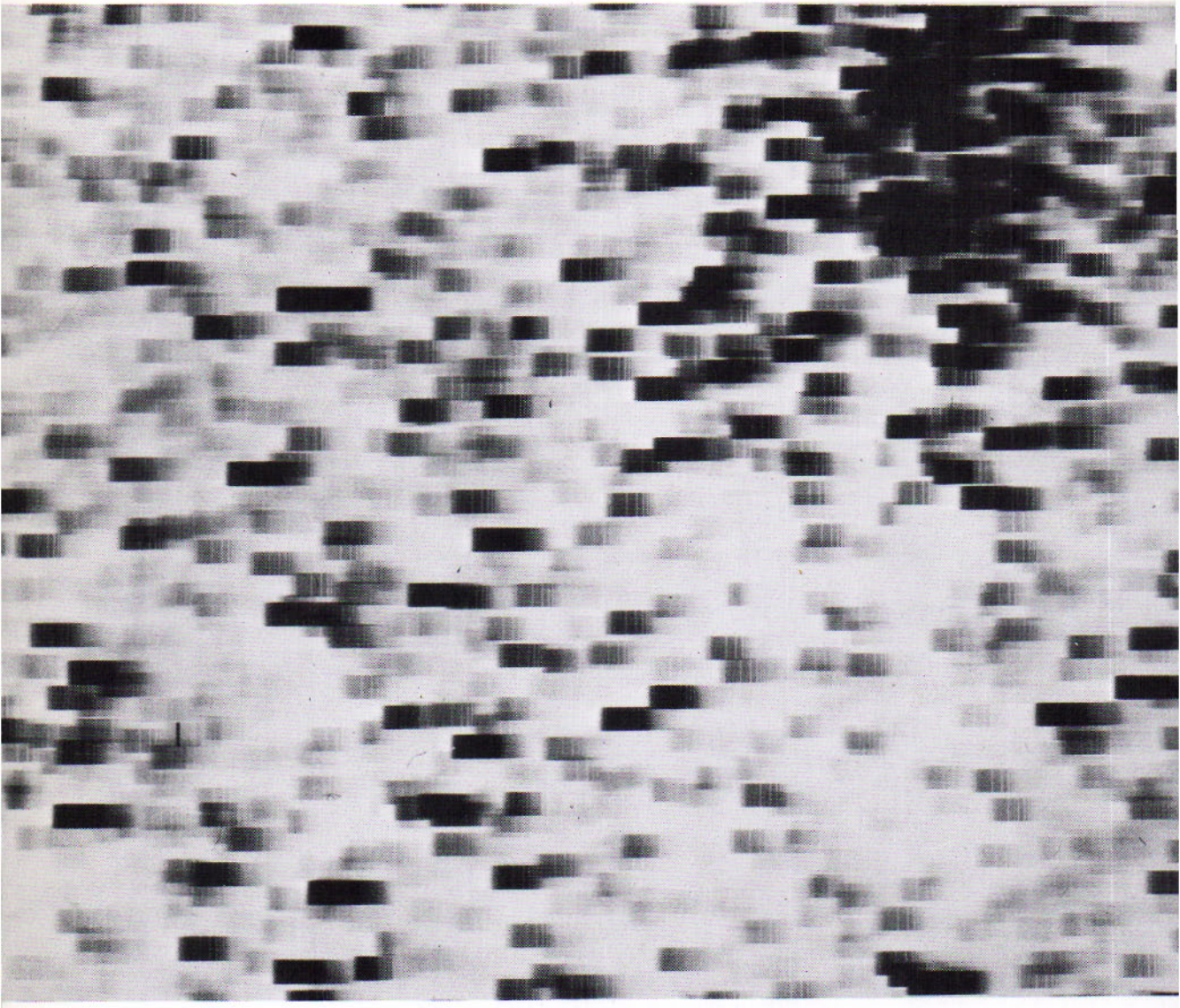


Figure 11.17
Objective prism plate of region NE of NGC 3532. Note bright emission line star on lower left. *(Photograph by A. G. Davis Philip using Michigan Curtis Schmidt telescope with Warner and Swasey ultraviolet prism at Cerro Tololo)*

until at B9 they have practically disappeared. Hydrogen lines increase in intensity throughout the subdivisions. Examples are Spica and Rigel.

Type A. Hydrogen lines attain maximum intensity at A2. Examples are Sirius and Vega. B and A stars are blue.

Type F. Hydrogen lines are declining. Lines of metals are increasing in intensity, notably the Fraunhofer H and K of ionized calcium. Canopus and Procyon are examples.

Type G. Lines of metals are prominent. These stars are yellow. The sun (type G2) and Capella are examples.

Type K. Lines of metals now surpass the hydrogen lines in strength. Bands of cyanogen and other molecules are becoming prominent. These stars are reddish. Examples are Arcturus and Aldebaran.

Type M. Bands of titanium oxide become increasing prominent up to their maximum at M7. The cooler members are classified by J. J. Nassau and associates to M10 by the strengthening of vanadium oxide bands. Betelgeuse and Antares are examples of this red type.

Types R and *N* show bands of carbon and carbon compounds, *type S* of zirconium oxide and lanthanum oxide.

MAGNITUDES OF THE STARS

A star's magnitude is its brightness according to certain accepted standards. These standards, sometimes confusing, have given way to increasingly more rigorous definitions as our ability to differentiate subtle brightness differences improves and as astronomers improve their ability to measure brightnesses in more and more rigorously defined wavelength bands.

11.18 The Scale of Magnitudes

The grading of the naked-eye stars in early times into 6 magnitudes (1.25) was intended primarily to assist in identifying the stars. There is no evidence that the choice of 6 groups, rather than some other number, was governed by any definite idea of fixed numerical relations between the groups. For many centuries afterward, the magnitudes of the stars were accepted as they appeared in Ptolemy's catalog. It was not until the comparatively recent times of the Herschels that stellar magnitudes began to enter as important factors into astronomical investigations, as in statistical studies of the structure of the stellar system.

About 1830, John Herschel concluded that a geometrical progression in the apparent brightness of the stars is associated with the arithmetical progression of their magnitudes. The problem was then to ascertain the constant ratio of brightness corresponding to a difference of 1 magnitude that would best represent the magnitudes already assigned to the lucid stars. Pogson, at Radcliffe in 1856, proposed the adoption of the ratio having the logarithm 0.4, a convenient value nearly equal to the average ratio derived from his observations and those of other astronomers. He adjusted the zero of this fixed scale so as to secure as good agreement as possible with the early catalog at the 6th magnitude. This is essentially the present scale.

Pogson's rule is a special case of a general psycho-physical relation established, in 1834, by the physiologist Weber and given a more precise phrasing later by Fechner. By Fechner's law, $R = c \log S$ where R is the intensity of a sensation, S is the stimulus producing it, and c is a constant factor of proportionality. Pogson had evaluated the constant in the corresponding relation: $m - n = c \log (l_n/l_m)$, where l_m and l_n are the apparent brightnesses of two stars having the magnitudes m and n, respectively. The constant is 2.5, or 1/0.4, or 1/log 2.512. If the difference, $m - n$, is 1 magnitude, $l_n/l_m = 2.512$.

11.19 Relation Between Brightness and Magnitude

We have seen that the ratio of brightness between two stars differing by exactly 1 magnitude is the number having the logarithm 0.4, which is about 2.512. A few values where the magnitude difference is a whole number are as follows:

Magnitude Difference	*Ratio of Brightness*
1.0 magnitude	2.512
2.0 magnitudes	6.31
3.0 magnitudes	15.85
4.0 magnitudes	39.8
5.0 magnitudes	100.0

In general, the ratio of apparent brightness, l_n/l_m, of two stars or other sources of light of magnitudes m and n can be derived by the formula:

$$\log (l_n/l_m) = 0.4(m - n).$$

It is to be noted that the number expressing the magnitude diminishes algebraically as the brightness increases, and that the choice of the zero point makes the magnitudes of the very brightest celestial objects negative. The apparent visual magnitude of the brightest star, Sirius, is −1.4, of the planet Venus at greatest brilliancy is −4.4, of the full moon is

−12.6, and of the sun is −26.8. The photographic magnitude of the faintest star recorded with the 200-inch telescope is +23.9.

The following examples illustrate some of the uses of the relation between apparent brightness and magnitude:

1. How much brighter is Sirius (magnitude −1.4) than a star of the magnitude +23.9?

Answer: $\log (l_n/l_m) = 0.4 \times 25.3 = 10.12$

l_n/l_m = about 13,200 million times.

2. Nova Aquilae, in the course of two or three days in June, 1918, increased in brightness about 45,000 times. How many magnitudes did it rise?

Answer: $\log (l_n/l_m) = \log 45{,}000 = 4.65 = 0.4(m - n)$

$m - n = 4.65/0.4$ = about 11.6 magnitudes.

3. The bright star Castor, which appears single to the naked eye, is resolved by the telescope into two stars of magnitudes 1.99 and 2.85. What is the magnitude of the two combined?

Answer: $\log (l_n/l_m) = 0.4 \times 0.86 = 0.344$

$l_n/l_m = 2.21$; $(l_m + l_n)/l_m = l_x/l_m = 3.21$

$\log (l_x/l_m) = 0.507 = 0.4(m - x)$

$m - x = 0.507/0.4 = 1.27$

$x = 2.85 - 1.27 = 1.58$, the combined magnitude.

11.20 Visual and Photographic Magnitudes

The *apparent magnitude* of a star refers to its observed brightness, depending on its actual brightness and its distance from us. The value for a particular star varies with the region of the spectrum to which the receiver of the star's radiation is sensitive. If the receiver is the eye alone or the eye at the telescope, it is also the visual magnitude that is determined.

Visual methods for determining magnitudes are now generally replaced by the more reliable photographic and photoelectric methods. By use of a suitable color filter and a specially stained plate, or still more accurately with a filter and photocell, it is possible to determine the magnitude on the visual scale. This *photovisual magnitude* (m_v) is usually meant when we refer to the visual magnitude.

The *photographic magnitude* (m_{pg}) of a star is its magnitude as shown in a blue-sensitive photograph. It is generally determined by the size of the star's round image, which is larger and also denser as the star is brighter, and by comparison with stars of known magnitudes in the same photograph. The desired magnitude is sometimes simply estimated by viewing the plate with an eyepiece. For greater accuracy the star's image is measured with special apparatus. A single measure on a photographic plate with an iris photometer can determine the brightness of a star with a probable error of about 0.03 magnitude.

Figure 11.20
Star field in Perseus (15.9). Note how star images are larger for brighter stars of the same spectral type. *(Lick Observatory Photograph by L. Herbig)*

$$\text{C.I.} = M_{pg} - M_v = m_{pg} - m_v$$

The difference between the photographic and photovisual magnitudes is defined as the *color index* (C.I.) of a star. The color index is further defined as having a value of zero for an AO star. From the radiation curve it is clear that stars having spectral types earlier than AO will have color indexes that are negative.

11.21 Photoelectric Photometry

The superior accuracy (0.01 magnitudes or better) of the photoelectric technique compared to visual and photographic techniques has led to its superseding these earlier methods. An important advantage of photoelectric photometry is that the photocathode responds directly proportionally to the intensity of light falling upon it. A disadvantage is that only one star can be measured at a time. It is also difficult to allow for the sky light in congested areas, as in the Milky Way. Where many stars are to be observed, the procedure is generally to set up magnitude standards in the area with the photomultiplier and then go on with photography.

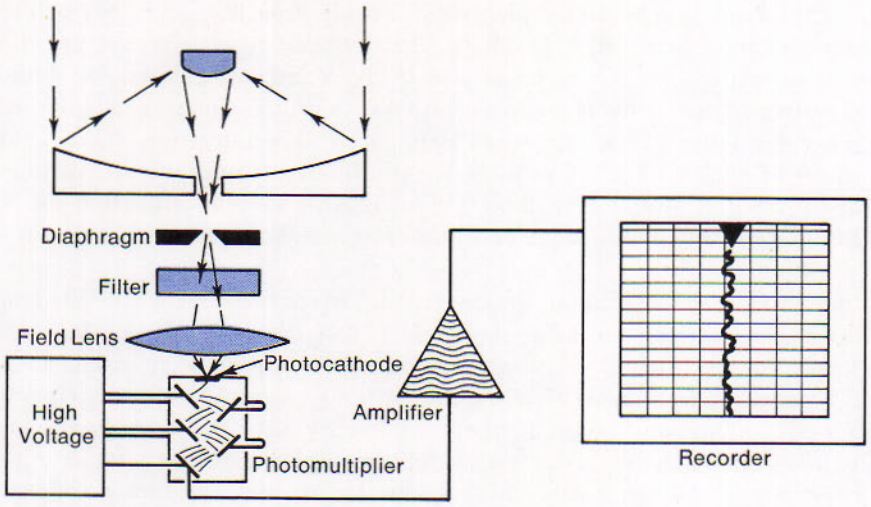


Figure 11.21
Schematic Drawing of a Photoelectric Photometer.

TABLE 11.II
The Brightest Stars

	NAME	SPECTRUM	APPARENT MAGNITUDE V	COLOR B–V	PARALLAX	ABSOLUTE VISUAL MAGNITUDE
1	Sirius	A1 V	−1.43	+0.00	0″.377	+1.5
2	Canopus	F0 Ia	−0.73	+0.15	.018	−4.4
3	Alpha Centauri, d	G2 V	−0.27*	+0.66	.760	+4.1
4	Arcturus	K2 IIIp	−0.06	+1.23	.090	−0.3
5	Vega	A0 V	+0.04	+0.00	.123	+0.5
6	Capella	(G0)	+0.09	+0.80	.073	−0.6
7	Rigel	B8 Ia	+0.15	−0.04	.005	−6.4
8	Procyon	F5 IV-V	+0.37	+0.41	.287	+2.7
9	Achernar	B3 V	+0.53	−0.16	.023	−2.7
10	Beta Centauri, d	B0.5 V	+0.66	−0.21	.016	−3.3
11	Betelgeuse, v	M2 Iab	+0.7	+1.87	.017	−2.9
12	Altair	A7 IV, V	+0.80	+0.22	.196	+2.3
13	Aldebaran, v	K5 III	+0.85	+1.52	.048	−0.7
14	Alpha Crucis, d	B0.5 V	+0.87*	−0.24	.015	−3.2
15	Antares, v, d	M1 Ib	+0.98*	+1.80	.019	−2.6
16	Spica	B1 V	+1.00	−0.23	.021	−2.4
17	Fomalhaut	A3 V	+1.16	+0.09	.144	+2.0
18	Pollux	K0 III	+1.16	+1.01	.093	+1.0
19	Deneb	A2 Ia	+1.26	+0.09	.006	−4.8
20	Beta Crucis	B0.5 IV	+1.31	−0.23	(.011)	−3.5
21	Regulus	B7 V	+1.36	−0.11	.039	−0.7
22	Adhara	B2 II	+1.49	−0.17	(.012)	−3.1
23	Castor, d	(A0)	+1.59*	+0.05	.072	+1.0
24	Shaula	B2 IV	+1.62	−0.23	(.026)	−1.3
25	Bellatrix	B2 III	+1.64	−0.23	(.026)	−1.3

d indicates a double star with a magnitude difference less than 5; combined magnitudes are given.
v indicates variable star.
( ) indicates estimated values.

The *photoelectric photometer* places various filters in the light path before the photomultiplier tube. Standard filter systems have been set up, the one in most general use being the U, B, V System (ultraviolet, blue, visual) introduced by H. Johnson and W. Morgan. Colors analogous to the color index (11.20) are then defined by the differences (U–B) and (B–V) in magnitudes. Other more highly differentiating systems, such as the u, v, b, y (ultraviolet, violet, blue, yellow) system advocated by B. Strömgren, are coming more and more into use.

11.22 The Brightest Stars

The 25 stars brighter than apparent visual magnitude +1.5 are listed in order of brightness in Table 11.II which first appeared in Sky & Telescope, August, 1957. The magnitudes, as given by H. L. Johnson, have been measured photoelectrically. The spectral types and luminosity classes are on the Morgan-Keenan system (12.20); they refer to the brighter components where the stars are visual doubles. Types from B to M are represented. The bluest stars, having the largest negative color indexes (11.24), are Alpha and Beta Crucis and Spica; the reddest are Betelgeuse and Antares.

These 25 stars are generally not the brightest in the sky because they are the nearest to us, but because they are intrinsically bright. Although Alpha Centauri, Sirius, and Procyon are among the nearest stars, others such as Rigel, Canopus, and Deneb are so remote that they must be highly luminous to appear so bright. The significance of the last column of the table is explained later in this chapter.

All but six of the brightest stars are visible at some times in the year throughout the United States. Those six become visible in their seasons south of about the following north latitudes: Canopus, 38°; Achernar, 33°; Alpha and Beta Centauri and Beta Crucis, 30°; Alpha Crucis, 28°.

11.23 Magnitude Standards

In many investigations that involve the apparent magnitudes of stars, the magnitudes must all conform to the same scale. Standard magnitude sequences have been set up as a means of control for all observers. They are sequences of stars in limited areas, having their magnitudes well determined and grading in brightness by small steps. The Mount Wilson north polar sequence of F. H. Seares long served as primary standards for photovisual and photographic magnitudes.

The greater accuracy of photoelectric photometry is employed in the newer sequences. The Mount Wilson and Palomar sequences in certain areas are based on magnitudes of selected stars of the original north polar sequence and are extended to the fainter magnitudes observed with the 200-inch telescope.

A system of standard magnitudes determined photoelectrically with

specified color filters has been established for the U, B, V filter system (11.21) by H. L. Johnson and W. W. Morgan for stars in various parts of the sky; for the red (*R*) at 7000 angstroms by R. Hardie, and infrared (*I*) at 8250 Å by G. E. Kron.

11.24 Star Colors

The various spectral classes discussed in sections 11.16 and 11.17 are differentiated by their colors. One color already mentioned (11.21) is B–V and another U–B. Both colors are adjusted at zero value for a main sequence star of spectral type AO, such as Sirius. Due to the definition of the magnitude scale it is clear that these colors increase as a star is redder, because a red star appears brighter in yellow light and therefore has a numerically smaller magnitude than in blue light. Thus, the reddest stars have a large positive value for their color and the blue stars have a slightly negative value for their colors.

Negative color values approach a certain limiting value. From Planck's radiation formula and the total radiating disk we can derive a relationship between the absolute magnitude at a given wavelength (M_λ), and the radius (R) and the absolute temperature (T) that involves certain constants (C_λ, χ).

$$M_\lambda = C_\lambda - 5 \log R + \frac{1.560}{\lambda T} + \chi$$

This formula can be evaluated at the visual and photographic wavelengths to give formulae for M_{pg} and M_v and the various constants can be evaluated in terms of the magnitude, radius and temperature of the sun. This gives us a very simple relationship for the Color Index (C.I.)

$$C.I. = M_{pg} - M_v = \frac{7200}{T} - 0.64$$

and we see that even for an infinitely hot source the color index will not be more negative than −0.64. A slightly different numerical result occurs for the color B–V. It is important to note that since we are using the color in the above formulae it is not essential to know the absolute magnitudes.

A *color index* or other similar measurement is a numerical measure of the star's color and therefore of its spectral type if nothing intervenes in space to redden the light. If the spectral type is also known and if the measured color index is greater than would normally be expected for a star of this type, the excess reveals the presence and effect of an intervening cosmic dusty medium, as we note in a later chapter. Values of *B−V* for the different spectral types are given in Table 12.II and may be found for the brightest stars from Table 11.II.

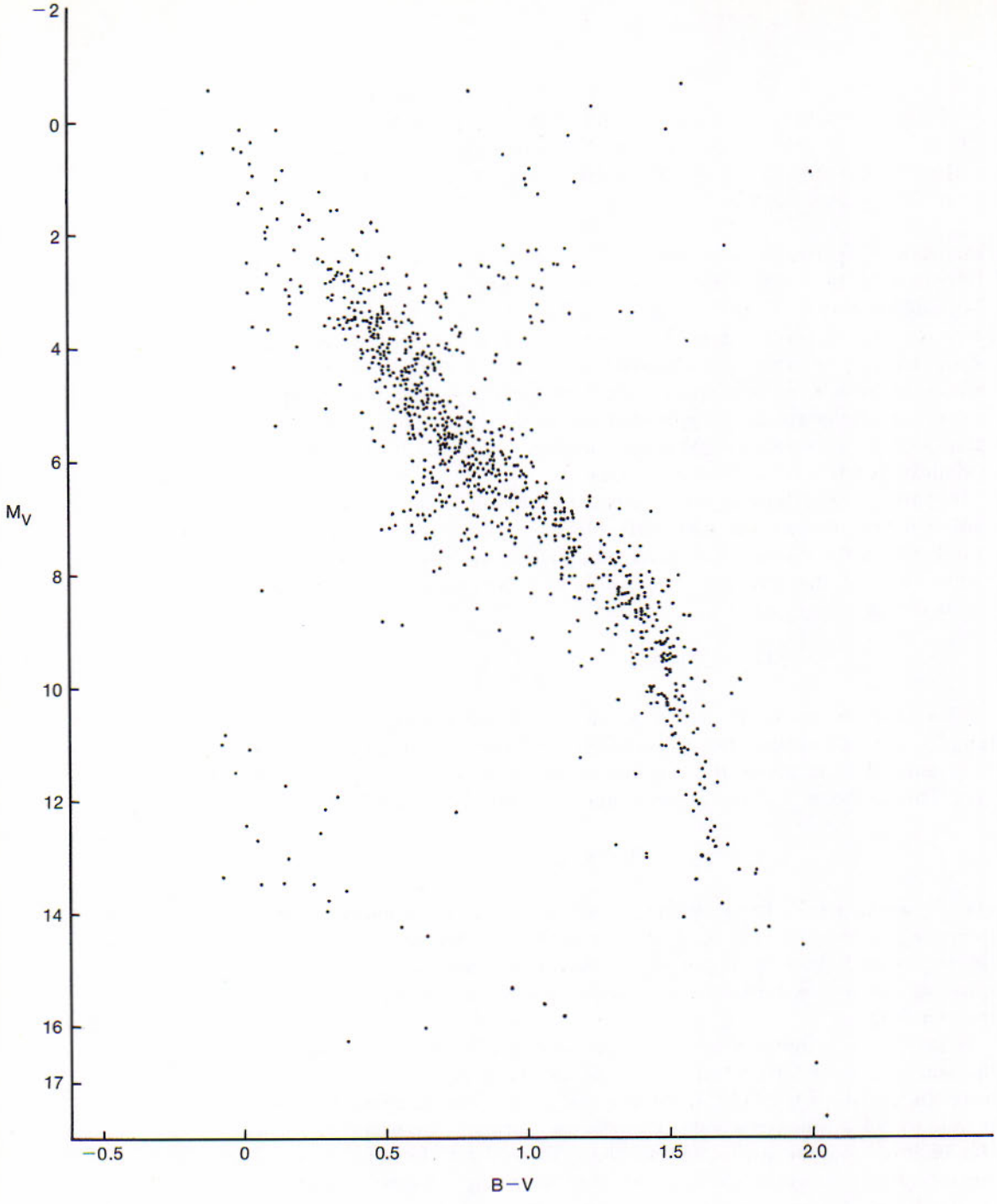


Figure 11.24A
Magnitude-color plot of the nearest stars. Note white dwarfs in the lower left portion of the diagram.

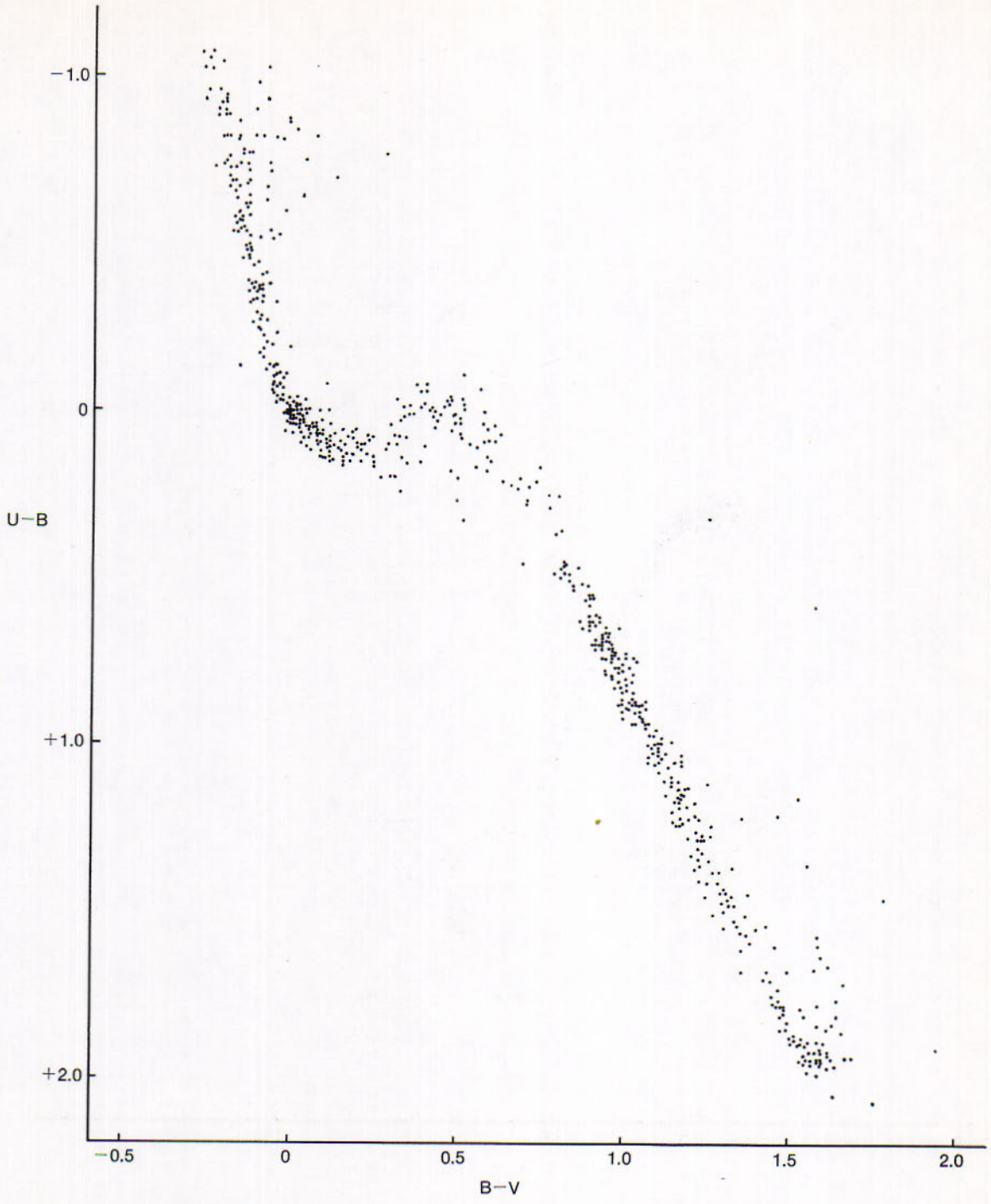


Figure 11.24B
Color-color plot of 708 bright stars. The main sequence is clearly defined.

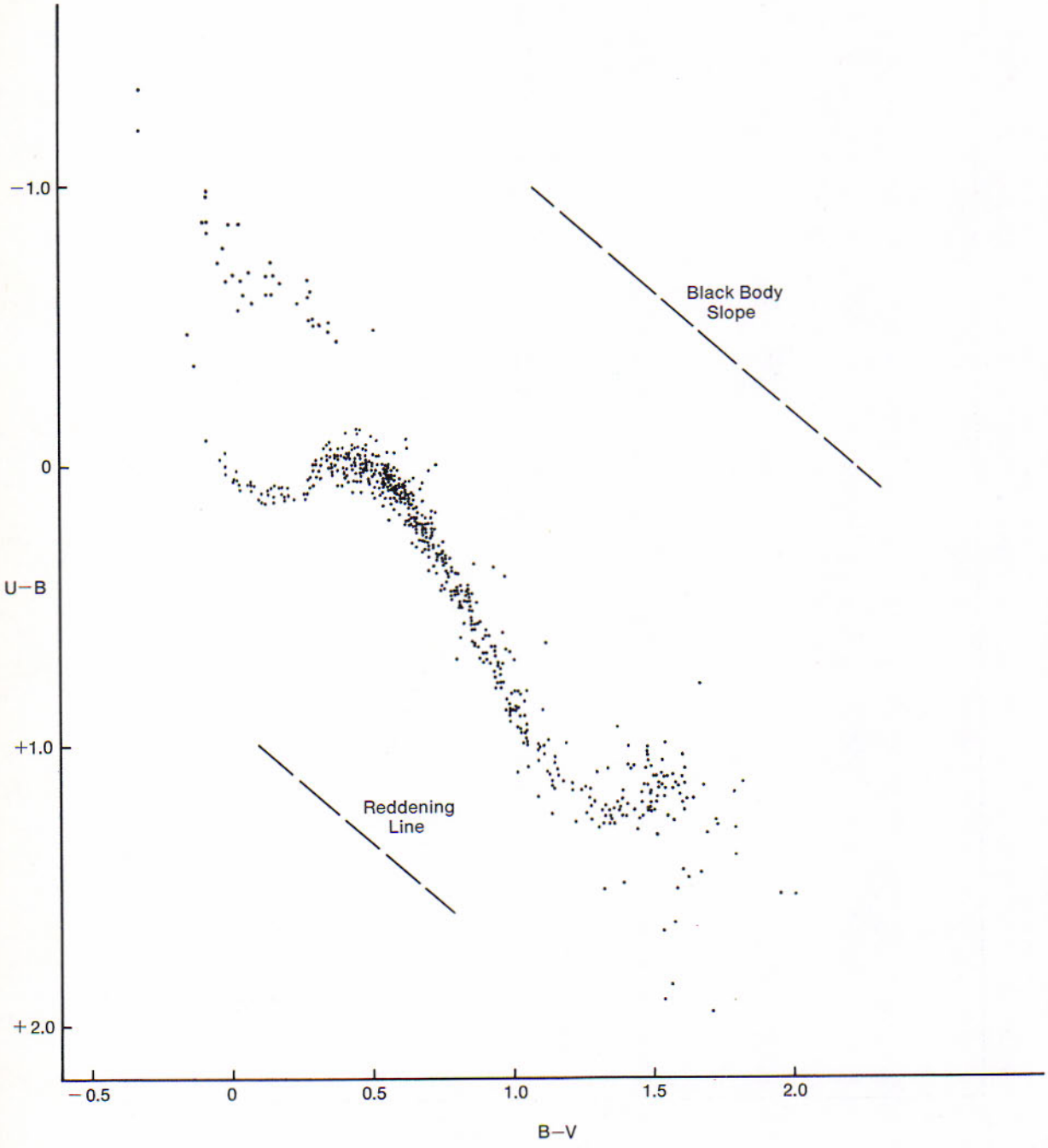


Figure 11.24C
Color-color plot of 662 of the nearest stars.

LUMINOSITIES OF STARS

The brightness of a star is one of its basic observable parameters. Stars of the same temperature but different luminosities may be at different distances or may have different surface areas or both. A star's luminosity is expressed in terms of the solar luminosity.

11.25 Absolute Magnitudes

The apparent magnitude of a star relates to its brightness as we observe it. This depends on the star's *luminosity*, or brightness at a specified distance, and on its actual distance. One star may appear brighter than another only because it is the nearer; thus the sun appears brighter than Capella. In order to rank the stars fairly with respect to luminosity, it is necessary to place them all at the same distance from us or, what amounts to the same thing, to calculate how bright they would appear if they were placed at the same distance. By agreement the standard distance is 10 parsecs, or 32.6 light years.

The *absolute magnitude* of a star is the apparent magnitude it would have at the distance of 10 parsecs (parallax 0".1). We should note quite explicitly that this is a *conventional* definition of absolute magnitude since any other distance would do just as well. The *absolute magnitude scale* is further defined by saying that a main sequence A0 star shall have an absolute magnitude of 0.0.

When the parallax, p", is known and the apparent magnitude, m, has been determined by observation, the absolute magnitude, M, can be calculated by the formula shown on the margin, where r is the distance in parsecs. This important formula is derived from the relation between brightness and magnitude (11.19) and the fact that the brightness of a point source of light varies inversely as the square of its distance. The absolute magnitude is of the same sort as the apparent magnitude employed in its calculation; it may be visual, photographic, or some other kind.

$$M = m + 5 + 5 \log p, \text{ or}$$
$$M = m + 5 - 5 \log r,$$

The type being referred to is written as M_v for absolute visual magnitude, M_{pg} for absolute photographic magnitude, etc. If a certain filter system is being referred to, e.g. the U, B, V system, we would write M_U, M_B and M_V.

When the absolute magnitudes, M_1 and M_2, of any two stars are known, the ratio of their luminosities, L_2 and L_1, is given by the formula on the margin, which is the same relation already given for the apparent magnitudes.

$$\log(L_2/L_1) = 0.4(M_1 - M_2),$$

11.26 Relative Luminosities

It is the custom to express the luminosity of a star in terms of the sun's luminosity, that is, as the number of times the star would outshine the sun if both were the same distance from us. This ratio can be calculated by substituting the absolute magnitudes of the sun and star in the preceding formula.

The sun's apparent visual magnitude, m, is -26.8; its parallax, p, on the same basis as those of the stars, is the radian, 206,265″, of which the logarithm is 5.314. By the first formula of the preceding section, the sun's absolute visual magnitude, M, is $-26.8 + 5 + 26.6$, or $+4.8$. At the standard distance of 10 parsecs the sun would appear as a star of nearly the 5th magnitude, only faintly visible to the naked eye.

The expression for the star's visual luminosity relative to that of the sun is accordingly:

$$\log (\text{luminosity}) = 0.4(4.8 - \text{star's absolute magnitude}).$$

Because the majority of the brightest stars are more remote than 10 parsecs, they must be more luminous than the sun. Indeed, this is true of all stars of Table 11.II, as is shown by their absolute magnitudes. Rigel and Deneb are of the order of 10,000 times as luminous as the sun. On the other hand, we have noted in Table 11.I that many stars nearer us than the standard distance of 10 parsecs are visible only with the telescope, so that they must be considerably less luminous than the sun. The conclusion is that the stars differ very greatly in luminosity. The foregoing formula is employed in the following examples.

(1) Compare the visual luminosities of Sirius and the sun. The absolute visual magnitude of Sirius is $+1.5$.

Answer: $\log L = 0.4(4.8 - 1.5) = 1.3$. Thus Sirius is 23 times as luminous as the sun.

(2) Compare the luminosities of Barnard's star (absolute visual magnitude $+13.2$) and the sun.

Answer: $\log L = 0.4(4.8 - 13.2) = -3.4 = 6.6 - 10$. Thus Barnard's star is 0.0004 as luminous as the sun.

11.27 Stars of the Main Sequence

When the absolute magnitudes of stars in our neighborhood are plotted against their spectral classes, as in Fig. 11.27, the majority of the points are arrayed in a band running diagonally across the diagram, referred to as a *spectrum-luminosity diagram*. The middle line of this band drops rather steadily along the spectral sequence from absolute magnitude -3 for B stars to fainter than $+10$ for M stars. This band is known as the *main sequence*.

The sun, a yellow star of type G2 and absolute visual magnitude $+4.8$, is a main-sequence star. It is about 100 times less luminous than the

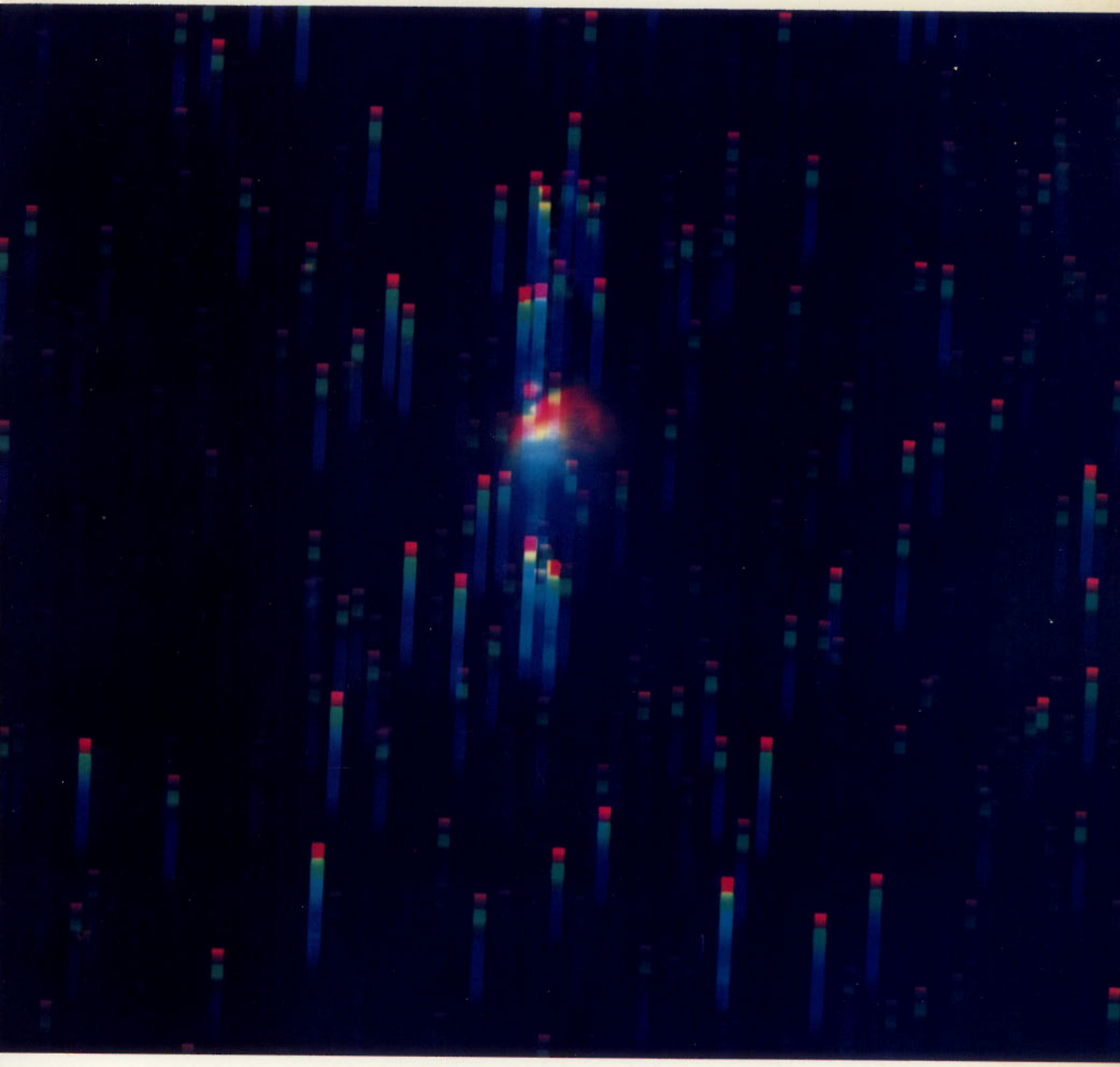


Plate I

Objective Prism Spectra. The field is centered on the Great Orion Nebula. The dispersion is north-south so the spectra can be widened by driving the telescope slightly slower than the sidereal rate. *(Department of Astronomy, University of Michigan)*

Plate II
Planet Mars. The orange hue is apparent and distinguishes Mars among the "naked-eye" planets. *(Photograph from the Hale Observatories)*

Plate III
Jupiter. Note one satellite on the right, and its shadow on the planet, slightly below and to the right of the famous red spot. *(Lick Observatory photograph)*

Plate IV
Saturn, the ringed planet. *(Lick Observatory photograph)*

Plate VI
The Dumbell, NGC 6853, an unusual planetary nebula. *(Lick Observatory photograph)*

Plate V
The Pleiades, and nebulosity in Taurus, NGC 1432. *(Photograph from the Hale Observatories)*

Plate VII
The Ring nebula in Lyra, NGC 6720. *(Photograph from the Hale Observatories)*

Plate VIII
Planetary nebula NGC 6543. *(Lick Observatory photograph)*

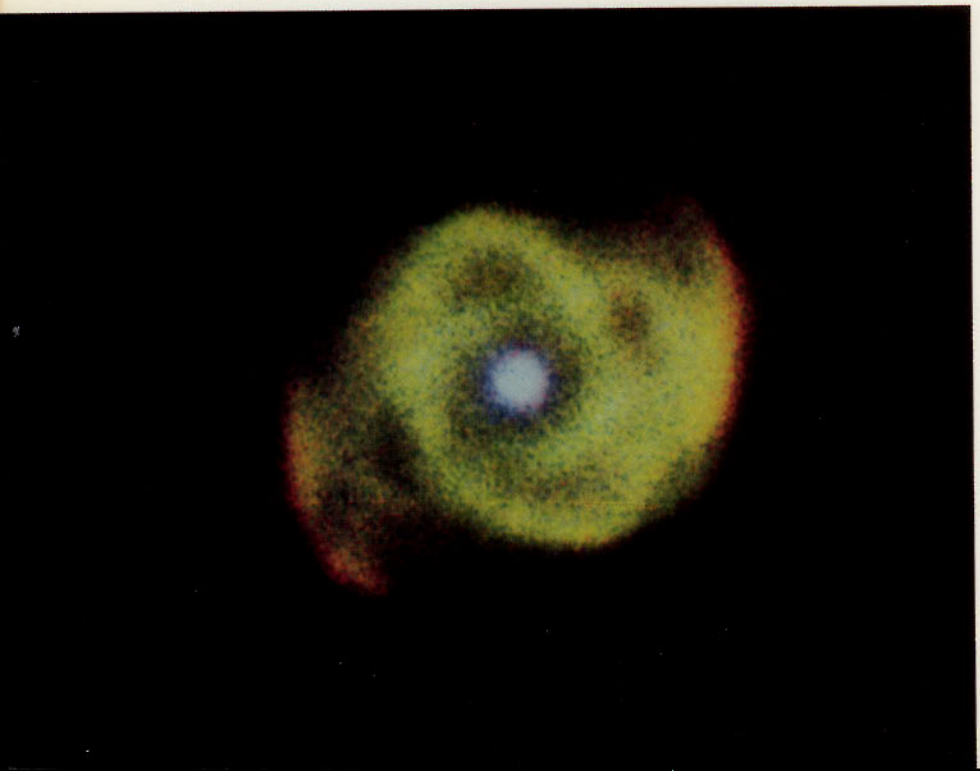


Plate IX

Crab Nebula in Taurus, NGC 1952. Note red filaments of hydrogen and blue haze of synchrotron emission *(Lick Observatory photograph)*

Plate X

Planetary nebula in Aquarius; NGC 7293. *(Photograph from the Hale Observatories)*

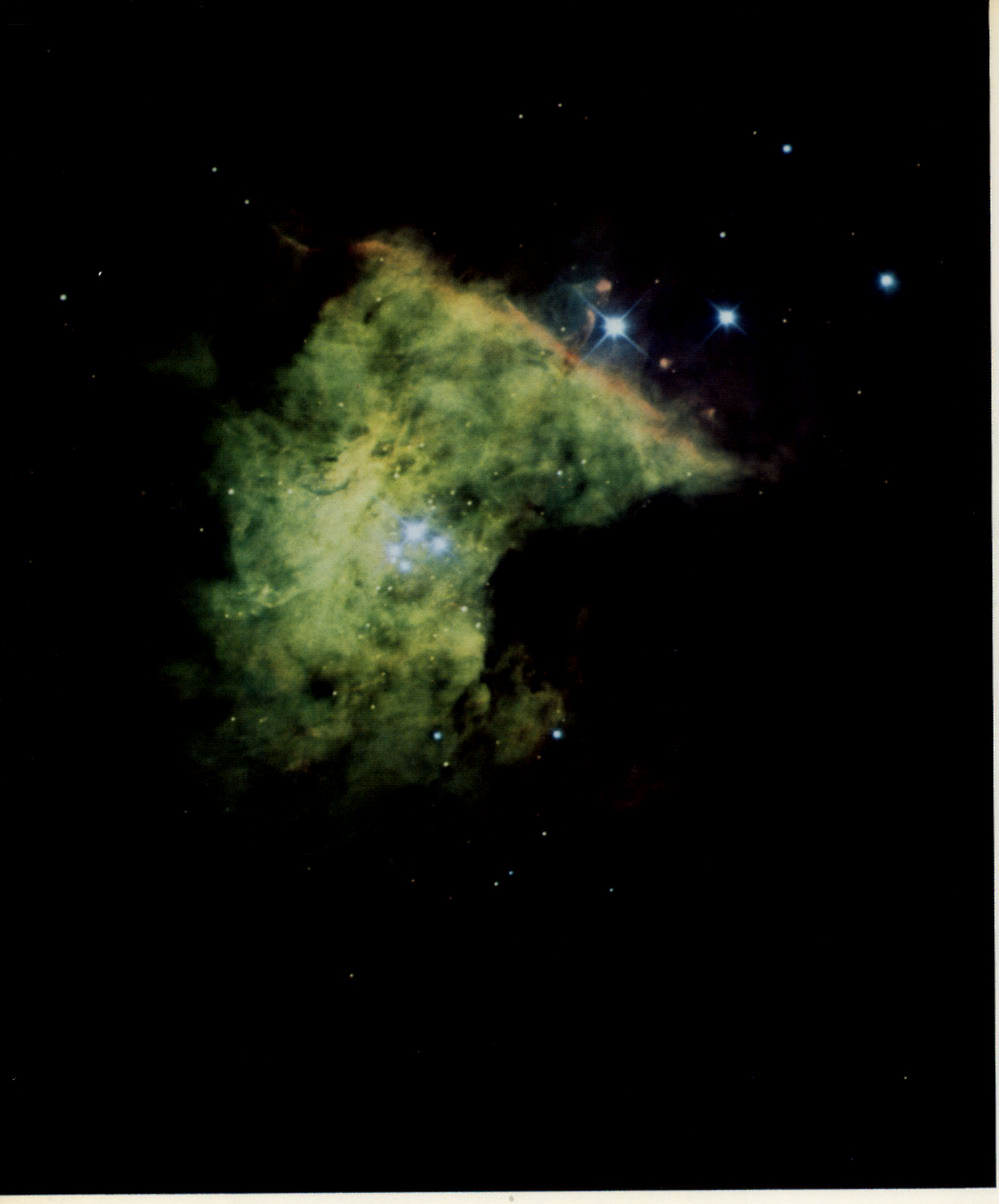


Plate XI
Orion nebula, NGC 1976, around the Trapezium. *(Lick Observatory photograph)*

Plate XIV
The Lagoon nebula, NGC 6523. *(Lick Observatory photograph)*

Plate XII
M51 or NGC 5194, a spiral nebula in Canes Venatici. *(Lick Observatory photograph)*

Plate XIII
M82 or NGC 3034, an unusual galaxy in Ursa Major. *(Lick Observatory photograph)*

Plate XV
Orion nebula, NGC 1976, inner region (white) is burnt out in this plate. Compare with Plate XI. *(Lick Observatory photograph)*

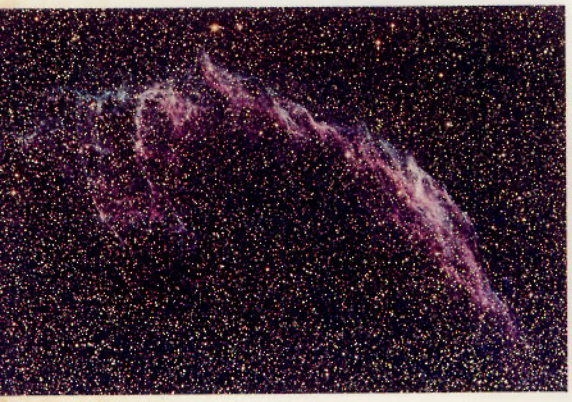


Plate XVI
Veil nebula in Cygnus, NGC 6992. *(Photograph from the Hale Observatories)*

Plate XVII
Rosette nebula in Monoceros, NGC 2237. *(Photograph from the Hale Observatories)*

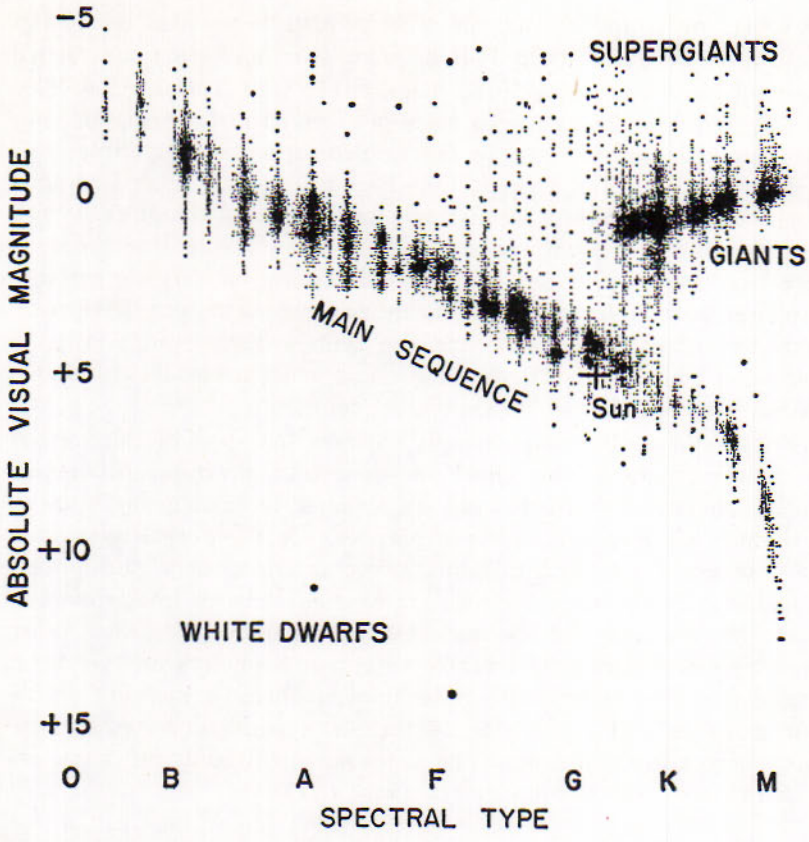


Figure 11.27
Spectrum-luminosity diagram of 6700 stars in the sun's region of the Galaxy. Each dot represents a star. *(Diagram by Lund Observatory)*

average blue star and the same amount brighter than the average red star of the sequence. Because the sun as viewed from a distance of 10 parsecs would be a faint star to the naked eye, the red stars of the sequence must generally be telescopic objects. Red stars such as Betelgeuse and Antares, however, are among the apparently brightest stars. More distant than 10 parsecs, they are much more luminous than the sun. These and other stars of high luminosity are represented by points that appear above the band of the main sequence.

11.28 Giant and Dwarf Stars

Since its introduction by H. N. Russell in 1913, the spectrum-luminosity diagram has played a leading part in directing the studies of the stars. Ejnar Hertzsprung had previously drawn attention to the sharp distinction between red stars of high and low luminosity and had named them giant and dwarf stars, respectively. The original term *dwarf stars* is commonly used today to denote main-sequence stars fainter than about absolute visual magnitude +1.

Figure 11.29
Sirius and companion photographed with the Sproul 24-inch refractor with a hexagonal diaphragm outside the objective. 1961 March 26: separation 9".2; east to right, north below. *(Photograph: S. L. Lippincott and J. K. Wooley)*

The spectrum-luminosity diagram is accordingly known as the Hertzsprung-Russell diagram, or the H-R diagram. It is now often superseded by the equivalent color magnitude diagram (11.24A), because the color indexes of stars may be measured more precisely than the spectral types can be estimated, and may also be determined to fainter magnitudes.

Giant stars, such as Arcturus and Capella, are decidedly more luminous than main-sequence stars of similar spectral types and therefore having about the same surface temperatures as these; they are brighter because they are larger than the main-sequence stars. *Supergiant stars* are extraordinarily large and luminous giants. Examples are Antares and Betelgeuse. *Subgiants* are between the giants and the main sequence, and *subdwarfs* are somewhat below this sequence. *White dwarf stars* are far less luminous than main-sequence stars of corresponding colors.

Since it is not good enough simply to infer the sizes of the various classes of stars, considerable effort has gone into direct measurements. The most convenient measurements are obtained from eclipsing binaries (14.19) which are also spectroscopic binaries. For the very largest stars we may make direct measurements using interferometric techniques. Michelson and Pease were able, using a Michelson beam interferometer, to obtain the diameters of six stars. Hanbury-Brown and Twiss, using an intensity interferometer, were able to obtain diameters for two main sequence stars. The interferometer techniques obtain a measure of the angular diameter and knowledge of the star's parallax is required to convert this to a linear measure. The same statement holds for diameters measured by the lunar occultation technique (6.13).

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

11.29 White Dwarf Stars

The name given to this star is deceptive because some are yellow and at least one is red. Although their masses average half the sun's mass, their diameters generally range between one half and four times the earth's diameter. Their densities average 2×10^5 times the sun's density and may have much greater values. The companion of Sirius was among the earliest of this remarkable type of stars to become known.

White dwarfs comprise an estimated 3 per cent of all the stars of our galaxy. Their low luminosity, however, permits only the very nearest ones to be observed. W. J. Luyten lists about two thousand stars of this type, the majority of which he has discovered. In his photographic search over more than a quarter of a century for stars of large proper motion, which are accordingly nearby, Luyten has found that most of the faint stars among them are red members of the main sequence. An occasional one has proved to be either a white dwarf or a subdwarf, between the main sequence and the white dwarfs in the color-magnitude diagram.

A white dwarf star is described by Greenstein as mainly a degenerate mass devoid of hydrogen, surrounded by a nondegenerate envelope 65

miles deep, and above this an atmosphere of a sort only a few hundred feet deep. *Degenerate matter* conforms to an equation of state different from the ordinary gas laws. According to the theory, the radius of a completely degenerate star is inversely proportional to the mass, which cannot exceed 1.4 times the sun's mass.

11.30
Spectra of White Dwarf Stars

Greenstein has reported on his studies of the spectra of 50 white dwarf stars, which he has already photographed with the 200-inch telescope in his undertaking to record the spectra of all that are available to apparent magnitude 16. Some spectra (Fig. 11.30A) show prominent helium absorption, others show strong dark hydrogen lines, and at least one spectrum has only the H and K lines of ionized calcium and a line of neutral magnesium. The lines are generally much widened by compression of the very dense stars, and in some cases (type DC) are not seen at all. The other preliminary types assigned to white dwarfs are DB, DA, DF, DG, and DK; these are in order of increasing color index of the stars. Although there is not a simple relation between the spectra and colors of these stars, the color-magnitude diagram (Fig. 11.30B) reveals a rather convincing sequence below the main sequence.

The theory of general relativity predicts that the lines in the spectrum of a star should be displaced to the red by an amount that is directly

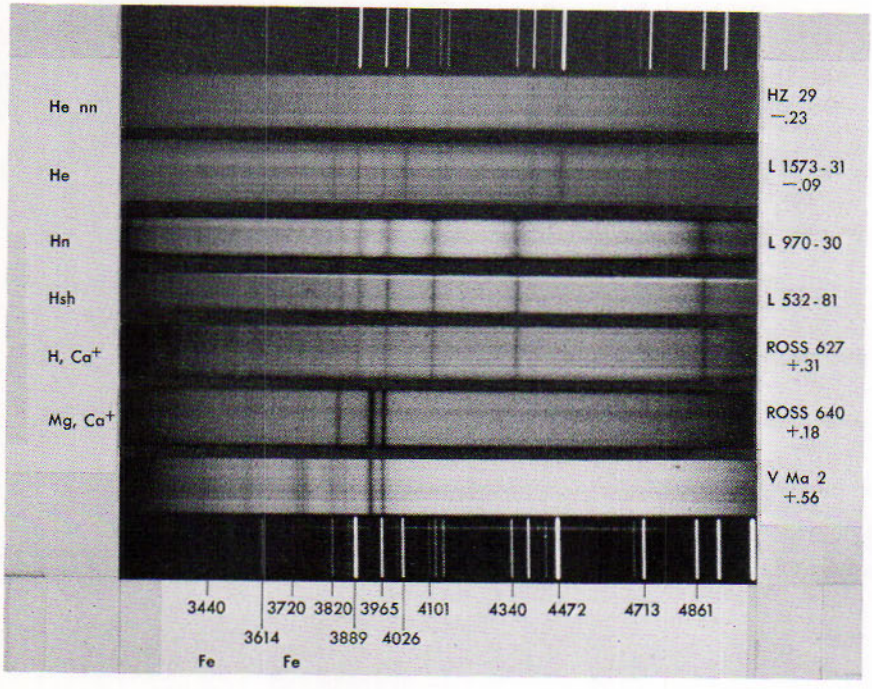


Figure 11.30A
Spectra of white dwarfs, with prominent elements in each spectra shown on the left. *(Photograph by J. L. Greenstein, California Institute of Technology)*

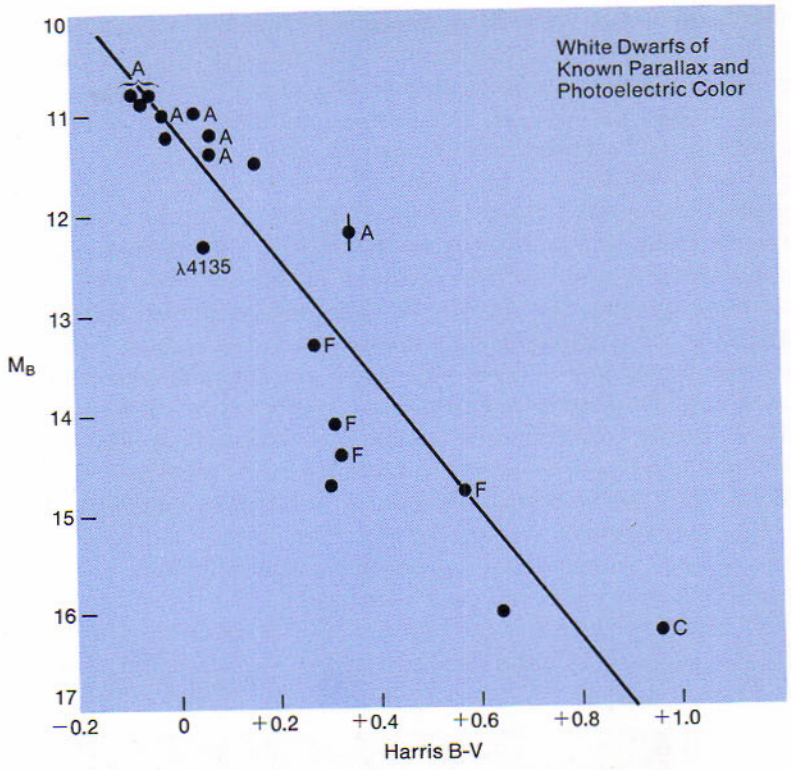


Figure 11.30B
Color-absolute magnitude diagram of white dwarf stars. *(Magnitudes and color indexes by D. Harris; Diagram by J. L. Greenstein)*

proportional to the cube root of the star's mean density. In the case of the sun the predicted displacement corresponds to a velocity of recession of only 0.6 km/sec and is masked by other effects. For the very dense white dwarfs the displacements should be much greater.

D. M. Popper has observed about the expected relativity shift of the lines in the spectrum of the 9th-magnitude white dwarf companion of the star 40 Eridani. His measured redshift is equivalent to a velocity of recession of 21 km/sec, which is considered in satisfactory agreement with the predicted value of 17 km/sec corresponding to a diameter of 22,500 kilometers for this star.

11.31 Luminosity Function

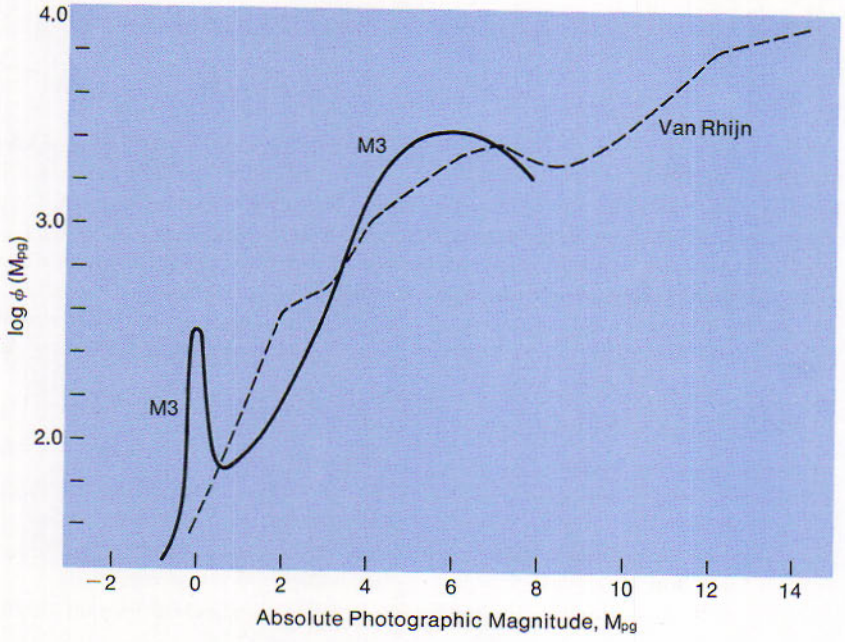
The relative numbers of stars for successive intervals of absolute magnitude in a sample volume of space constitute the *luminosity function* for that sample. The broken line of Fig. 11.31 refers to samples in the sun's vicinity, studied especially by P. J. van Rhijn in Holland and S. W. McCuskey at Warner and Swasey Observatory. The supergiants, at the left

of its bright end, are scarce. Beyond the faint end, at the right, the numbers are still increasing. For stars nearest the sun indications are that the maximum number is reached at absolute photoraphic magnitude +15.5.

Stars of low luminosity are evidently in the great majority in a sample volume of space around the sun. As we see them in the sky, however, the stars of high luminosity are the most numerous. The reason is that the less luminous stars must be nearer in order to be visible.

As examples, consider the lucid stars. The greatest distance, r in parsecs, at which a star of absolute visual magnitude M is visible to the naked eye is found by the formula (11.25): $\log r = (6.2 - M + 5)/5$. Here we suppose that the faintest lucid star is of apparent visual magnitude +6.2 and that no cosmic dust intervenes. By this formula, the limiting distance is 1740 parsecs for a supergiant star of absolute magnitude −5 and is reduced to 17 parsecs for a star as bright as the sun. The limit would be only 0.17 parsec, or slightly more than half a light

Figure 11.31
Relative Numbers of Stars of Different Luminosities. The logarithms of the numbers for successive intervals of absolute magnitude for stars in the sun's vicinity are represented by the broken line, and for stars of the globular cluster M 3 by the full line. *(Diagram by Allan R. Sandage)*



year, for a star having $M = +15$; such a star would have to be much nearer us than the nearest known star in order to be visible to the naked eye.

The full curve of the figure refers to stars of a different sample, studied by Sandage in the globular cluster M 3 (16.16).

REVIEW QUESTIONS

1. What is meant by:
   a) heliocentric parallax
   b) annual parallax
   c) trigonometric parallax
   d) absolute parallax
2. Show with the aid of a diagram what the parallactic ellipse looks like for a star at:
   a) R.A. 18^h, Decl + 65°
   b) R.A. 6^h, Decl + 25°
   c) R.A. 18^h, Decl + 20°
3. For a given star, at what points in the earth's orbit is the parallax best determined? Can you suggest some reasons for measuring the parallax at points other than these best points?
4. Define a star's proper motion. How is it measured? Why is a star with a large proper motion likely to prove interesting?
5. How is a star's radial velocity measured?
6. The Large Magellanic Cloud shows almost no annual change in radial velocity. Why?
7. Vega exhibits a small annular variation about its mean value of −19 km/sec. Explain.
8. In what two ways do we determine the solar motion? How is the solar motion used?
9. Describe the principle features of the Harvard spectral classification system.
10. Show that a magnitude difference of 5 corresponds to a difference of a factor of 10 in the distances for two stars of the identical spectral type.
11. Radio astronomers have a term equivalent to magnitudes called decibels, where the constant of proportionality is 10. In relating terms, one magnitude corresponds to how many decibels?
12. In the telescope the visual binary 85 Pegasus resolves into two stars of apparent magnitudes of 5.8 and 8.8. What is the combined magnitude of the unresolved pair?
13. If the color index of a star is 0.0 what is the temperature of the star? What type of star does this correspond to?
14. What is meant by the term main sequence?
15. If white dwarfs are as numerous as 3% of all stars, why is it that we see so few?
16. What would be the distinction of a star having an absolute magnitude of +15 and an apparent magnitude of +10?
17. Construct and discuss the spectrum-luminosity diagram obtained from Table 11.I.

REFERENCES

Dufay, Jean, *Introduction to Astrophysics: The Stars*, trans. O. Gingerich, New York, Dover Publications, 1964.

Page, Thornton, and W. L. Page, *Starlight*, New York: The Macmillan Co., 1967.

van de Kamp, Peter, *Basic Astronomy*, New York, Random House, 1952.

Wood, Frank B., ed., *Photoelectric Astronomy for the Amateur*, New York: The Macmillan Co., 1963.

FOR FURTHER STUDY

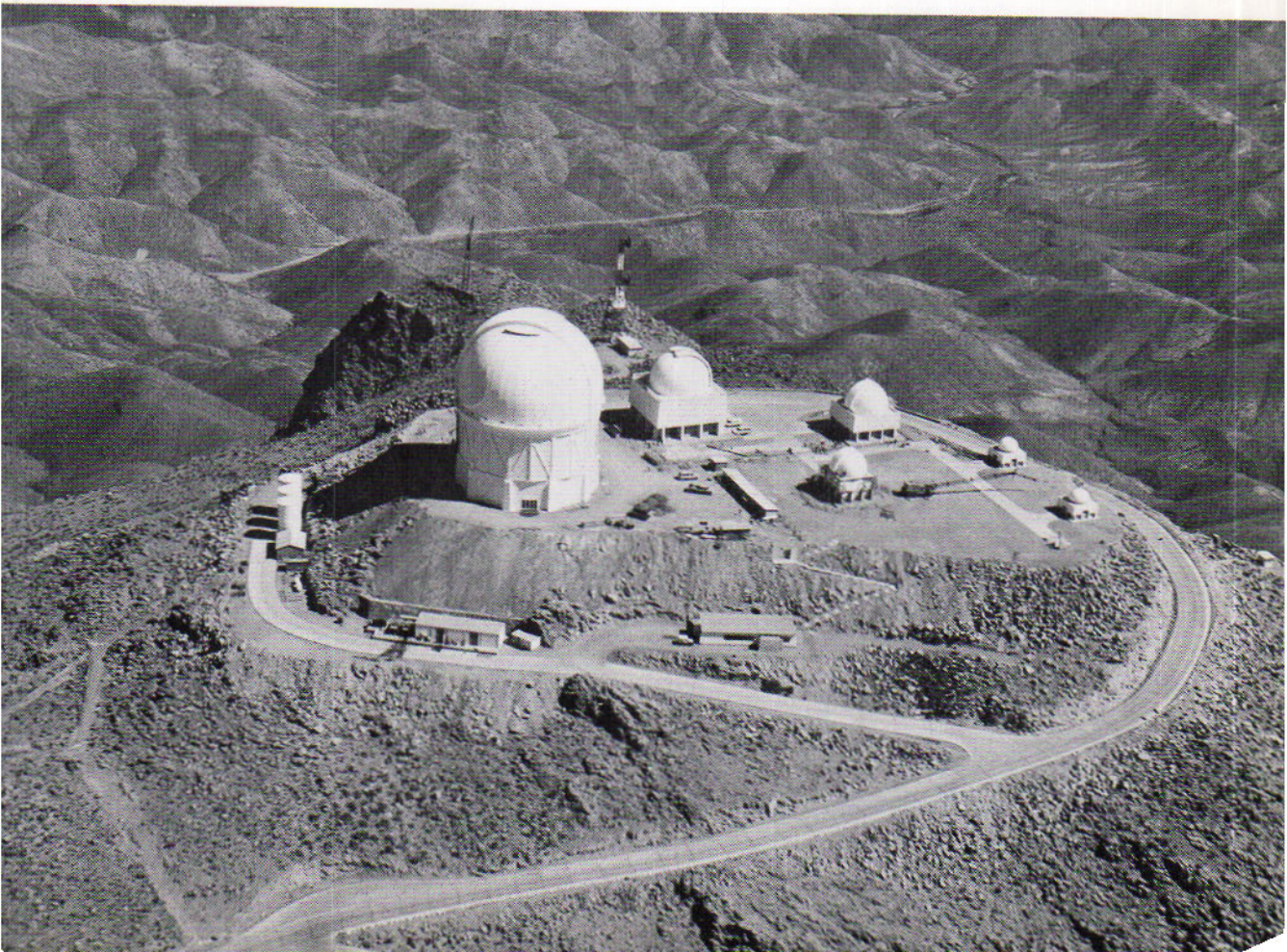
Allen, Clabon Walter, *Astrophysical Quantities*, 2nd ed., London: Athlone Press, 1955.

Strand, K. A., *Basic Astronomical Data*, Chicago: University of Chicago Press, 1968.

Swihart, Thomas L., *Astrophysics and Stellar Astronomy*, New York: John Wiley and Sons, 1968.

van de Kamp, Peter, *Principles of Astrometry*, San Francisco: Freeman, Cooper & Co., 1967.

Cerro Tololo Inter-American Observatory in the Andes, near La Serena, Chile. *(AURA Photograph)*




12

STELLAR ATMOSPHERES AND INTERIORS

ATOMIC STRUCTURE AND RADIATION — STELLAR ATMOSPHERES — EXTENDED ATMOSPHERES AND ENVELOPES — INTERIORS OF THE STARS

Like the sun, the stars are globes of intensely hot gas. Their radiations emerge from their photospheres and filter through their atmospheres, where dark lines are formed in their spectra. The atmospheres around some stars are extended enough to imprint bright lines in the spectra, and larger envelopes may be visible directly with the telescope.

This chapter describes evidence given by stellar spectra concerning the exteriors of the stars. It then considers what the interiors may be like to produce the exterior phenomena and what processes in the interiors may supply the energy to keep the stars shining. These subjects are introduced by a brief account of some relations between the constituents of a gas and the radiations they absorb and emit.

The "Owl" nebula in Ursa Major. *(Photograph from the Hale Observatories)*

ATOMIC STRUCTURE AND RADIATION

The study of stellar atmospheres and interiors draws heavily upon atomic and nuclear physics. Here we take a simplified concept of the atom to explain our observations. Similar concepts are used to explain molecular spectra.

12.1 Constituents of the Atom

Atoms are the building blocks of all material. They are composed essentially of electrons, protons, and neutrons. The *electron* is the lightest of these constituents; its mass is 9.1055×10^{-28} grams and it carries unit negative charge of electricity. The *proton* is 1836.57 times as massive as the electron and carries unit positive charge. The *neutron* has about the same mass as the proton and is electrically neutral.

The nucleus of the atom ranges progressively from the single proton of the ordinary hydrogen atom to compact groups of protons and neutrons in the heavier atoms. Each added proton contributes one unit to the positive charge on the nucleus. In the normal atom the nucleus is surrounded by negatively charged electrons equal in number to the protons, so that the atom as a whole is electrically neutral.

Among the products that issue from atoms when they are vigorously bombarded or disintegrate spontaneously are positrons and photons. The *positron* has the same mass as the electron but carries unit positive charge. The *photon* is a unit bundle of energy. There are also mesons and particles having masses greater than that of the proton. All these constituents and products of atoms are considered to be wave formations, but are often pictured as particles.

The following descriptions of some atomic processes employ the conventional model of the atom proposed, in 1913, by N. Bohr. They begin with the atom of hydrogen, the simplest and also the most abundant in the universe.

12.2 Model of the Hydrogen Atom

The normal hydrogen atom consists of one proton attended by a single electron. In the Bohr model the electron revolves around the proton analogously to a planet revolving around the sun. The force holding the electron in its orbit is that of the attraction between the unlike electric charges which, like the gravitational force, is inversely proportional to the square of the distance between them. Whereas a planet in a two-body system would remain always in the same orbit, the electron may be found at different times in a variety of possible orbits.

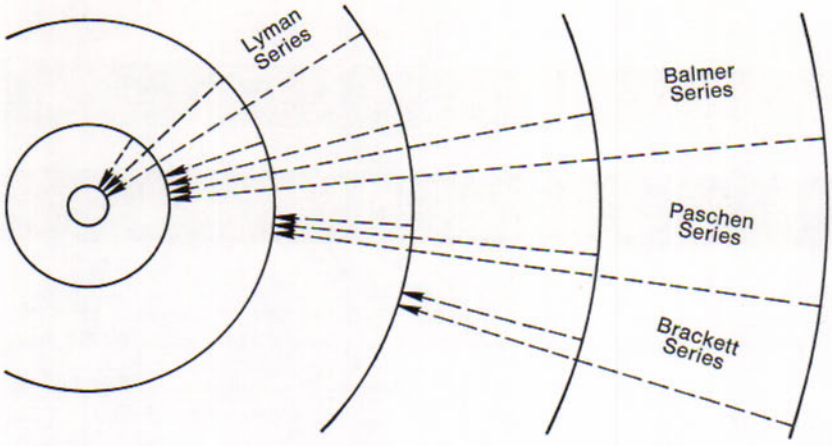


Figure 12.2
Conventional representation of the hydrogen atom. Possible orbits of the electron around the nucleus are shown as circles.

The radii of the permitted orbits around the proton (Fig. 12.2) are proportional to the squares of the integers; that of the innermost orbit is about half an angstrom. The atom does not absorb or emit radiation as long as the electron remains in the same orbit. It absorbs energy, as from radiation that strikes it, only when it can find the right quantity to raise its electron exactly to a higher orbit against the attraction of the proton. At once the electron falls back again, releasing as radiation the energy it absorbed. Here is justification of the rule previously stated (4.8), that a gas abstracts from light passing through it the same wavelengths that the gas itself emits.

$\Delta t \sim 10^{-8}$ secs

The relation between the gain or loss of energy by the atom in such a transition and the frequency of the radiation absorbed or emitted is $E_1 - E_2 = h\nu$. E_1 is the energy of the atom when the electron is in one orbit, and E_2 is the energy when the electron is in another orbit. The difference between the two states is the photon, $h\nu$, which is absorbed and then emitted. The quantity of this unit bundle of energy is given by the constant h times the frequency, ν, of the radiation. Thus the frequency of the radiation absorbed or emitted is higher, and the wavelength is shorter, as the energy difference between the two orbits is greater.

$\nu\lambda = c$

12.3 Series of Hydrogen Lines

A prominent feature of the spectra of the hotter stars as ordinarily photographed is a long series of hydrogen lines. Beginning with the Fraunhofer C line in the red, the lines appear along the spectrum at diminishing intervals, like a succession of telegraph poles receding in the distance, until they close up in the near ultraviolet. These lines of the *Balmer series* (Fig. 12.3) are designated in order: Hα, Hβ, Hγ, and so on. More than 30 are identified in stellar spectra and in the sun's chromosphere.

Balmer, a Swiss physicist, derived in 1885 an empirical formula by

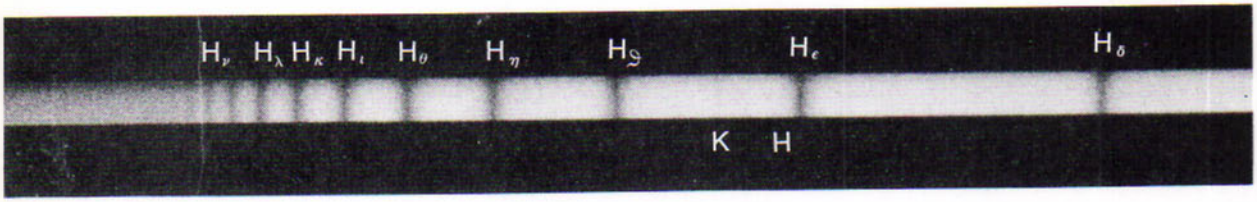


Figure 12.3
Hydrogen lines in the spectrum of an A0V Star. The Balmer series is shown from Hδ to the limit of the series in the deep ultraviolet. *(Courtesy of* Catharine *Garmany)*

which the wavelength of any line in the series may be calculated. In the more general form, which applies to other hydrogen line series as well, the formula is:

$$1/\lambda = R(1/m^2 - 1/n^2),$$

where λ is the wavelength of the line in centimeters; R is about 109,678cm^{-1}, m is the number of the orbit in the Bohr model from which the electron is raised to produce a dark line and to which it returns to produce a bright line in the spectrum; n is any whole number greater than m, representing the number of the orbit to which the electron is raised. A single hydrogen atom can promote only one electron transition at any instant; but in a gas containing very many atoms all possible transitions are likely to be in progress.

In the Balmer series, $m = 2$ and n is put equal to 3, 4, 5, and so on. Four other hydrogen series have been observed in celestial and laboratory spectra. These are the *Lyman series* ($m = 1$) in the extreme ultraviolet, the *Paschen series* ($m = 3$), the *Brackett series* ($m = 4$), and the *Pfund series* ($m = 5$), all in the infrared. The lines in the spectra of some other chemical elements are also arrayed in series.

12.4 Chemical Elements

Table 12.I lists the names, symbols, atomic numbers, and atomic weights of 96 chemical elements. Numbers 93 to 96, and more recently to 103, were produced in radiation laboratories.

The *atomic number* of an element is the number of protons in the nucleus and also the number of electrons around the nucleus of the normal atom. All atoms having the same atomic number belong to the same chemical element.

The *atomic weight* is the mass of the atom. The unit of mass employed in the table is one sixteenth the mass of the average oxygen atom, taken as weight 16.0000; the value of the unit is 1.660×10^{-24} grams. The atomic weight given here is frequently the average for two or more different kinds of atoms, or *isotopes*, of the same element, which differ in mass because their nuclei have different numbers of neutrons.

The *mass number* of an atom is its atomic weight rounded off to the nearest whole number, the sum of its protons and neutrons. Thus the mass numbers of hydrogen, helium, and lithium corresponding to their weights in the table are, respectively, 1, 4, and 7. In formulas of nuclear reactions a particular atom is conveniently designated by its symbol having the atomic number as subscript and the mass number as superscript. As an example, the iron isotope $_{26}Fe^{56}$ contains 26 protons and 30 neutrons.

12.5 The Electron Shells

Turning from the simple hydrogen atom to more complex ones, we find less confusion of electron transitions than might at first be expected. Most of the electrons are so firmly held in filled shells that they are not easily raised to higher levels.

The *shells* are the same as the Bohr hydrogen orbits. They are filled in order of distance from the nucleus when they acquire $2n^2$ electrons, where n is the number of the Bohr orbit. Thus the first shell is filled by 2 electrons, the second by 8, the third by 18, and so on. A filled shell will not receive additional electrons and is reluctant to release any that it possesses. As examples, Fig. 12.5 shows the outer structures of a number of lighter atoms having their electrons in the lowest possible orbits.

The normal hydrogen atom (atomic number 1) has its single electron in the first orbit, from which it can be raised with moderate effort. The helium atom (number 2) has its two electrons filling the first shell, which is not easily broken. The lithium atom (number 3) has two electrons locked in the inner shell and a third in the second orbit, from which it is easily removed. In the successively heavier atoms: beryllium, boron, carbon, and so on, electrons are added one at a time to match the added protons in the nucleus. In the neon atom (number 10) the second shell is filled. The sodium atom (number 11) has two filled shells and the additional electron in the third orbit ready to jump out on slight provocation.

Some of the atoms represented in the figure have shells in the making and others have only filled shells, as in the cases of the chemically inactive elements helium and neon. Evidently the amount of energy the atom must absorb to raise an electron from its normal position to a higher level depends on the number of protons attracting from within

TABLE 12.I
The Chemical Elements

ELEMENT	SYMBOL	ATOMIC NUMBER	ATOMIC WEIGHT*	ELEMENT	SYMBOL	ATOMIC NUMBER	ATOMIC WEIGHT*
Hydrogen	H	1	1.0080	Indium	In	49	114.82
Helium	He	2	4.003	Tin	Sn	50	118.70
Lithium	Li	3	6.939	Antimony	Sb	51	121.76
Beryllium	Be	4	9.013	Tellurium	Te	52	127.61
Boron	B	5	10.812	Iodine	I	53	126.91
Carbon	C	6	12.012	Xenon	Xe	54	131.30
Nitrogen	N	7	14.007	Cesium	Cs	55	132.91
Oxygen	O	8	16.0000	Barium	Ba	56	137.
Fluorine	F	9	18.999	Lanthanum	La	57	138.92
Neon	Ne	10	20.	Cerium	Ce	58	140.13
Sodium	Na	11	22.991	Praseodymium	Pr	59	140.91
Magnesium	Mg	12	24.313	Neodymium	Nd	60	144.25
Aluminum	Al	13	26.98	Promethium	Pm	61	(147)
Silicon	Si	14	28.09	Samarium	Sm	62	150.36
Phosphorus	P	15	30.975	Europium	Eu	63	151.96
Sulfur	S	16	32.066	Gadolinium	Gd	64	157.25
Chlorine	Cl	17	35.454	Terbium	Tb	65	158.93
Argon	A	18	39.949	Dysprosium	Dy	66	162.50
Potassium	K	19	39.103	Holmium	Ho	67	164.94
Calcium	Ca	20	40.08	Erbium	Er	68	167.27
Scandium	Sc	21	44.96	Thulium	Tm	69	168.94
Titanium	Ti	22	47.90	Ytterbium	Yb	70	173.04
Vanadium	V	23	50.95	Lutetium	Lu	71	174.98
Chromium	Cr	24	52.00	Hafnium	Hf	72	178.50
Manganese	Mn	25	54.94	Tantalum	Ta	73	180.95
Iron	Fe	26	55.85	Tungsten	W	74	183.86
Cobalt	Co	27	58.94	Rhenium	Re	75	186.3
Nickel	Ni	28	58.71	Osmium	Os	76	190.2
Copper	Cu	29	63.55	Iridium	Ir	77	192.2
Zinc	Zn	30	65.37	Platinum	Pt	78	195.09
Gallium	Ga	31	69.72	Gold	Au	79	197.0
Germanium	Ge	32	72.60	Mercury	Hg	80	200.60
Arsenic	As	33	74.92	Thallium	Tl	81	204.38
Selenium	Se	34	78.96	Lead	Pb	82	207.20
Bromine	Br	35	79.912	Bismuth	Bi	83	208.99
Krypton	Kr	36	83.80	Polonium	Po	84	210
Rubidium	Rb	37	85.48	Astatine	At	85	(211)
Strontium	Sr	38	87.63	Radon	Rn	86	222
Yttrium	Y	39	88.91	Francium	Fa	87	(223)
Zirconium	Zr	40	91.22	Radium	Ra	88	226.05
Niobium	Nb	41	92.91	Actinium	Ac	89	227
Molybdenum	Mo	42	95.95	Thorium	Th	90	232.05
Technetium	Tc	43	(99)	Protactinium	Pa	91	231
Ruthenium	Ru	44	101.1	Uranium	U	92	238.04
Rhodium	Rh	45	102.91	Neptunium	Np	93	(237)
Palladium	Pd	46	106.4	Plutonium	Pu	94	(239)
Silver	Ag	47	107.874	Americium	Am	95	(241)
Cadmium	Cd	48	112.41	Curium	Cm	96	(242)

* Atomic weights, except those in parentheses, are from the *Journal of the American Chemical Society* (1956). The numbers and names of transuranic elements more recently produced are: 97 berkelium, 98 californium, 99 einsteinium, 100 fermium, 101 mendelevium, 102 nobelium, 103 lawrencium.

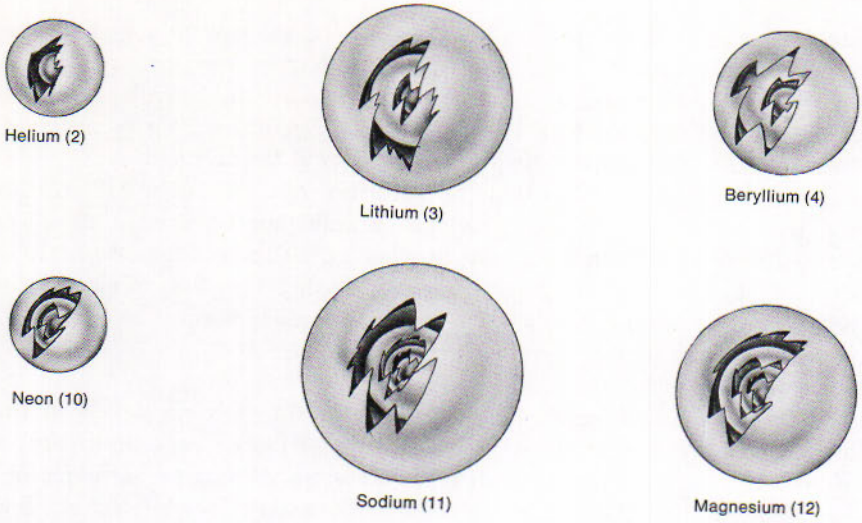


Figure 12.5
Shell model electron structure of some lighter atoms.

and the arrangement of the other electrons in the superstructure. Each kind of atom presents its peculiar demands in this respect. Here we have the explanation of the principle (4.8) that each gaseous element produces its characteristic pattern of spectrum lines.

12.6 Neutral and Ionized Atoms

The *neutral atom* has its full quota of electrons, equal to the number of protons in the nucleus, so that it is electrically neutral. The *ionized atom* has generally lost one or more electrons. It has absorbed enough energy to transfer these electrons beyond the outermost orbit. They have become *free electrons,* free to dart about independently until they are captured by ionized atoms.

The *singly ionized atom* has lost a single electron and has thereby acquired a single unit positive charge. The *doubly ionized atom* has lost two electrons and has an excess of two positive charges. Each succeeding ionization requires a greater amount of energy to promote it. The extent of the ionization is indicated by adding a Roman numeral to the symbol for the element (number I designates the neutral atom). Thus singly ionized helium is written He II.

The removal of an electron leaves the superstructure of the atom similar to that of the next lower atomic number. As an example, the singly ionized helium atom has one electron left and in this respect resembles the neutral hydrogen atom. The important difference between the two is that the helium nucleus has two positive charges and thus holds the electron 4 times as tenaciously as does the hydrogen nucleus with its

single charge. The energy required to raise the electron to a higher level is accordingly 4 times as great, so that the spectrum line produced by the transition has 4 times the frequency, or one fourth the wavelength of its hydrogen counterpart. The lines of ionized helium ordinarily observed correspond to the Brackett series of hydrogen in the infrared.

Thus the pattern of lines in the spectrum of a gas, whether it is in the laboratory or surrounding a star, reveals not only what chemical elements are represented but also to what extent their atoms are ionized. Certain negative ions are permissible, where the atom has an electron in addition to its normal number. The negative hydrogen ion is an important source of continuous absorption in stars like the sun.

12.7 The Energy-Level Diagram

The Energy-Level diagram (Fig. 12.7A) substitutes energy levels in the atom for the orbits of the original Bohr model. The present theory of atomic structure assigns to each energy level a distinctive wave pattern having properties that can be expressed mathematically. A spectrum line corresponds to a transformation from one wave pattern to another.

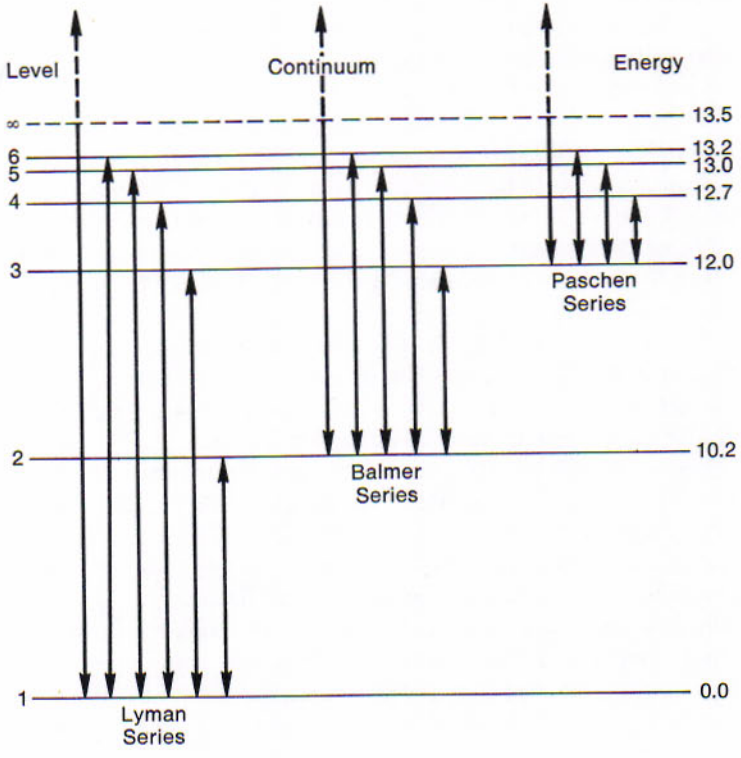
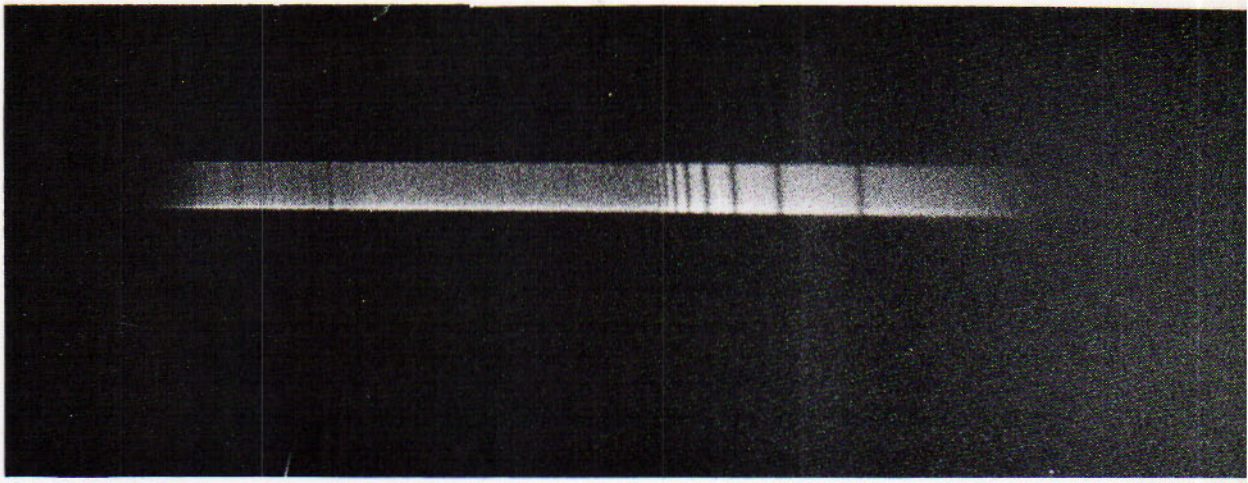


Figure 12.7A
Energy-level diagram of the hydrogen atom. The levels correspond to the Bohr orbits. The numbers at the right are proportional to the energy required to raise the electron from the lowest level.



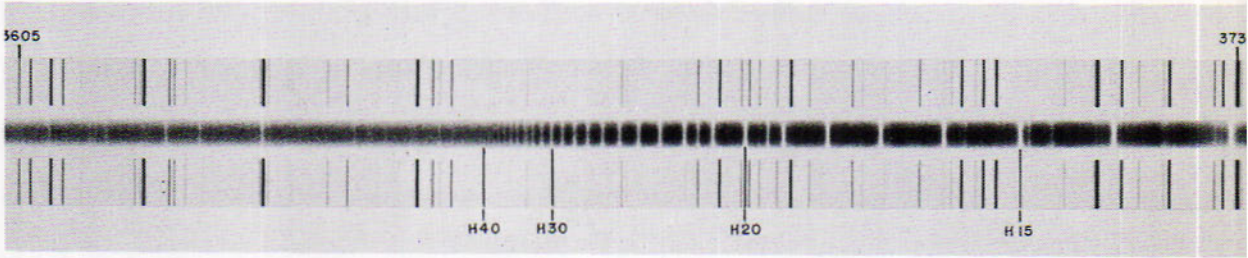


Figure 12.7B (Top) The Balmer series of hydrogen in Sirius. MgII doublet is visible at the left. *(NASA Gemini 12 Photograph courtesy of K. Heinze)*

Figure 12.7C (Bottom) Spectrum of the star HD 193182, showing Balmer series down to and including the continuum. *(Photograph from the Hale Observatories)*

The lowest level, corresponding to the innermost Bohr orbit, is the base of electron transitions producing the Lyman series of lines in the hydrogen spectrum. Level 2 is the base of the Balmer series, 3 of the Paschen series, and 4 of the Brackett series. The vertical lines in the figure represent transitions of electrons from lower to various higher levels and conversely, and the lengths of these lines are proportional to the energy expended or released in the transitions.

The various transitions are not equally likely. For example, an electron in the third orbit of hydrogen has a higher probability of going directly to the ground state giving up a Lβ photon than it does of going to the second orbit giving up a Hα photon. Consideration of orbit lifetimes and transition probabilities complicates spectral interpretation and is a fascinating and challenging area of astronomy.

Certain transitions have extremely small probabilities and are referred to as *forbidden transitions.* Such transitions play a very important role in astrophysics and actually are not absolutely forbidden. They occur most easily in high temperature, low density conditions and hence are common in astronomy. The most famous lines arising from forbidden transitions

are the oxygen lines first seen in planetary nebulae and, of course, the hydrogen 21-cm line.

The higher the level, the less is the additional energy required to raise the electron to the next higher level against the weakened attraction of the more distant nucleus. Near the top level only a little more energy is needed to remove the electron completely and thus to ionize the atom.

A *continuum,* or continuous absorption, extends from the head of the Balmer series far into the ultraviolet (Fig. 12.7A) in the spectra of the hotter stars. It is produced when the energy absorbed by electrons at the second level is in excess in various amounts of that barely necessary to remove them from the atoms. Conversely, the continuum in emission is produced when free electrons having different velocities are captured by the atoms and fall to the second energy level. A dark or bright continuum may also appear beyond the limits of other line series.

12.8 Molecular Spectra

A *molecule* is here defined as a combination of two or more atoms, whereas in chemistry it may also include separate atoms. The astrophysical distinction between the atom and the molecule arises from the difference between the spectra of the two. Atomic spectra contain only lines. Molecular spectra are characterized by series of bands. Each *band* is many angstroms long and is composed of systematically spaced lines.

As the simplest case, the diatomic molecule consists of two positive ions surrounded by electrons. It may be pictured as a rotating dumbbell having an elastic rod. Its energy at any instant is derived from: (1) Its rotation on an axis at right angles to the line joining the two ions; (2) its vibration along that line; (3) its electronic energy states analogous to those of the atom. The spectrum of the molecule may be interpreted by extensions of models of the atom.

Molecules are generally identified by comparing a stellar spectrogram with a laboratory spectrogam of the suspected molecule. Since molecular spectra are often temperature sensitive to a very great extent, the obtaining of the proper laboratory spectrum is not a simple task. In the case of microwave spectra of molecules it is even more difficult and laboratory

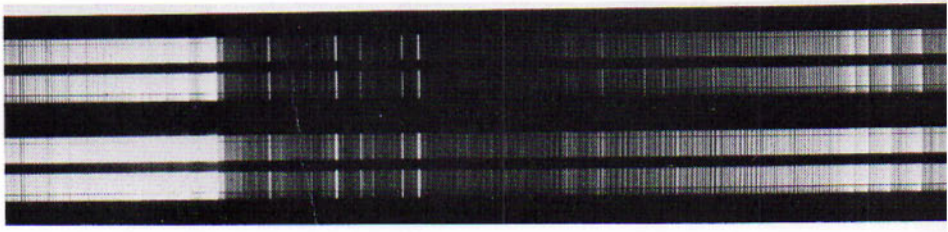


Figure 12.8
Cyanogen bands in the spectrum of the carbon arc. These bands are conspicuous in the spectra of some red supergiant stars. *(Photograph from the Hale Observatories)*

spectra of molecules of astronomical importance in this region of the spectrum are very small in number.

Molecules such as the oxides of titanium and zirconium, compounds of carbon and water vapor appear prominently in the spectra of the cooler stars. Molecules of water vapor, carbon dioxide, and oxygen are abundant in the earth's atmosphere where their dark bands obscure celestial spectra over large ranges of wavelength. In order to overcome this handicap and to observe these molecules in other planets and stars, astronomers are using high-level aircraft, balloons and satellites. The observation of molecules in inter-stellar space is discussed in Chapter 16.

STELLAR ATMOSPHERES

Although the absorption and the continuous spectrum of a star are produced in the same layers, it will be useful for our purpose to think of the dark lines as originating in the atmosphere above the star's photosphere. We are concerned here with effects of temperature and density in stellar atmospheres as shown by the spectra.

12.9 Temperature by the Radiation Laws

The *effective temperature* of a star is the temperature that a perfect radiator of the same size must have to produce the same output of radiation. It is calculated by the radiation laws (10.4) from the observed quantity of the star's radiation. One calculation, based on Stefan's law, employs the radiation of all wavelengths as measured by its heating of a thermocouple, and resembles the evaluation of the sun's temperature from the solar constant. This is often referred to as *bolometric temperature.*

A second calculation of the temperature, by Planck's formula, is based on the intensities of the star's radiation in various wavelengths. The intensities are measured in different parts of the available regions of the spectrum, and most accurately by use of the photoelectric cell attached to a spectroscope. In such procedures it is necessary to allow for whatever absorption of the radiation occurs on the way from the star to the observer.

12.10 Temperatures and Colors

The effective temperatures of stars at intervals of spectral type in Table 12.II are as given by P. C. Keenan and W. W. Morgan, except the first one. These are tied to an adopted surface temperature of 5730° for the sun, type G2; they may be in error by several hundred degrees, according to these investigators, so that the last significant figure is valid only in comparing the relative values.

TABLE 12.II
Effective Temperatures and Color Indexes

SPECTRUM	TEMPERATURE	COLOR B − V	SPECTRUM	MAIN SEQUENCE Temperature	MAIN SEQUENCE Color B − V	GIANTS Temperature	GIANTS Color B − V
		mag.			mag.		mag.
O5	50,000°K	−0.32	F5	6,700°K	+0.4	6,350°K	....
B0	27,000	−0.30	G0	6,000	+0.60	5,350	....
B5	15,600	−0.16	G5	5,520	+0.68	4,650	....
A0	10,400	0.00	K0	5,120	+0.82	4,350	+1.01
A5	8,200	+0.15	K5	4,350	+1.18	3,600	+1.52
F0	7,200	+0.30	M0	3,750	+1.45	3,400	+1.56
F5	6,700	+0.45	M5	....	+1.69	2,710	.....

The temperature diminishes as the spectral type progresses from O to M. The temperature change itself is mainly responsible for the succession of spectral patterns, as has been noted before; the explanation of this relation is given in following sections. Yellow and red stars of the main sequence have temperatures considerably higher than those of giant stars of corresponding types. It will be seen presently that the character of the spectrum is determined by the density as well as the temperature of the stellar atmosphere.

The colors, blue minus visual magnitude, of these types in the table are as determined photoelectrically by H. L. Johnson and W. W. Morgan. With diminishing temperature of the stars the most intense radiation is increasingly shifted toward the red in accordance with Wien's law, and the color increases; it is set at zero for type A0. Thus the bluest stars have negative color indexes and the reddest ones have the largest positive indexes as has been shown earlier (11.24).

12.11 Excitation and Ionization of Atoms

The *normal atom* has its electrons as close as possible to the nucleus. When the atom absorbs energy so that an electron is raised to a higher level, it becomes an *excited atom.* If the atom absorbs energy enough to remove an electron, it becomes ionized. The energy may be provided by a photon or a fast-moving particle. As the temperature of the gas is

greater, its particles dart about more rapidly and their collisions are more vigorous. Such collisions cause *thermal excitation and ionization* of the atoms.

The physicist M. N. Saha of India, in 1920, developed the rules of thermal ionization of atoms. His work ranks along with the Bohr model among the classics of atomic investigation. Saha presented an analogy between the evaporation of a liquid and the removal of electrons from atoms.

12.12 Thermal Ionization

When a covered container partly filled with a liquid is put in a sufficiently warm place, the liquid begins to evaporate. The amount of liquid diminishes until the space above it becomes saturated, so that just as much vapor is returning to the liquid as is coming out of it. The extent of the evaporation before this steady state is reached depends on 3 conditions.

1. The amount of liquid evaporated is greater as the temperature is higher.
2. The evaporation is greater at a particular temperature as the space above the liquid is more nearly empty at the start.
3. It is greater at a particular temperature and vapor pressure for some liquids than for others; compare, for example, the evaporation of ether and of water.

Consider the parallel case of the removal of electrons from the atoms of a gas. (1) The extent of the ionization is greater at higher temperatures, where the particles are colliding more vigorously. (2) It is greater in a rarer gas where fewer free electrons are available to replace the ones removed. (3) It is greater in similar conditions for atoms to which electrons are more loosely bound. We now examine the third relation.

12.13 Energy Required to Ionize an Atom

This energy is conveniently pictured in terms of an electron colliding with an atom. It is equivalent to the kinetic energy an electron acquires when it is accelerated across a potential difference of a specified number of volts. This *ionization potential* is accordingly expressed as a number of *electron volts*. We may here, if we choose, simply note the numbers themselves. An atom having a small ionization potential is ionized at a lower temperature, where it is subjected to more moderate collisions, than an atom having a larger number.

Table 12.III gives the ionization potentials of some chemical elements which are represented by prominent lines in stellar spectra. The column headed I refers to the neutral atom, II to the singly ionized atom, and III to the doubly ionized atom. Neutral sodium with a single electron in its outer shell has a small number, as would be expected. This atom should be excited or singly ionized in the atmospheres of the cooler stars. Neutral

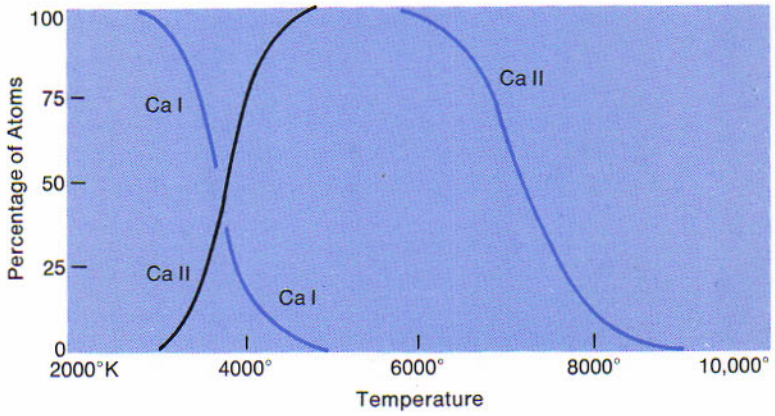
TABLE 12.III
Ionization Potentials of Selected Elements

ELEMENT	SYMBOL	STAGE OF IONIZATION		
		I	II	III
Helium	He	24.6	54.4	....
Nitrogen	N	14.5	29.6	47.4
Oxygen	O	13.6	35.1	54.9
Hydrogen	H	13.6	....	....
Carbon	C	11.3	24.4	47.9
Silicon	Si	8.1	16.3	33.5
Iron	Fe	7.9	16.2	30.6
Magnesium	Mg	7.6	15.0	80.1
Titanium	Ti	6.8	13.6	27.5
Calcium	Ca	6.1	11.9	51.2
Strontium	Sr	5.7	11.0	43.0
Sodium	Na	5.1	47.3	71.6

helium with its two electrons locked in the innermost shell has the largest number in this column. Its lines should be prominent only in the spectra of the hotter stars, where lines of ionized atoms of other elements should also appear. Note that the numbers in the table increase with successive stages of ionization.

As an example, Fig. 12.13 shows the effect of increasing temperature on the spectra of calcium atoms in the atmospheres of main-sequence stars.

Figure 12.13
Effect of temperature on calcium in the atmospheres of main-sequence stars. All calcium atoms are neutral (Ca I) at temperatures up to about 3000°K, where singly ionized atoms (Ca II) begin to appear. The second ionization begins at about 6000°. *(Adapted from a diagram in* Atoms, Stars and Nebulae *by L. Goldberg and L. H. Aller)*



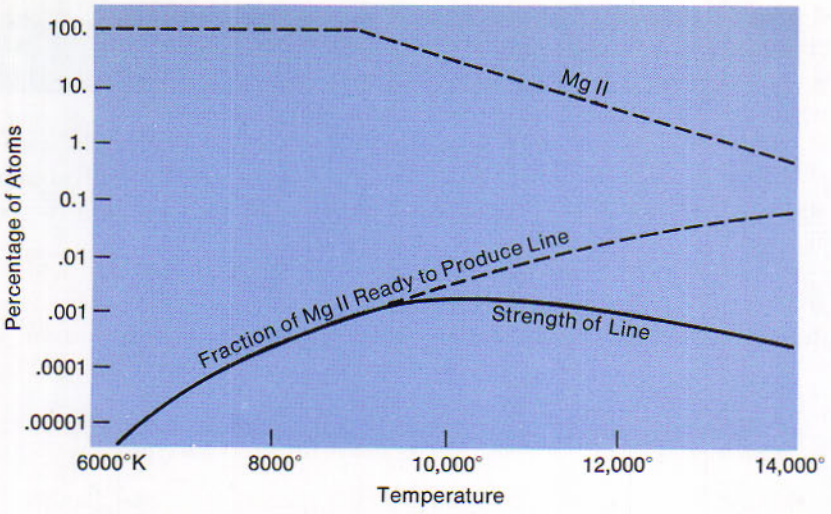


Figure 12.14
Relative strength of Mg II absorption Line 4481 Å at different temperatures. Showing the percentage of magnesium atoms in stellar atmospheres that are singly ionized at different temperatures, the fractions of these that are ready to produce the line, and the strength of the line. *(Diagram by C. H. Payne-Gaposchkin)*

12.14 Effect of Temperature on Stellar Spectra

Saha's theory of thermal ionization of atoms has been amplified and given more precise form. For a gas at an assigned temperature and pressure the Saha and Boltzmann formulas permit the calculation of the fraction of atoms in any stage of ionization and of these the fraction at any level of excitation. It is accordingly possible to predict how the patterns of lines in the spectra of stars must change with increasing temperature of the stars. Conversely, the pattern observed in the spectrum of a particular star informs us of the surface temperature of that star.

Consider the dark magnesium line at wavelength 4481 Å, which is prominent in the spectra of blue stars such as Sirius. To produce this line the magnesium atom must have lost one electron and must have a second electron already raised to the level next above the lowest possible one. The strength of the line thus depends on the fraction of the magnesium atoms in the star's atmosphere that are singly ionized and on the fraction of these having the second electron raised to the required level.

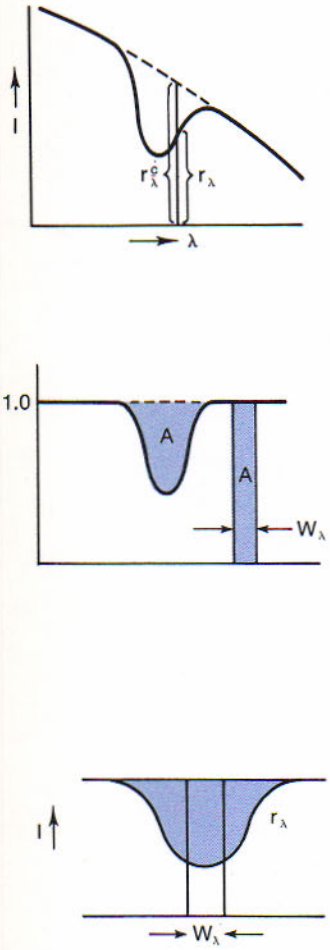
Fig. 12.14 shows how the calculated fractions and the resulting strength of the line vary with temperature. This magnesium line begins to appear at about the sun's temperature, attaining its greatest strength around 10,000°K, and declines at higher temperatures.

We have seen that stellar spectra can be arrayed in a single sequence, except for some branching at the ends. Along the sequence from the red to the blue stars, as we observe again in Fig. 12.15, the patterns of bands and lines gradually change. It is in order to explain the spectrum changes as effects of increasing temperatures of the stellar atmospheres. Later we will consider the effects of the different densities of these atmospheres as well.

12.15 Atomic Abundances

It should be clear from the foregoing that the appearance of certain spectral lines and their strength is a function of temperature and of the number of atoms involved as well. All things being equal, astronomers cover this by assuming local thermodynamic equilibrium, and the relative strengths of lines will then give the ratio of abundance of the various elements. Since it is not a simple task to accomplish this with different strength lines, we resort to two devices: the *equivalent width* and the *curve of growth.*

The absorption lines appear in the continuum of the star which is generally a gradually curving level, sloped one way or the other. If we draw in the level of the continuum across the absorption line we can



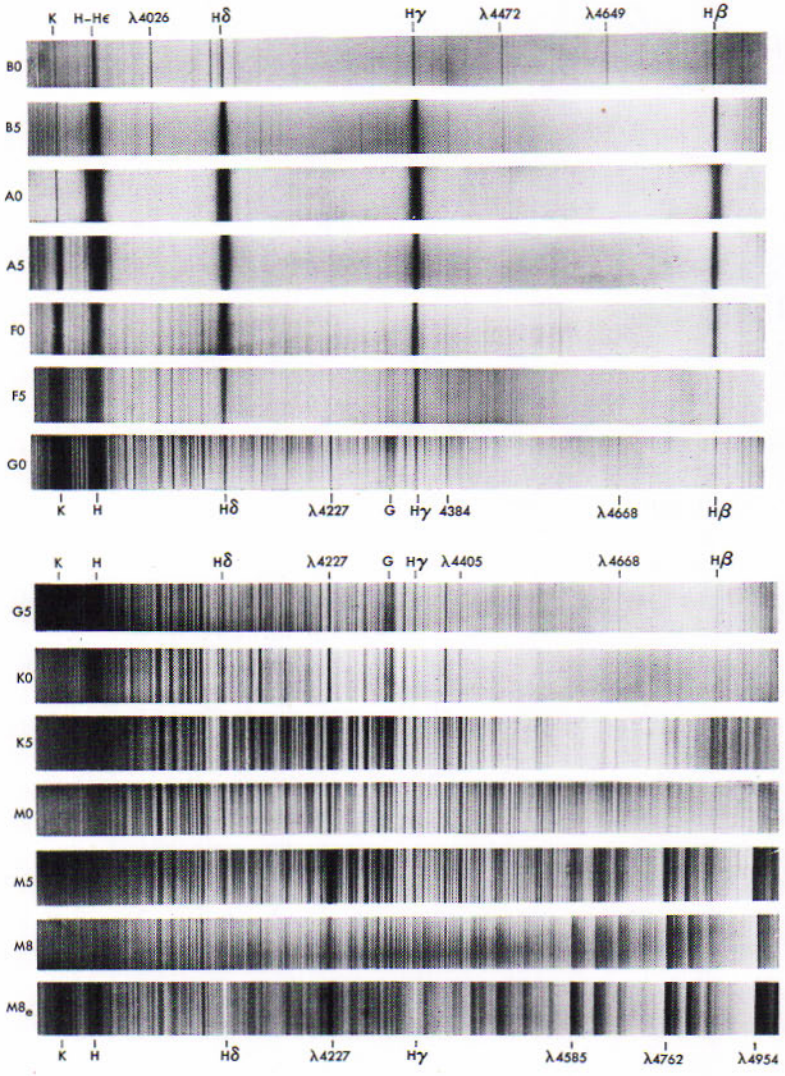
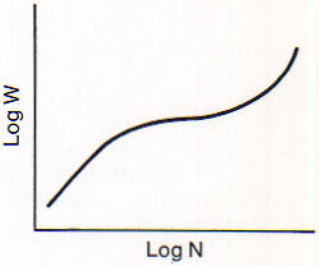


Figure 12.15
Sequence of Stellar Spectra. Types B to M. *(University of Michigan Observatory photographs)*

then form the ratio of points (r_λ) along the profile to the continuum (r_λ^c) at the same wavelength. This procedure normalizes the continuum to unity and displays the undistorted profile of the line. Normally, the line does not go down to zero intensity so we form a rectangular line with zero intensity having the same area (A) as our line. This rectangular line has a width on the wavelength scale and the width is referred to as the line's equivalent width (W_λ) measured in angstrom units.



For any given star the number of atoms involved can be plotted as the abscissa and the equivalent width as the ordinate. This curve is called the curve of growth. In actual practice the procedure is quite delicate and involves calculations well beyond the scope of this book but the principle is straightforward. In this way we arrive at Table 12.IV adapted from work by A. Unsold.

We learn from Table 12.IV that the abundance of a given element does not vary very greatly from object to object. Of course, there are some celestial objects that have abnormal abundances, but in the great majority of cases in the Milky Way and other galaxies there seems to

TABLE 12.IV
Percentage Abundances of the First Twenty Elements

		G2V (SUN)	BoV	PLANETARY NEBULAE
1	H	86.000000	88.12	84.72
2	He	13.700000	11.63	15.08
3	Li	T	—	—
4	Be	—	—	—
5	B	—	—	—
6	C	.000282	.000143	.000424
7	N	.000068	.000260	.000268
8	O	.000509	.000569	.000847
9	F	—	—	T
10	Ne	—	.000842	.000337
11	Na	.000001	—	—
12	Mg	.000026	.000044	—
13	Al	.000002	.000002	—
14	Si	.000031	.000040	—
15	P	T	—	—
16	S	.000014	.000018	.000085
17	Cl	—	—	.000003
18	A	—	.000556	.000007
19	K	T	—	—
20	Ca	.000002	—	—

T indicates trace.

be a common mixture of elements. It seems very probable that the entire mixture originated at the same time very early in the history of the Universe and has not altered very greatly since that time.

12.16 Interpretation of the Spectral Sequence; the Coolest Stars

At the relatively low temperatures of the red stars the spectra show lines of neutral atoms and molecular bands. The lines are particularly those of elements, such as sodium, calcium, and iron, which have easily excited atoms, as indicated by their low ionization potentials (Table 12.III). Hydrogen lines are present despite the higher ionization potential of these atoms from their lower levels, because hydrogen is a very abundant element.

The bands are especially of carbon compounds, titanium oxide, and zirconium oxide; these provide a basis for classifying the red stars. The presence of carbon bands and the absence of titanium oxide bands characterize the spectra of types R and N. Titanium oxide bands are prominent in the spectra of M stars and vanadium oxide bands strengthen for the very coolest ones. Zirconium oxide is conspicuous and titanium oxide is usually absent in the S stars.

The division of red stars into three branches is ascribed to difference of chemical composition. In this view the atmospheres of N stars have more carbon than oxygen; the carbon combines with the available oxygen to form unobservable carbon monoxide and then produces other compounds. The atmospheres of M and S stars contain more oxygen than carbon. After exhausting the carbon, the oxygen has combined with titanium and vanadium or with zirconium, which are believed to have different abundances in these two types of stars.

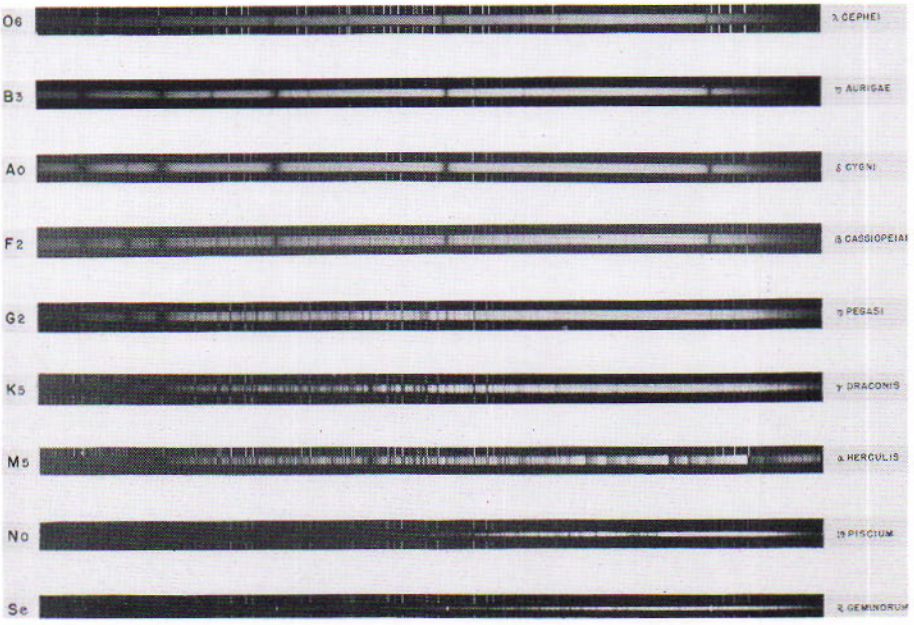


Figure 12.16
Principal types of stellar spectra: O, B, A, F, G, K, M, N, and S. The spectral region shown extends from λ3900 Å (left) to above λ4800 Å (right). *(Photograph from the Hale Observatories)*

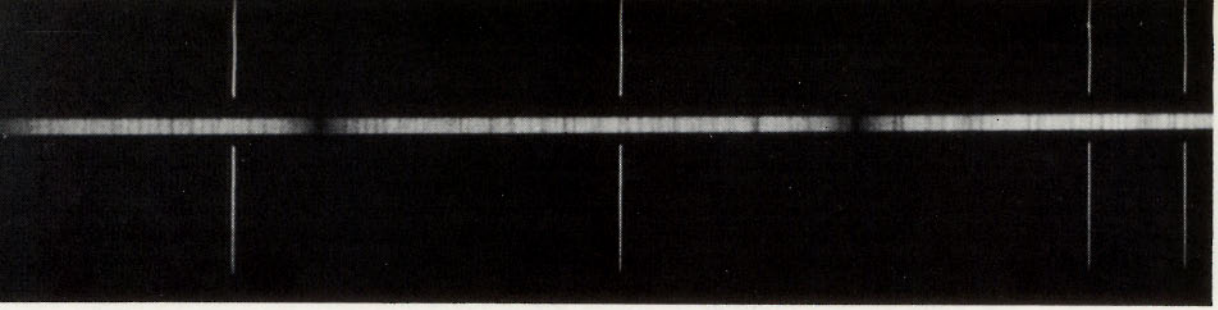


Figure 12.17
The H and K lines of calcium in the spectrum of a late F star.

12.17 Stars of Intermediate Temperature

At the higher temperatures of the yellow stars the compounds are being disrupted and their bands are no longer prominent in the spectra. Titanium oxide bands have disappeared at type K0; but cyanogen (CN), hydrocarbon (CH), hydroxyl (OH), and other combinations are still present at the temperature of the sun.

The Fraunhofer H and K lines (Figure 12.17) of singly ionized calcium dominate the spectra of yellow stars. Here all calcium is singly ionized (Fig. 12.13) and the second ionization is about to begin. Hydrogen lines are becoming stronger as more of these atoms are excited. The complex patterns of neutral metals, such as iron and magnesium, are still conspicuous, but these lines are fading in the hotter F stars. With the removal of an electron the atom produces a different set of lines, as we have seen; and the strong lines of many ionized metals lie in the far ultraviolet that is cut out by the earth's atmosphere. Thus with increasing temperature and degree of ionization in stellar atmospheres the visible spectrum becomes less complex.

12.18 Spectra of the Hottest Stars

The hydrogen lines are most conspicuous in type A2. They decline in still hotter stars as more of the atoms become ionized. Having had only one electron to lose, the singly ionized hydrogen atom cannot absorb light. The lines of this abundant element persist even in type O where only 1 in 100,000 atoms of hydrogen remains neutral, as L. Aller points out. Neutral helium is latest to appear, in type B9; its lines become strongest at B3 and quickly fade, being replaced by lines of ionized helium in O stars.

In addition to hydrogen and helium, the spectra of very hot stars show prominent lines of doubly ionized oxygen, nitrogen, and carbon, a simple pattern visually because most strong lines are in the far ultraviolet. The hottest stars are O5. At the theoretical upper limit, type O0 at a temperature of 100,000°K or more should show no lines at all in the ordinarily observable regions of the spectrum.

The changing patterns along the spectral sequence are therefore caused mainly by changing surface temperatures of the stars. At any stage in the sequence the prominence of a particular set of lines is conditioned by the excitation and ionization potentials of the atoms which produce them. It remains to consider the effect of different densities of stellar atmospheres on the spectra they form.

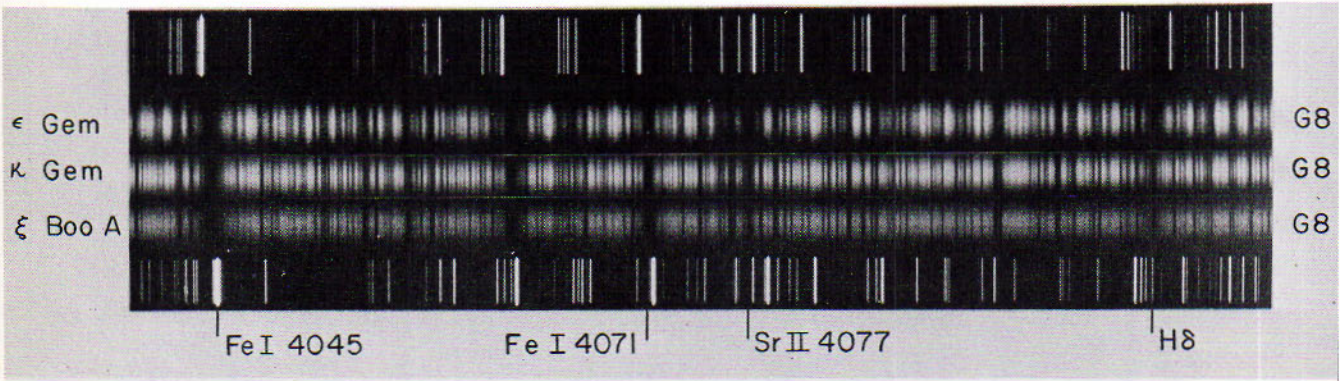
12.19 Spectra of Giant and Main-Sequence Stars

The effective temperatures of giant stars are lower than those of main-sequence stars of the same spectral type (Table 12.II). The reason is given by the theory of thermal ionization. According to this theory the degree of ionization in stellar atmospheres increases with the temperature, and at any specified temperature is greater when the pressure of the gases is low. The atmospheres of giant stars are less dense than those of main-sequence stars. Thus the giants attain a particular degree of ionization and the corresponding type of spectrum at a lower temperature.

Although the spectra of giant and main-sequence stars of the same type are often similar in general appearance, certain lines are stronger in the main-sequence stars and certain other lines are weaker. For example, in Fig. 12.19 the hydrogen lines are stronger in the main-sequence star, while lines of neutral iron are enhanced in the supergiant. Lines due to ionized metals in general are stronger and sharper in supergiants than in main-sequence stars. The reason for the lines due to ionized metals being stronger is straight forward. The atmosphere of a supergiant is less dense than that of the main-sequence star; thus an atom, once ionized, stays that way longer and hence equilibrium is established where more of the atoms are ionized.

Figure 12.19 demonstrates these differences for type G stars. Strontium is a sensitive indicator for the later type stars such as types F, G and K and in the very late stars neutral calcium and the cyanogen molecule furnish the principal differences. These differences have led to the stars being classified by their luminosity within a given spectral type. It is clear that if two stars have basically the same spectral type, i.e. the same temperature, then if one star is much larger than another star the former star must be brighter than the latter star.

Figure 12.19
Spectra of three G8 stars of different luminosity classes. Subtle differences differentiate between supergiants (Ib), giants (III), and dwarfs (V). *(Lick Observatory Photograph)*



12.20 The Yerkes Classification of Stellar Spectra

This classification, done by W. W. Morgan and P. C. Keenan, assigns *luminosity classes* as well as the lettered Draper types. These are numbered in order beginning with the most luminous stars, which have the least dense atmospheres. The numeral I refers to supergiant stars, Ia for the more luminous and Ib for the less luminous supergiants; II refers to bright giants, III to normal giants, IV to subgiants, and V to main-sequence stars. Thus the two-dimensional designation in Table 11.II for Deneb is A2 Ia; Antares, M1 Ib; Epsilon Canis Majoris, B2 II; Aldebaran, K5 III; Procyon, F5 IV; Vega, A0 V.

This classification is clearly based upon the visual region of the electromagnetic spectrum. It is apparent that for very cool thermal sources the "observable" spectrum will fall in the far infrared and/or radio region of the spectrum and that special classifications will be needed for such objects. These classifications are usually based upon the slope of the continuum at specified frequencies and are analogous to using color indexes.

12.21 Spectroscopic Parallaxes

We have seen that more luminous yellow and red stars are cooler than are less luminous stars of the same spectral type. Although this effect is partly compensated by the rare atmospheres of the more luminous stars so that the spectra remain about the same through the type, the lines of certain elements show conspicuous differences of intensity. For example, the line of ionized strontium at 4215 Å becomes stronger and that of neutral calcium at 4277 Å becomes weaker as the stars are more luminous. These and other criteria are the basis of *spectroscopic parallaxes*, the means of determining the distances of stars by examining their spectra.

In Fig. 12.21A the intensity ratio of the sensitive strontium line to a neighboring constant iron line is represented for stars of the same type

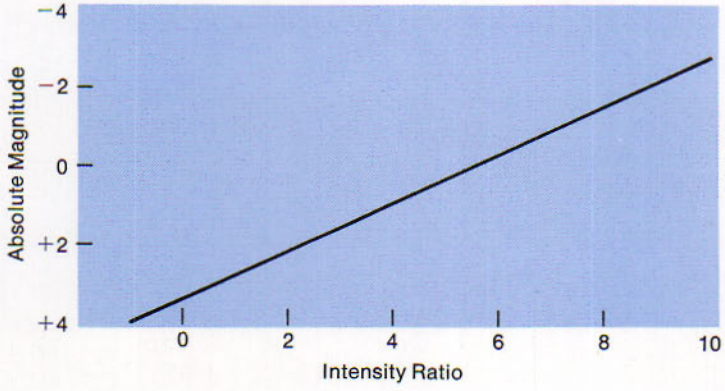
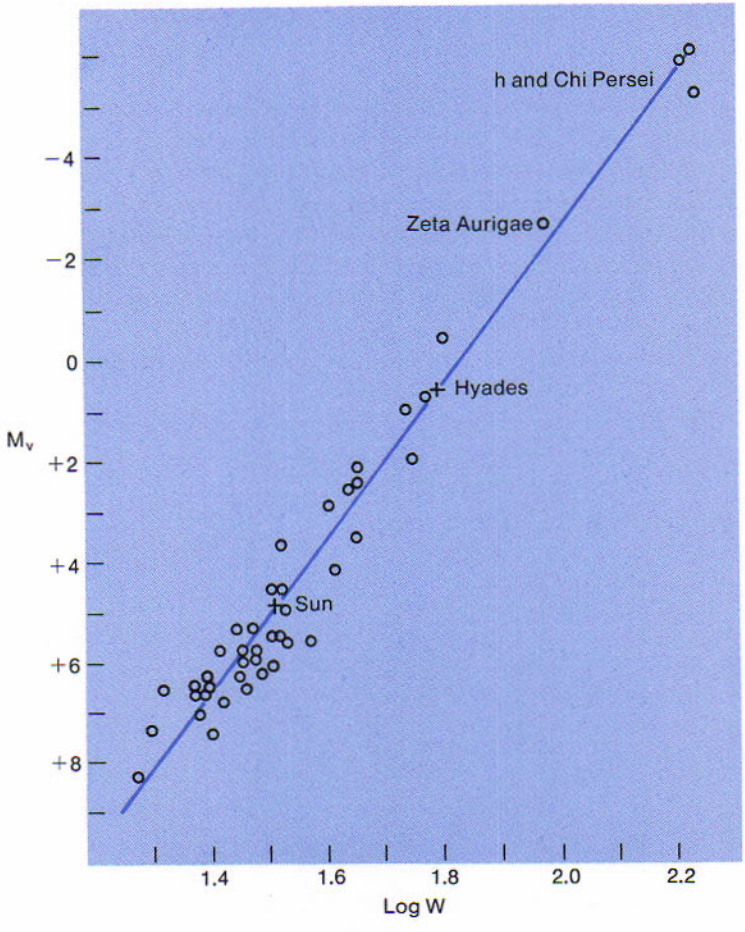


Figure 12.21A
Intensity Ratio of Ionized Strontium Line at 4215 Å Relative to Neutral Iron Line at 4260 Å in Stellar Spectra of the Same Type. *(Diagram by Dorrit Hoffleit, Harvard Observatory)*

and of known absolute magnitude. We note that the two lines are equally intense for a star of absolute magnitude +2.7; the strontium line is 4 times as intense for magnitude +0.7 and 8 times as intense for magnitude −2.0. Whenever such a relation has been established for two lines, the absolute magnitude, M, can be determined from the spectrum of any star of that particular type. The star's parallax, p, can then be found by the formula (11.25): $M = m + 5 + 5 \log p$.

As an example, the strontium line at 4215 Å is twice as intense as the iron line at 4260 Å in the spectrum of a star of the type represented in the diagram. The star's apparent magnitude m, is +7.0 and is not increased numerically by intervening dust. Required the parallax of the star.

Figure 12.21B
Calibration curve for the Wilson-Bappu effect. *(Courtesy of O. C. Wilson)*



From Fig. 12.21A the star's absolute magnitude is +2.0. Thus log $p = (M - m - 5)/5 = (2.0 - 7.0 - 5.0)/5 = -2.0 = 8.0 - 10$. The parallax is 0".01, so that the star's distance is 100 parsecs.

The parallaxes of several thousand stars have been determined by this and other spectral criteria, beginning with the pioneer work of W. S. Adams and A. H. Joy at Mount Wilson. The probable error of a spectroscopic parallax is 15 per cent of its value, whereas that of a direct parallax is around 0".002 regardless of its value. Thus the two methods should be equally reliable for a parallax of 0".013. The direct parallax is likely to be the more dependable for stars nearer than 25 parsecs, and the spectroscopic parallax for more distant stars.

An interesting distance measure akin to the spectroscopic parallax is called the method of Wilson and Bappu and is often referred to as the *Wilson-Bappu effect* after the astronomers who discovered it. It has long been known that the H and K absorption lines often show bright emission cores in the late stars (G, K and M) especially the giants. O. C. Wilson and M. K. V. Bappu studying these stars discovered that the equivalent width of the bright reversals correlated with the star's absolute magnitude (Figure 12.21B). Thus the absolute magnitude of a given star can be obtained by measuring the equivalent width of the bright reversal and then the distance can be derived by using the formula cited above and in section 11.25. As dependable as this method is there is as yet no rigorous physical explanation for it.

EXTENDED ATMOSPHERES AND ENVELOPES

Many hot stars have envelopes so much larger than the photospheres that considerable portions lie outside the cones joining the photospheres to the observer. Much of the light of the surrounding gases is not balanced by their absorption of the radiations from the stars within them. Their bright-line spectra appear superposed on the dark-line spectra of the stars themselves. Extended atmospheres are recognized around type B emission stars, P Cygni stars, and Wolf-Rayet stars by their bright-line spectra. Envelopes around certain novae (13.20) are extensive enough to be visible directly with the telescope, and those of the nearer planetary nebulae are still more conspicuous. Mira and Alpha Herculis are examples of red stars having extended atmospheres.

12.22 Type B Emission Stars

These stars have spectra in which bright lines of hydrogen and sometimes of other elements are present, frequently being superposed on corresponding broader dark lines. Around 4000 such stars are known. About 10 per cent of B0 and B3 stars show emission lines in their spectra and a progressively smaller proportion from B5 to A4 do so. In some cases the bright lines have made their appearance in the spectra of what seemed before to be normal B stars and have subsequently disappeared.

Otto Struve has explained that these emission lines occur generally in the cases of rapidly rotating stars and are widest where the axes of the stars are nearly at right angles to the line of sight. Gases emerge, particularly from the equators where gravity is most diminished by the swift rotations, and form unstable extended atmospheres rotating around the stars. The bright lines in the spectra are broadened by the Doppler effect of the rotation, some of the light coming from parts of the atmosphere that are approaching us and some from parts that are receding from us.

12.23 Pleione, a Shell Star

Narrow dark lines of hydrogen and other elements make their appearance in the spectra of some blue emission stars in addition to the bright lines. The narrow lines then present a striking pattern far into the ultraviolet. Examples of such stars, designated as *shell stars,* are Pleione, 48 Librae, and Gamma Cassiopeiae.

Pleione is one of the brighter members of the Pleiades cluster. Bright hydrogen lines, which had been observed in its spectrum since 1888, disappeared in 1905, so that the spectrum resembled that of an ordinary B5 star. The reappearance of broad bright lines and presently of narrow dark lines as well was reported in 1938. Both sets of lines grew stronger until 1946 and weakened thereafter, until in 1951 they were almost gone.

After the long period of earlier stability, there began a slight but persistent acceleration of the atoms outward from the star's normal atmosphere. The acceleration of the streams of atoms increased until the outward velocities were 50 km/sec or more, becoming greater for the ultraviolet lines of the Balmer series of hydrogen than for those in the visible region. The first result was an emitting extended atmosphere; the second was the absorbing shell at a distance from the photosphere which has been estimated as 2 or 3 times the radius of the star. Eventually the supply of atoms failed and the shell blew away.

Some stars such as Gamma Cassiopeiae have had temporary shells of shorter duration. Others have shells which are more stable than that of Pleione, and their dark lines show the same velocity as from the star within them. In all cases these lines are narrow because they are absorbed at high levels where the density is low, and they are not broadened by the rotation.

Gamma Cassiopeiae ejected a temporary shell in 1934. After the shell

dissipated its spectrum looked like the prototype of a B2 III star. Sometime in 1964 or 1965 it began ejecting a new shell. Spectrograms taken in the fall of 1965 showed a flat filled in spectrum which changed to the characteristic broad emission lines in 1966 and increased in intensity.

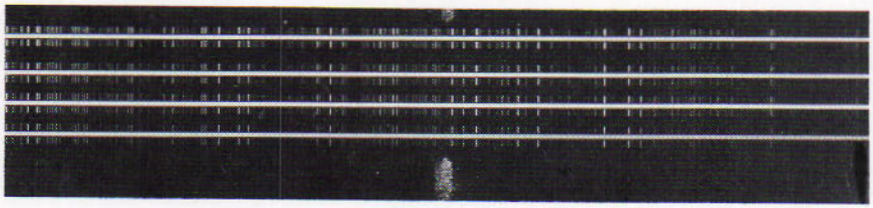


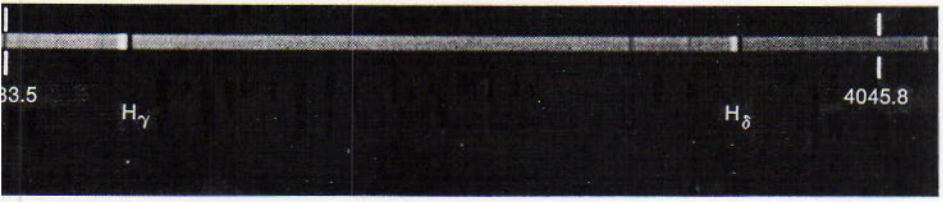
Figure 12.23
Spectrogram of Gamma Cassiopeiae, which normally appears as a typical B star. Here its shell is filling in all lines, leaving the spectrum completely featureless. *(Leander McCormick Observatory Photograph by H. J. Wood)*

12.24 The P Cygni Stars

These comprise a group of blue emission stars showing features quite different from the ones already mentioned. Like the others, these stars are mainly type B and a few are early A; but their bright lines have dark lines at their violet edges rather than near their centers. Hydrogen lines are the most prominent. The prototype P Cygni burst out like a nova twice in the 17th century and thereafter became invariable at the 5th apparent magnitude.

The picture here, according to P. W. Merrill, is that of gases passing out through a shell which itself does not expand. The bright lines in the spectrum are somewhat widened, some parts coming from the gas on our side of the star, which is approaching us, and other parts from the gas beyond the star, which is receding from us. The dark lines displaced to the violet are absorbed by the part of the envelope immediately in front of the star, where the gas has the maximum speed of approach.

Figure 12.24
Spectrum of P. Cygni. *(Leander McCormick Observatory Photograph by P. Giguere)*



12.25 Wolf-Rayet Stars

These stars are named after two astronomers at the Paris Observatory who discovered, in 1867, the first 3 known stars of this remarkable type. About 200 are now recognized; the brightest is the 2nd-magnitude southern star Gamma Velorum. Formerly included with the dark-line stars of type O, these stars are now grouped by themselves in type W. Of average absolute magnitude −5, they are among the hottest stars, having

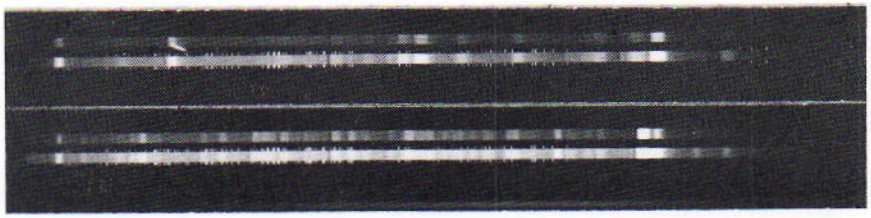


Figure 12.25
Spectra of typical Wolf-Rayet stars. Long and short exposures of HD192163 (top) and HD192103 (bottom) taken at the Lick Observatory. *(Photographs courtesy of Lindsey Smith)*

surface temperatures of 50,000°K. Averaging twice the sun's diameter, they are surrounded by much larger atmospheres.

The spectra contain broad bright lines on a much fainter continuous background and sometimes bordered by weak and diffuse dark lines at their violet edges. The prominent lines are mainly of helium, oxygen, silicon, nitrogen, and carbon in various stages of ionization; but where nitrogen is conspicuous, carbon is practically absent, and vice versa. Hydrogen lines are surprisingly weak.

One interpretation of the Wolf-Rayet stars pictures a star surrounded by a gaseous envelope where the material might seem to be streaming outward as fast as 3000 km/sec, as indicated by the widths of the bright lines. An alternate interpretation is that the lines may be widened mainly by random motions in the envelope. A number of these stars are spectroscopic binaries and a few are known to be eclipsing stars. It is suspected that all may be components of close pairs, the W stars revolving with larger type O companions, a situation that may provide an important clue to the still mysterious behavior of Wolf-Rayet stars.

12.26 Planetary Nebulae

Planetary nebulae are so named because the nearer ones appear with the telescope as greenish disks somewhat remindful of the disks of Uranus and Neptune. They are gaseous envelopes around central stars, having diameters 20,000 to more than 100,000 times the earth's distance from the sun. Planetary nebulae are of three general types, as O. C. Wilson describes them: (1) Ring-like, often having a bright inner ring and a fainter outer ring; (2) amorphous, consisting of "roundish blobs"; (3) irregular.

Several hundred planetary nebulae are recognized, including 86 new ones found by A. G. Wilson and G. O. Abell in the Palomar Sky Survey. All invisible to the naked eye, they appear with the telescope in a variety of sizes. They range in this respect from the relatively nearby ring-like NGC 7293 in Aquarius (see Plate X), having half the apparent diameter of the moon, to objects so reduced by distance that they can be distinguished from stars only by their peculiar spectra. Many faint planetaries were discovered by R. Minkowski in a systematic search for objects showing the red line of hydrogen and very little continuous spectrum, and by K. G. Henize by the same procedure in the southern hemisphere.

The Ring nebula in Lyra (PlateVII) is among the most familiar of the planetaries. Unlike some other ring-like planetaries, it is a true ring, as R.

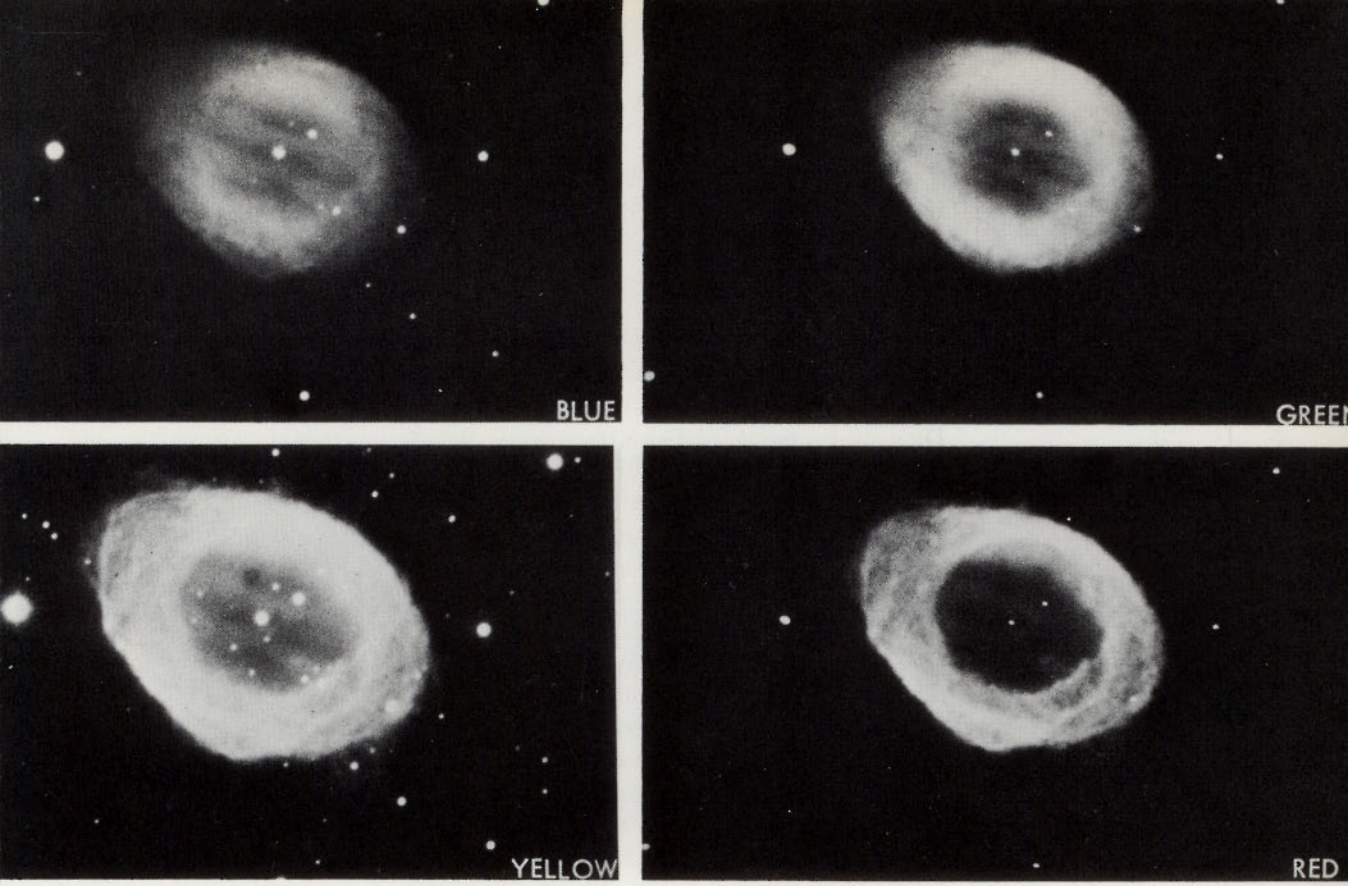


Figure 12.26A
The "Ring" nebula in Lyra, photographed in blue (t.l.), green (t.r.), yellow (b.l.) and red (b.r.) lights. See also Plate **VII** *(Photograph from the Hale Observatories)*

Figure 12.26B
The "Owl" nebula, NGC 3587, in Ursa Major, photographed on February 9, 1910. *(Photograph from the Hale Observatories)*

Minkowski and D. E. Osterbrock have demonstrated. This ring is 20 times as bright as the space inside it, whereas a ratio of 2 would be predicted if the nebula were actually a hollow ellipsoidal shell. An example of a true ring seen edge-on might well be NGC 650 and 651, cataloged as a double nebula. The meaning of the NGC numbers is explained later (15.1).

The central stars of planetary nebulae are about as massive as the sun but are much smaller than the sun, so that they have high densities. They are among the hottest stars, having surface temperatures of 50,000°K, or more. Thus they furnish a rich supply of ultraviolet radiation to illuminate the nebulae. Because their radiation is mainly in the ultraviolet, these stars are less easy to see than are the nebulae, but they come out clearly in the blue photographs.

12.27 Expansion of Planetary Nebulae

When the image of a planetary nebula is centered on the slit of the spectroscope, the bright lines of the spectrum tend to be double in the middle. The part of the line formed in the near side of the nebula is displaced toward the violet, showing that the gas is approaching us; and the part from the far side is displaced toward the red, showing that the gas is there receding from us. Thus the planetary nebulae are moving outward from the central stars. The speeds of expansion in the radii range from 10 to 50 km/sec and may have decelerated moderately since the outburst of the material from the central stars. After lifetimes of 20,000 years, the nebulae begin to disintegrate by breaking into separate clouds of gas.

It will be noted later (13.20) that planetary nebulae differ from the expanding envelopes of novae in their slower rates of expansion and their much longer lives. Unlike supernovae, planetary nebulae are not radio sources or, perhaps more correctly, have not been detected by radio means. There is also a pronounced difference in the amount of material involved; the mass of a planetary nebula is about one tenth of the star's mass, whereas the mass of a nova envelope does not exceed 1/10,000 of the mass of the star.

THE INTERIORS OF THE STARS

The study of stellar interiors is the paragon of analytical reasoning and the application of nuclear physics. Nuclear reactions, specifically the conversion of hydrogen to helium, are the source of the enormous energy released by stars.

12.28 A Star in Equilibrium

Studies of the interior of a main-sequence star such as the sun proceed on the assumption that the star is in mechanical equilibrium under its own gravity. By this we mean that the weight of the gas above any level in the interior is exactly supported by gas and radiation pressures at that level. The assumption is reasonable, because the star would shrink if the internal pressure were inadequate or else expand if it were excessive; and the sun is not observed to be doing either at present.

The weight of overlying gas to be supported at a particular level depends on the mass of that gas and the acceleration of gravity. These values can be derived for a star of known mass and radius if it is also known how the density of the material diminishes with distance from the center of the star. The pressure required to balance the weight depends on the temperature and composition of the gas at that level. If the composition is known, then it becomes possible to calculate the temperature there and, similarly, at all other levels in the star's interior.

In order to preserve the balance between weight and pressure, the rate at which the energy is liberated in the central region of the star and is passed on up to the surface must remain equal to the rate of radiation at the surface. The required rate of energy liberation is determined by the luminosity of the star. If this rate diminishes, the star will cool and contract; if it increases, the star will grow hotter and expand. One important objective of studies of stellar interiors is to learn the process by which the energy is liberated and how it is likely to affect the future of the star.

12.29 Gas and Radiation Pressures

The *gas pressure* at any point inside a star is produced by the turmoil of the gas particles in that vicinity. The relation between the pressure, p, the density, ρ, and the absolute temperature, T, of the gas is given by the gas law: $p = R\rho T/\mu$, where R is a constant and μ is the mean weight of the gas particles. These particles are constituents of shattered atoms. In the laboratory a gas compressed to a density exceeding one tenth the density of water ceases to conform to this law; but in the very hot interior of a star, where atomic superstructures are disrupted, the law continues to operate at densities far exceeding that of water.

The *radiation pressure* results from the outward flow of radiation through the star. Radiation pressure at any point is directly proportional to the rate of radiation there, which by Stefan's law (10.4) varies as the fourth power of the temperature. Thus with increasing temperature the ratio of radiation to gas pressure increases. The share of radiation pressure in keeping the star inflated, however, is considered negligible where the star's mass does not exceed 2 or 3 times the mass of the sun.

$P_{rad} \propto B \propto T^4$

12.30 Effect of Chemical Composition

The atoms in the deep interiors of the stars are highly ionized. The result is a confusion of atomic nuclei, partly or entirely separated from their electrons, and the electrons themselves. For example, a neutral atom of iron (atomic weight 55.8 and number 26) would become 27 particles of average weight 55.8/27, or 2.1, if entirely ionized. The average values for the "metals," meaning here all the elements except hydrogen and helium, are so nearly the same that the proportions in which they occur are not important in the gas law. What has to be determined is the proportion between the metals and the lightest gases, hydrogen and helium, whose particles have average weights of only 0.5 and 1.3, respectively.

Consider two stars having the same size, density, and density distribution, the first star composed mainly of iron and the second entirely of hydrogen. By the formula the temperature at a particular point in the first star would be 2.1/0.5, or about 4, times that at the corresponding point in the second star in order to produce the same gas pressure. If the sun were composed mainly of metallic gases, its central temperature would be 40 million degrees K. If it contained only hydrogen, the central temperature would be 10^7 degrees and it would shine only 1 per cent as brightly as the metallic sun.

Thus chemical composition is an important factor in determining the interior temperatures and also the luminosity of a star. The procedure has been to find by trial and error a composition giving a central temperature appropriate to the observed luminosity of the star. Calculations of this sort have recently assigned a smaller percentage of metals, and accordingly a lower central temperature, than they did previously.

12.31 The Interior of the Sun

In the sun, where the liberation of energy may not yet have produced pronounced changes, the chemical elements are believed to be in about the same proportions from the chromosphere almost to the center. Hydrogen contributes nearly 75 per cent of the mass, helium about 25 per cent, and the metals 1 per cent or so in the models recently discussed by B. Strömgren. The temperature increases rapidly from the surface to a central value of 13 million degrees K. The gases around the center are 90 times as dense as water. M. Schwarzschild has employed a more recent model of the sun by R. Weymann, where the hydrogen content at the center is reduced to 50 per cent. The central temperature is thereby increased to 15 million degrees and the density to 134.

Energy is liberated in the deep interior of the sun in the form of very high-frequency radiation. It is passed on upward mainly by absorption and emission by the gas particles, until it reaches the surface and escapes into space. In this process the frequency is gradually stepped down, so that much of the energy emerges as visible sunlight.

Figure 12.31
A large modern digital computer. Such computers are used to calculate models of stellar interiors. *(Photograph of the B6500 computer system, courtesy Burroughs Corporation)*

12.32 What Keeps the Stars Shining?

A major problem of astronomy has been to locate the stores of energy which supply the radiations of the stars during their long lives. The stars cannot continue to shine simply because they are extremely hot; energy must come from some source to keep them so highly heated. Nor are the stars kept heated by combustion. Even in the cooler regions of the sun's atmosphere the atoms are generally not combining. The sun is too hot to burn.

We dismiss any idea that adequate amounts of energy are being stoked into the sun from outside. Meteors must fall into the sun in great numbers, it is true, and their impacts are great as they arrive with speeds approaching 600 km/sec. A yearly fall of meteoric material 3500 times the earth's mass would be needed to supply as much energy as the sun radiates in a year. The actual amount falling into the sun can be only a very small fraction of this requirement.

Another early idea of how the sun and other stars keep shining seemed more promising. The idea was that all the necessary heat may be supplied by their contraction.

12.33 The Contraction Theory

In 1854, the German physicist H. L. F. von Helmholtz explained that a yearly contraction of 43 meters in the sun's present radius would supply enough heat to keep the sun shining. The process could not continue indefinitely. After a few million years, as it then seemed, the sun would become so compressed that it could contract no further; then the sun would quickly cool, bringing to us the "end of the world."

The theory that the sun shines only because of its contraction does not conform with the present cosmic time-scale. To have continued shining as it does today as the result of contraction alone, the sun must have shrunk to its present size in less than 50 million years. That duration of the sunshine may have seemed quite long enough a century ago, but it is far too short today when geologists are dating the beginning of life on the earth at least 10^9 years in the past.

Thus contraction is not the only factor in the continued shining of the sun and stars. In current theories of stellar evolution it enters as a means of heating youthful stars to the point where their internal supplies of energy can be released; it assists in important ways thereafter and ultimately assumes control when these supplies are exhausted.

Having failed to find adequate supplies of energy elsewhere, the scientists now look more confidently within the atoms. The inquiry is guided by a relation derived from the theory of relativity.

12.34 Relation Between Energy and Mass

$E = mc^2$

Early in the present century, Albert Einstein showed that energy and mass are related by the formula: Energy equals mass times the square of the speed of light. The energy is expressed in ergs, the mass in grams, and the speed of light in centimeters a second. By this formula there is a vast amount of energy in a very small quantity of matter. For example, the energy released by the destruction of a single gram of matter could lift a weight of 60 million kg from the earth's surface to the height of 2 km above it.

We now approach the answer to the question: What keeps the stars shining? In the present view, the energy required to keep the stars shining is converted from the excess material when atoms of lighter chemical elements are built up into atoms of heavier ones in the stellar interiors.

12.35 Fusion of Hydrogen into Helium

The great abundance of hydrogen in the stars has directed the search for processes which can transform hydrogen into heavier elements. At the central temperature of the sun the collisions of gas particles are vigorous enough to unite hydrogen nuclei into helium nuclei. This transformation

can operate on a scale sufficient to keep the sun shining at the present rate. The arithmetic of the operation is as follows:

The relative weight of the hydrogen nucleus is 1.0076 and that of the helium nucleus is about 4.003. Where hydrogen in stellar interiors is being converted to helium, each combination of 4 hydrogen nuclei into a single helium nucleus involves a mass loss of about $4.030 - 4.003 = 0.027$, or seven tenths of 1 per cent of the original mass. The unrecovered mass is released as energy.

Two means of transforming hydrogen into helium in the stars have received much attention. They are the proton-proton reaction and the carbon cycle. The first of these is considered the more effective at the temperatures in the sun and the redder stars of the main sequence. The carbon cycle becomes effective at somewhat higher central temperatures.

12.36
Proton-Proton Reaction

This reaction combines 6 hydrogen nuclei to form a helium nucleus and at the end puts 2 hydrogen nuclei back into circulation. One process of this type, proposed by W. A. Fowler and C. C. Lauritzen, is represented by the following succession of formulas, in which the subscripts are the atomic numbers and the superscripts are the atomic weights to the nearest whole number. The symbol β^+ denotes a positron, or positive electron, and γ denotes a γ ray, or a unit of high-frequency radiation. The steps are:

$$ {}_1H^1 + {}_1H^1 \rightarrow {}_1H^2 + \beta^+ \quad (1) $$

$$ {}_1H^2 + {}_1H^1 \rightarrow {}_2He^3 + \gamma \quad (2) $$

$$ {}_2He^3 + {}_2He^3 \rightarrow {}_2He^4 + 2{}_1H^1 \quad (3) $$

In this reaction: (1) Two protons, or normal hydrogen nuclei, combine to form a deuteron. One proton becomes a neutron and a positron, which soon unites with a negative electron in the vicinity to add to the energy. A neutrino formed in this step promptly escapes from the scene. (2) The deuteron combines with a third proton to form helium of weight 3, with the release of a γ ray. (3) Two helium-3 isotopes combine to produce an ordinary helium nucleus and 2 protons. The available energy released in the formation of each helium nucleus equals about 4×10^{-5} ergs.

The fusion of hydrogen into helium, which provides the energy of the hydrogen bomb, may become the principal source of power on the earth as well as in the stars. Although the quantity of deuterium in a liter of ordinary water is very small, it has an energy content equivalent to that of 350 liters of gasoline. The amount of deuterium in the oceans is enough to supply all foreseeable demands for power in the world for billions of years to come if a method of controlling the fusion reaction can be discovered.

12.37 The Carbon Cycle

This process, known as the carbon cycle because carbon is its promoter, was suggested, in 1938, by the physicist H. A. Bethe, who also proposed the chain of the proton-proton reaction in a form somewhat different from the one given in the preceding section. The cycle is:

$$
\begin{aligned}
{}_6C^{12} + {}_1H^1 &\rightarrow {}_7N^{13} + \gamma && (1)\\
{}_7N^{13} &\rightarrow {}_6C^{13} + \beta^+ && (2)\\
{}_6C^{13} + {}_1H^1 &\rightarrow {}_7N^{14} + \gamma && (3)\\
{}_7N^{14} + {}_1H^1 &\rightarrow {}_8O^{15} + \gamma && (4)\\
{}_8O^{15} &\rightarrow {}_7N^{15} + \beta^+ && (5)\\
{}_7N^{15} + {}_1H^1 &\rightarrow {}_6C^{12} + {}_2He^4 && (6)
\end{aligned}
$$

In the turmoil of the star's interior: (1) A carbon nucleus of weight 12 combines with a proton to form radioactive nitrogen, with the release of a γ ray. (2) The nitrogen degenerates into carbon of weight 13 and a positron. (3) The carbon combines with a second proton and forms ordinary nitrogen, with the release of a γ ray. (4) This nitrogen combines with a third proton to form radioactive oxygen with the release of a γ ray. (5) The oxygen degenerates into nitrogen of weight 15 and a positron. Here and in the second step a neutrino is released. (6) The heavy nitrogen combines with a fourth proton to produce the original carbon and a helium nucleus.

In this cycle 4 hydrogen nuclei unite to form a helium nucleus, and the excess mass is released as energy. The carbon is recovered and can be used repeatedly. The energy made available in the formation of a single helium nucleus is about the same as in the former process.

12.38 The Unmixed Model of a Star

Consider a main-sequence star, such as the sun, which contains at the outset the same mixture of chemical elements throughout and where conversion of hydrogen to helium has begun in the interior. Vigorous stirring is not to be expected, except near the photosphere and in the central core where the conversion is occurring.

When the hydrogen in the core is nearly exhausted, the fusion into helium spreads to surrounding regions. The core contracts, until it may become hot enough for the transformation of helium into carbon, oxygen, magnesium, and still heavier elements. These processes are considered in Chapter 16.

REVIEW QUESTIONS

1. What term is applied to an atom possessing all of its electrons but all of which are not at their lowest permitted levels?
2. What term is used to describe an atom that has lost one or more of its electrons?

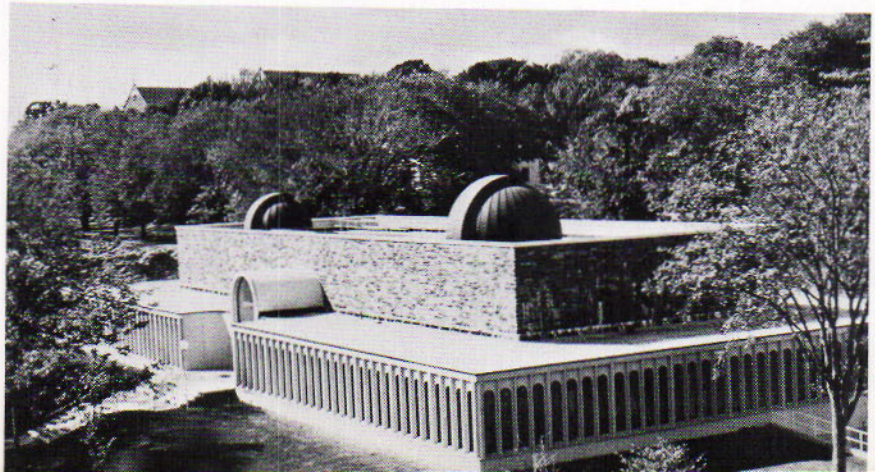
3. What does the symbol HII imply?
4. What are "forbidden transitions"?
5. Suppose that an X-ray source is imbedded in a cloud of hydrogen that is "optically thick" to X-radiation, i.e. no X-ray photon can escape the cloud, what emissions (or absorptions) would you expect to see?
6. The principal series of the sodium atom, analogous to the Lyman series of hydrogen, has a pair of lines for each transition. The wavelengths of the first excited state are 5890 Å and 5896 Å, the former being much stronger than the latter. What can we infer from this information?
7. Define the terms "equivalent width" and "curve of growth".
8. Why is it that we do not see molecules in the spectrum of Sirius?
9. Dynamical considerations lead astronomers to estimate that as much as 50% of the mass of our Milky Way Galaxy is thus far unidentified. Can mass from stars, planetary nebulae, and the like explain this? Discuss this in some detail.
10. Why does radiation pressure have more importance in an F-type star than in a K-type star?
11. What were the fallacies of the Contraction Theory?
12. Discuss the current energy producing reactions working in the stars.

FOR FURTHER READING

Glasstone, S., *Sourcebook on the Space Sciences,* Princeton: D. Van Nostrand Co., 1965.
Semat, Henry, and H. E. White, *Atomic Age Physics,* Rinehart & Co., New York, 1959.
Struve, O. and V. Zebergs, *Astronomy of the 20th Century,* New York: The Macmillan Co., 1962.

REFERENCES

Flugge, S., ed., *Handbuch der Physik,* vols. 50–54, [Astrophysics], Springer-Verlag, Berlin.
Greenstein, Jesse, L., ed., *Stellar Atmospheres,* Chicago: Univ. of Chicago Press, 1960.
Herzberg, Gerhard, *Atomic Spectra and Molecular Structure,* Princeton: D. Van Nostrand Co., 1950.
Swihart, T. L., *Astrophysics and Stellar Astronomy,* New York: John Wiley & Sons, 1968.



Princeton University Observatory, Peyton Hall. *(Princeton University Photograph)*


13

INTRINSICALLY VARIABLE STARS

OBSERVATION AND EVALUATION — PULSATING STARS — RED VARIABLE STARS — ERUPTIVE STARS

Variable stars are stars that vary in brightness and frequently in other respects as well. The second edition (1958) of the *General Catalogue of Variable Stars* lists 14,708 variables in our galaxy. All variable stars, according to the *Catalogue,* may be divided into three main classes: eclipsing, pulsating, and eruptive variables, each of which is subdivided into several types. We consider here stars that are intrinsically variable, from causes inherent in the stars themselves, leaving the eclipsing stars for the following chapter on binary stars.

Nova Hercules on March 10, 1934. *(Lick Observatory Photograph)*

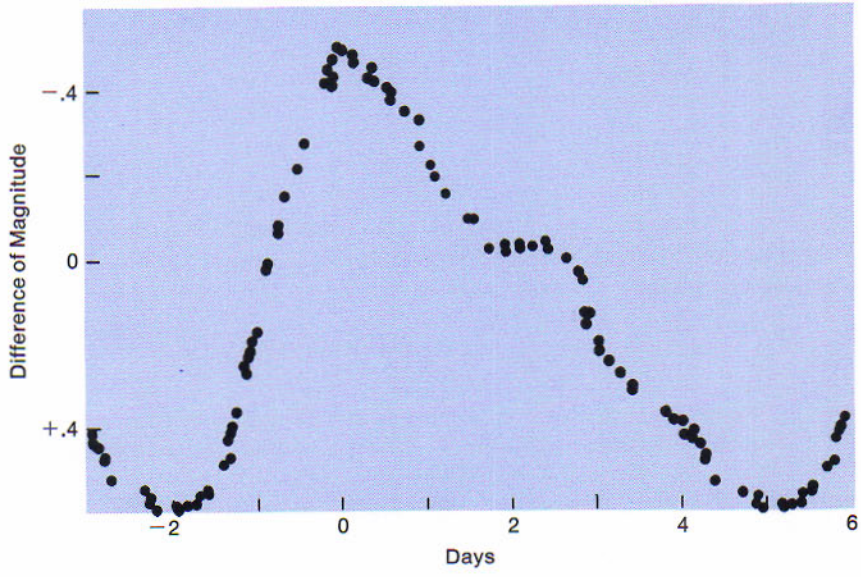
OBSERVATION AND EVALUATION

13.1 The Light Curve

The curve representing the array of points where the observed magnitudes of the star are plotted against the times of their observations is called a *light curve.* If the same variation is repeated periodically, the times of successive maximum and minimum brightness can be derived eventually, and from these are found the *elements* of the light variation. They are the *epoch,* or time of a well-defined place on the curve, and the *period* of the variation. The light curve for a single cycle may then be obtained more precisely by plotting all the observed magnitudes with respect to *phase,* or interval of time either in days or in fractions of the period since the epoch preceding each observation.

As an example, the elements for the cepheid variable star Eta Aquilae are: Maximum brightness = 2,414,827.15 + $7^d.1767 \cdot E$, where the first number is the epoch expressed in Julian days and the second is the period in days. In order to predict the times of future maxima we have simply to multiply the period by E = 1, 2, . . . and to add the results successively to the original epoch.

Figure 13.1
Light curve of Eta Aquilae, a classical cepheid with a period of about seven days *(Photoelectric light curve by C. C. Wylie)*



The *Julian Day* (3.23) is the number of days that have elapsed since the beginning of the arbitrary zero day at noon Greenwich mean time on January 1, 4713 B.C. It is a device often used in astronomical records to avoid the complexity of the calendar system. When the interval between two events is required, especially where they are widely separated in time, it is easier to take the difference between the Julian dates.

13.2 The Observation of Variable Stars

Once a star is discovered, it is given a designation that follows a plan devised before the use of the most powerful telescopes and the resulting discovery of many thousands of stars. This explains why the plan, originally a simple one, has become increasingly complex. It works as follows: Unless the star already has a letter in the Bayer system (1.24), it is assigned a capital letter, or two, in the order in which its variability is recognized, followed by the possessive of the Latin name of its constellation. For each constellation the letters are used in the order: R, S, . . . , Z; RR, RS, . . . , RZ; SS, . . . , SZ; and so on until ZZ is reached. Subsequent variables are AA, AB, . . . , AZ; BB, . . . , BZ; etc. By the time QZ is reached (the letter J is not employed), 334 variable stars are so named in the constellation. Examples are R Leonis, SZ Herculis, and AC Cygni. Following QZ the designations are V 335, V 336, and so on; an example is V 335 Sagittarii.

The next step is to recognize the character of the star's variability and to decide whether its continued study is likely to contribute to knowledge of such stars or of stars in general. The magnitudes at the various phases of its fluctuation are determined by comparison with neighboring stars of known magnitudes.

Studies requiring the highest precision are made with the photoelectric cell and may be extended by use of filters for transmitting different colors of the starlight to the cell. For many purposes the magnitudes are determined photographically, either measured by appropriate means or simply estimated in photographs with reference to sequences of known magnitudes. The photometric determinations are often supplemented by studies of the star's spectrum, where the character and displacements of the lines may give additional information.

Visual observations of the magnitudes are employed as well, especially where the light variation is of large range and somewhat irregular. Amateur astronomers with small telescopes find here an interesting and useful field and can make very valuable contributions to the science. The American Association of Variable Star Observers (AAVSO), of which Margaret W. Mayall is the director, has contributed many more than a million magnitude determinations.

PULSATING STARS

Many supergiant and giant stars are variable in brightness because they are alternately contracting and expanding, becoming hotter and cooler in turn. These pulsating stars include prominently the cepheid and RR Lyrae variables.

13.3 Cepheid Variable Stars

The name of one of the earliest recognized variable star, Delta Cephei, is affixed to a class of variables called cepheid variable stars. They are of two types: classical cepheids and type II cepheids.

Classical cepheids are so called because they resemble the prototype. About 625 are known in our galaxy, where they congregate near the central line of the Milky Way. Their periods range generally from 1 to 50 days and are most commonly around 5 days. Those of shorter periods are generally steady in period and form of light curve. The increase in brightness is likely to be more rapid than the decline, and the maximum of the light curve is often more sharply defined than is the minimum. The visual range of the variation is frequently around one magnitude.

Classical cepheids are yellow supergiants, not redder than type G0 at their maxima. Very rare in space, their high luminosities raise them to prominence in the sky out of proportion to their actual number. About a dozen are visible to the naked eye; the brightest are Polaris, Delta Cephei, Eta Aquilae, Zeta Geminorum, and Beta Doradus. Polaris has the smallest range of all, which is 0.17 magnitude in the ultraviolet and 0.04 in the infrared.

Type II cepheids have been recognized more frequently in the globular clusters and near the center of our galaxy. Their light curves have broader maxima and are more nearly symmetrical than those of the classical cepheids. Their spectra often show discontinuities in the hydrogen lines as do the RR Lyrae stars. Their periods are mostly from 12 to 20 days. An example is W Virginis.

13.4 RR Lyrae Variable Stars

The RR Lyrae variable stars are named after RR Lyrae. These were first observed in the globular clusters (15.12). They are often called *cluster variables,* although they are now recognized in greater numbers outside the clusters. The periods of their light variations are around half a day, ranging from 1½ hours to about a day, and are sometimes slowly changing. Especially in the clusters the light curves are nearly symmetrical for the shorter periods; at periods of about half a day they change abruptly to curves having very steep upslopes and extreme amplitudes, effects which moderate as the periods are longer. Variations in magnitude are generally

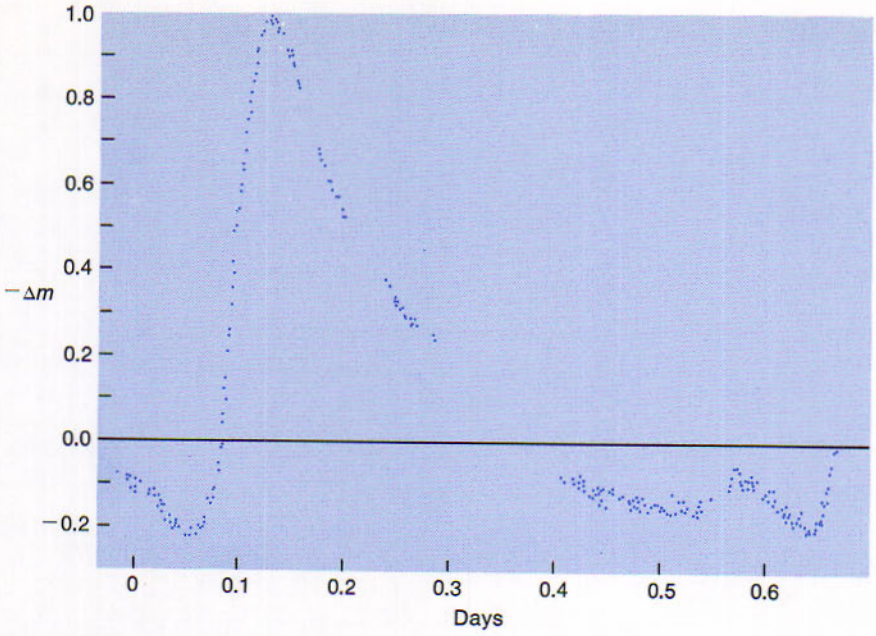


Figure 13.4
The light curve of RR Lyra according to Walraven. The rapid rise to maximum light is characteristic of these variables. *(Reproduced by permission from* Astronomische Bulletin Nederlands*)*

less in ultraviolet than in blue light, because of increased absorption by the Balmer continuum (12.7) as the stars rise to their higher temperatures.

These stars are blue giants, which vary between types A0 and F5. They occupy by themselves a small section of the horizontal giant branch in the color-magnitude diagram. Although they far outnumber the classical cepheids, their lower luminosity makes them less conspicuous in the sky. Not one is bright enough to be visible to the naked eye; the brightest examples are the prototype, RR Lyrae, and VZ Cancri, both of the 7th apparent magnitude.

13.5 The Spectra of Cepheids and RR Lyrae Stars

Two features which are especially significant for the interpretation of these stars are evident in their spectra.

1. *The spectral type is variable.* From maximum to minimum light the spectral type advances. In the case of Delta Cephei the change is from F4 to G2, signifying a drop of about 1500°C in surface temperature. Thus the surface of this star is hotter at maximum brightness and cooler at minimum.
2. *The spectrum lines oscillate* in the period of the light variation. The curve representing the variation of radial velocity with time is not

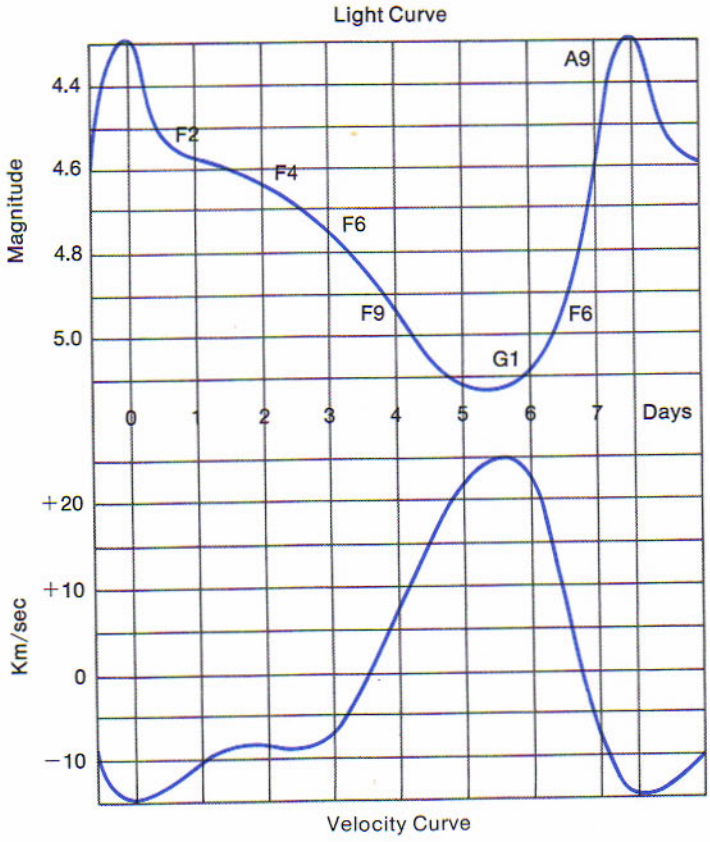


Figure 13.5
Light and velocity curves of W Sagittarii. Maximum brightness of this classical cepheid occurs at about the time of greatest negative velocity (of approach); minimum light at about greatest positive velocity (of recession). *(Diagram by R. H. Curtiss)*

far from the mirror image of the light curve. Near maximum brightness of the star the lines are displaced farthest toward the violet, and near minimum brightness they are displaced farthest toward the red end of the spectrum. This is the Doppler displacement owing to the motion in the line of sight of the atmosphere in front of the star. At maximum light these gases are approaching us and at minimum light they are receding from us.

The correspondence between the velocity and light curves is only approximate. The differences between the two depend on the part of the spectrum that is examined; also with increasing period of the variation the greatest velocity of approach tends to lag behind the light maximum, as A. H. Joy has shown.

These oscillations of the spectrum lines seemed in earlier times to signify the mutual revolutions of the stars with companions; but there

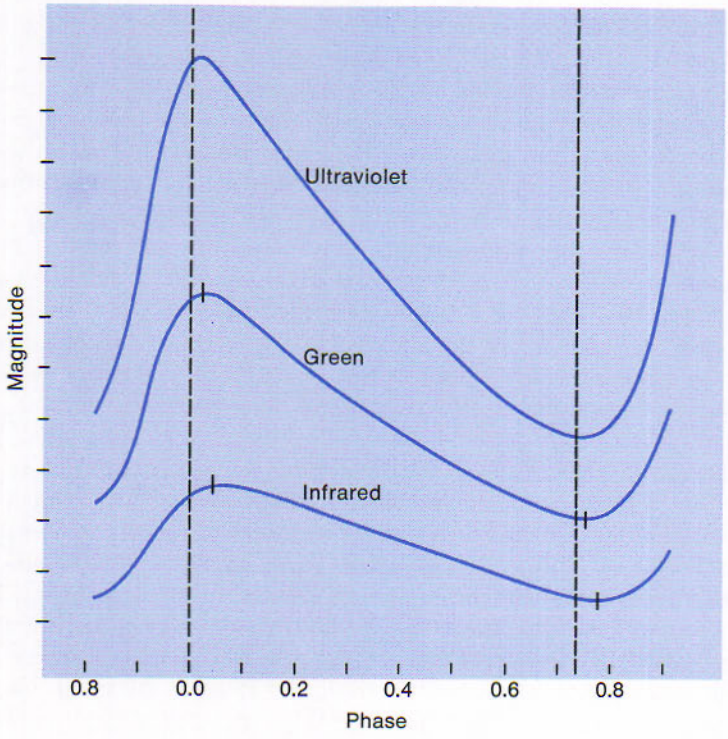
was no convincing idea of how the light variations could be caused by the revolutions. Suggestions that both spectrum and light variations could be shown by single pulsating stars received little attention until H. Shapley, in 1914, advocated the pulsation theory.

13.6
Light Variation of Delta Cephei

Light curves of Delta Cephei at three wavelengths are shown in Fig. 13.6. The light of the different colors was transmitted by appropriate filters placed in front of a photoelectric cell. The range of the light variation is 1.48 magnitudes in the ultraviolet and only 0.43 in the infrared, showing that the star is bluer at maximum than at minimum brightness. Maximum and minimum light come progressively later with increasing wavelength of the light that is observed. They are 0.27 day later in the infrared than in the ultraviolet. Thus when the star has begun to grow fainter at the shorter wavelengths it is still brightening at the longer ones.

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

Figure 13.6
Light curve of Delta Cephei in three colors. The range of light variation is narrower and both maximum and minimum brightness occur later in the longer wavelengths. *(Determined photoelectrically by J. Stebbins).*



13.7 Pulsating Stars

In the original form of the pulsation theory the star was supposed to expand and contract alternately all in phase. On this theory the star should be hottest and therefore brightest and bluest when it is most contracted. Actually it reaches that state almost a quarter of the period later when its spectrum lines are farthest displaced toward the violet, showing that the gases which cause the lines are moving outward fastest. The original plan was oversimplified.

The current pulsation theory, proposed in 1938 by M. Schwarzschild at Princeton, is more flexible and can lead to more complex results. The star's interior is supposed to pulsate in unison as before. Compressional waves run upward through the outer layers, reaching the higher levels later than the lower ones. Greatest compression of the gases in and above the photosphere may occur when the waves are moving outward fastest. The delay in the time of maximum brightness in light of longer wavelength, as is clearly shown in the case of Delta Cephei (Fig. 13.6), is consistent with the theory in its newer form. Although the cause of the pulsation is not clearly understood, it is supposed to be a transitory disturbance in the life of a star as it evolves across the top of the color-magnitude diagram. The region of instability in the color-absolute visual magnitude diagram, according to Sandage, is a strip of width about $0^m.2$ in color index. This strip extends from $M = +4$, $B - V = 0.0$, to $M = -6$, $B - V = 0.7$; it includes all cepheid and RR Lyrae variable stars.

A remarkable confirmation of the pulsation theory was first observed independently by O. Struve at McDonald Observatory and R. F. Sanford at Mount Wilson in the spectrum of RR Lyrae. Midway between the times of minimum and maximum brightness of the star, one set of narrow dark hydrogen lines formed at a higher level in the star's atmosphere is displaced toward the red, showing that these gases are still moving inward. A second set of broad dark lines from a lower level is displaced to the violet; there the gases have begun to move outward again. In addition, the appearance of narrow undisplaced bright hydrogen lines represents energy released by collision of the two strata. Similar features were later observed in the spectra of W Virginis and other pulsating stars.

13.8 Period-Luminosity Relation

Cepheid variables fluctuate in longer periods as the stars are more luminous. The relation is illustrated by the curves of Fig. 13.8, which show how the logarithm of the period increases as the median absolute magnitude is brighter. (The *median magnitude* is the average between the magnitudes at maximum and minimum brightness.) First established by H. Leavitt in 1908 and improved by Shapley in 1917, the period-luminosity relation has especial importance in providing a cosmic distance scale, because these very luminous stars are visible at great distances. Cepheids

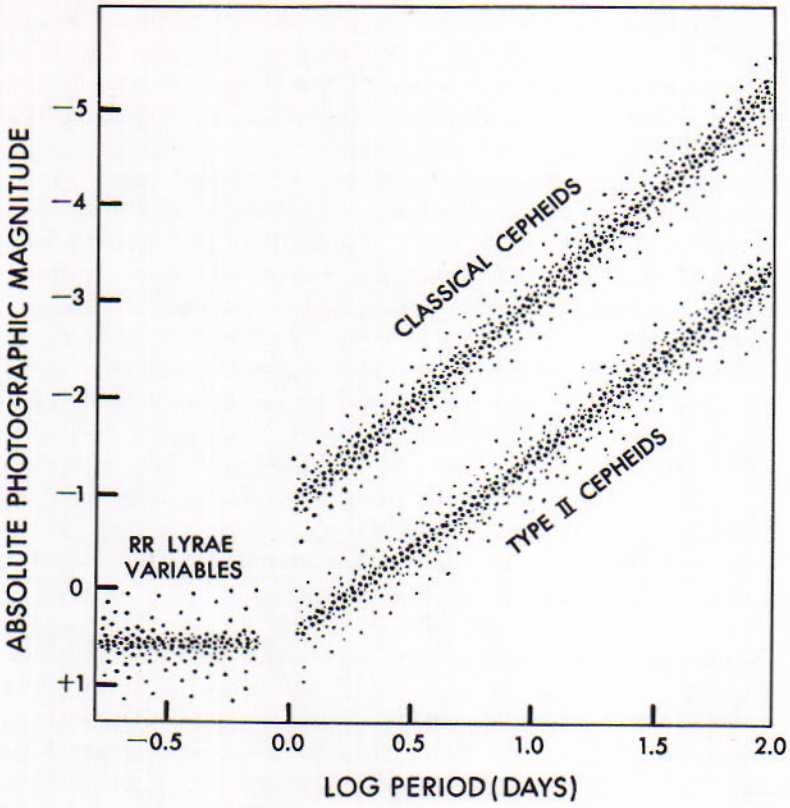


Figure 13.8
Period-luminosity relations for cepheids. The median absolute photographic magnitude of a cepheid can be read from the appropriate band when the period of the light variation and the median surface temperature of the star are known. The median magnitudes of all RR Lyrae variables are around +0.6. *(From preliminary data supplied by H. C. Arp)*

observed in photographs of 30 galaxies with the 200-inch telescope are available as distance indicators if we can be sure that the relation for these galaxies is the same as for our own galaxy. Two modifications of the earlier single curve are made necessary by more recent investigations.

1. The absolute magnitudes of classical cepheids are brighter than was formerly supposed. The curve in the diagram of Fig. 13.8 is raised 1.5 magnitudes above the curve previously used. The correction was announced by W. Baade in 1952. An unexpected means of verifying this correction and of supplying data for a second one was later

provided by J. B. Irwin's discovery of classical cepheids in galactic clusters, for which distances are accurately known; the first correction multiplied by the factor 2 the distances already determined from the periods of classical cepheids. The curve for the type II cepheids replaced the original curve.

2. The period-luminosity relation for both types of cepheids must be represented by bands at least 1 magnitude wide rather than by simple curves. This modification resulted from H. C. Arp's studies of cepheids in the Small Magellanic Cloud, and was explained by A. R. Sandage. From the known relation between the period and mean density he deduced that the absolute magnitude of a cepheid is related not only to the period of its pulsation but also to the star's mean surface temperature, or color. This would apply to all pulsating stars. The correctness of the explanation is confirmed by studies by R. P. Kraft and others of the few known cepheids in galactic clusters, where the surface temperatures can be accurately determined. These studies also show that for stars having the larger variations in brightness the absolute magnitudes are represented more nearly by the central curves of the bands of the diagram.

$P\sqrt{\rho} = \text{a constant}$

When the period and mean color of a cepheid of either kind is observed, the star's median absolute magnitude, M, can be read from the appropriate band of the diagram. When the median apparent magnitude, m, is also observed, the star's distance, r in parsecs, can be calculated by the formula (11.25): $\log r = (m - M + 5)/5$. In the use of this and similar photometric formulas, allowance must be made for any dimming of starlight by intervening cosmic dust, which would otherwise cause the distance to come out too great.

RR Lyrae stars are assumed to have a median photographic magnitude of $+0.6$ in the average regardless of period. Wherever one of these stars is found, its approximate distance is given by the formula: $\log r = (m - 0.6 + 5.0)/5$. The relation of luminosity to period and surface temperature is also shown by a band in Fig. 13.8; but the width of the band and the magnitude of its central line are not yet certainly established. RR Lyrae stars are useful for determining distances within our own galaxy. They are not bright enough to be observed with present telescopes beyond the nearest galaxies.

Effects of differences in surface temperatures of the stars are neglected in order to simplify the examples of distance calculations that follow.

(1). The classical cepheid SY Aurigae has a period of 10 days. Its apparent photographic magnitude varies from 9.8 to 11.0. Determine its approximate distance, supposing that no cosmic dust intervenes.

Answer: The median apparent magnitude is 10.4. The logarithm of the period is 1.0, and the corresponding absolute magnitude from the period-luminosity curve is −3.1. Thus log r = (10.4 + 3.1 + 5.0)/5 = 3.70. The resulting distance is 5000 parsecs.

(2). Required the approximate distance of the RR Lyrae variable RX Eridani, which has a period of 0.6 day and median apparent magnitude 9.2. We again suppose no intervening dust.

Answer: The median absolute magnitude of an RR Lyrae variable is assumed to be +0.6. Thus log r = (9.2 − 0.6 + 5.0)/5 = 2.72. The distance is 520 parsecs.

13.9 Beta Canis Majoris Variables

These are pulsating stars of spectral types B1 and B2. The few known examples have been studied intensively, especially by O. Struve and associates. They form a band in the color-magnitude diagram, which begins with Gamma Pegasi near the main sequence and extends to the left and above that sequence as far as the prototype. The periods of the light variations increase progressively in this direction from 3½ to 6 hours.

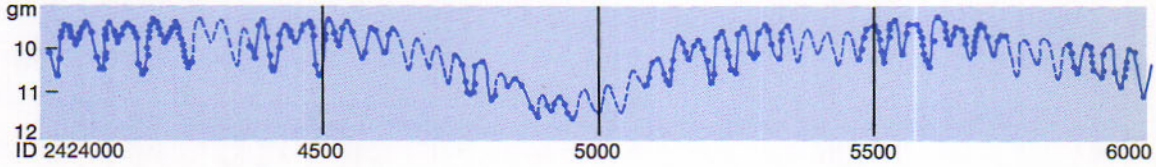
The light variations have small and often variable ranges from cycle to cycle and the spectrum lines oscillate in the same periods, as with other pulsating stars.

13.10 RV Tauri Variables

Among the yellow and reddish variable stars, around types G and K, which do not conform to the principal patterns, 40 or more form a sort of connecting link between the cepheids and the red variables. They are known as RV Tauri stars after one of their members. Their variations in light in a semiregular manner are ascribed at least partly to pulsations. Like the cepheids, these stars are redder at minimum light and are separated into type I and II groups.

RV Tauri itself can serve as an example, although the pattern is not the same throughout the group. Its light curve (Fig. 13.10) shows a semiregular variation of about a magnitude in a period around 79 days. A cycle comprises two maxima of nearly equal brightness and two unequal minima; meanwhile, the spectrum varies between G4 and K5. In this case there is also a superimposed variation of more than 2 magnitudes in a period of about 1300 days.

Figure 13.10
Light curve of RV Tauri. *(Diagram by Leon Campbell and Luigi Jacchia)*



RED VARIABLE STARS

Many red giants and supergiants are variable in brightness. Their variability involves temperature variations of the photospheres and may in some stars be enhanced by variable veiling by clouds of solid or liquid particles alternately forming and dispersing in the outer regions of their atmospheres. These variable stars are of two general types: (1) Mira-type, or long-period, variables, having an approach to regularity; (2) irregular variables, where the fluctuations are less readily predictable. The red giant variables have been often regarded as pulsating stars, and are so classified in the General Catalogue of Variable Stars.

Some red stars and certain others of the main sequence are also irregularly variable in brightness.

13.11 Mira-Type Variables

These are variable in cycles of from 3 months to more than 2 years, and most commonly around 10 months. Their periods and ranges in brightness have an approach to regularity which is remindful of the cycles of sunspot numbers. The range in their light variations averages 5 magnitudes and may be as much as 10 magnitudes in the extreme case of Chi Cygni. This star at maximum brightness in its different cycles is sometimes plainly visible and at other times is still invisible to the naked eye. Meanwhile, the total radiations of Mira-type variables, as determined by their heating of a thermocouple at the focus of a telescope, vary only about 1 magnitude.

The spectra are mainly of type M, but S, R, and N are represented as well. They contain the dark lines and bands characteristic of red stars, and also bright lines, particularly of hydrogen, which have been the means of discovering many of these variable stars. The bright hydrogen lines, which are first formed at lower levels than the dark ones, often make their appearance about midway between minimum and maximum brightness of the stars, become most intense about a sixth of the period after maximum, and disappear around the next minimum.

As with the cepheids, the spectral type moves toward the blue end of the sequence from minimum to maximum light, and the surface temperature increases accordingly. We can best consider the problem presented by the Mira-type variables by referring to the prototype as a particular example.

13.12 Mira

Mira (Omicron Ceti) itself is the best-known and at times the brightest of these variable stars; its light variations have been observed for more than 3½ centuries. It was in fact the first variable star to be recognized, aside from two or three novae, and was therefore called *stella mira*. Mira

is a red supergiant at least 10 times as massive as the sun, having a diameter 300 times as great, and an average density only about a 3-millionth that of the sun. Its maximum brightness ranges in the different cycles generally from the 3rd to the 5th apparent visual magnitude, and its least brightness from the 8th to the 10th magnitude. The average period of its light variation is 330 days.

From minimum to maximum brightness of the star the spectrum of Mira varies from M9 to M6. The surface temperature rises from 1900° to 2600°K, increasing by the factor 1.37. By Stefan's law (10.4) the rate of the star's total radiation increases as the 4th power of this quantity, or by the factor 3.5. Yet the visible light increases an average of 5 magnitudes, or by the factor 100. The large difference between these factors is ascribed to the temperature change and the diminished veiling of the star by its molecular bands and clouded envelope as the star becomes hotter.

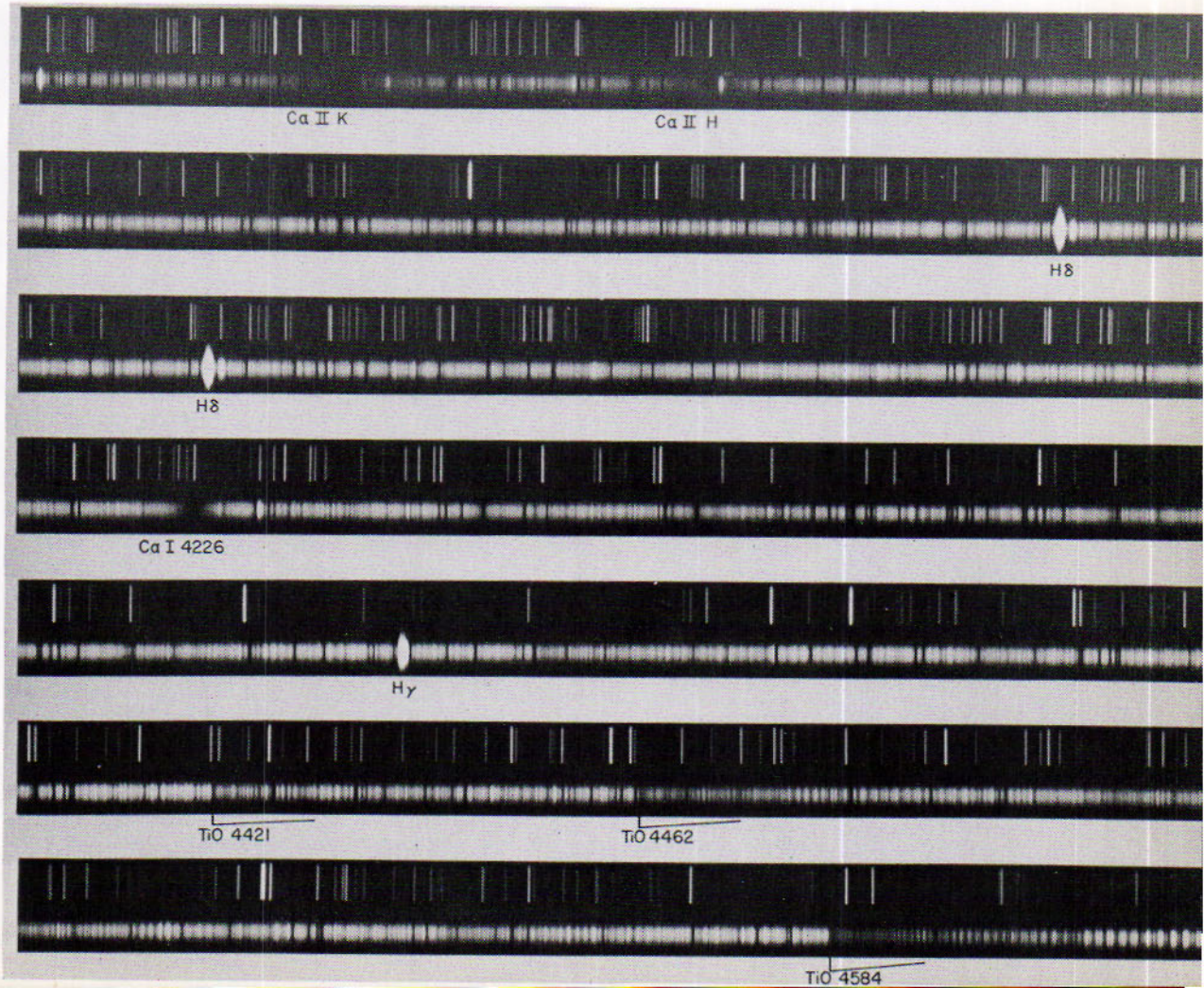


Figure 13.12
Spectrum of Mira. The titanium oxide bands are a prominent feature. Note the hydrogen lines in emission characteristic of long period variables. *(Lick Observatory photograph)*

A. H. Joy's long-continued studies of the spectrum of Mira have served to guide the thinking about the cause of its variability. The dark lines show a slight longward displacement around maximum light, especially at the brighter maxima. The behavior of the bright lines has seemed more significant for the interpretation. They are most displaced toward the violet around maximum light of the star, as with the cepheids, but there is a one-way shifting during a cycle rather than an oscillation. Indeed, there is little evidence from the spectrum or from measurements with the interferometer that Mira is alternately expanding and contracting.

13.13 The Nature of Mira-Type Variability

P. W. Merrill has also called attention to some unusual changes that have occurred in the spectra of long period variables. In somewhat earlier times the pulsation theory was invoked to interpret the variability of the red as well as the blue and yellow giant stars. More recently, this type of pulsation has come to be considered as a trigger mechanism or else as inoperative in these red stars. In a lecture in 1955, Merrill expressed his preference for a main hypothesis of "hot fronts," perhaps like shock waves, moving outward successively from below the photosphere and disappearing at the higher levels as a new disturbance forms below. Merrill's description of the proposed process may be summarized as follows:

The impact of each disturbance at the visible levels of the star is spread over an interval of weeks or months as the wave travels outward with moderate speed. The earlier arrival of its infrared radiations brings a general warming and brightening of the photosphere. The ultraviolet radiations arrive later to cause the bright lines in the spectrum.

At length comes the kinetic impact of the wave itself, which then proceeds outward and after months of travel reaches the uppermost layers of the atmosphere. Here the dissipating wave may cause the gases to condense into droplets which evaporate slowly in much the same way that a slowly disappearing train of white particles is set up by a jet airplane above us. The variable veiling of the photospheres, along with the varying strength of the titanium oxide bands in the spectra, may contribute to the large ranges in brightness which characterize the Mira-type variables.

13.14 Irregular Variable Stars

Many red supergiant and giant stars vary irregularly in brightness in narrower limits often not exceeding half a magnitude. Betelgeuse is the brightest of these. Another example, the type M5 supergiant Alpha Herculis, varies unpredictably between visual magnitudes 3.0 and 4.0. Its distance from us is 500 light years and its diameter is 500 times the sun's diameter. A. J. Deutsch's studies of the spectrum have shown that this red star has an envelope extending out at least as far as its visual companion, a distance of 700 astronomical units. The material of the envelope is moving outward at a rate exceeding the velocity of escape from

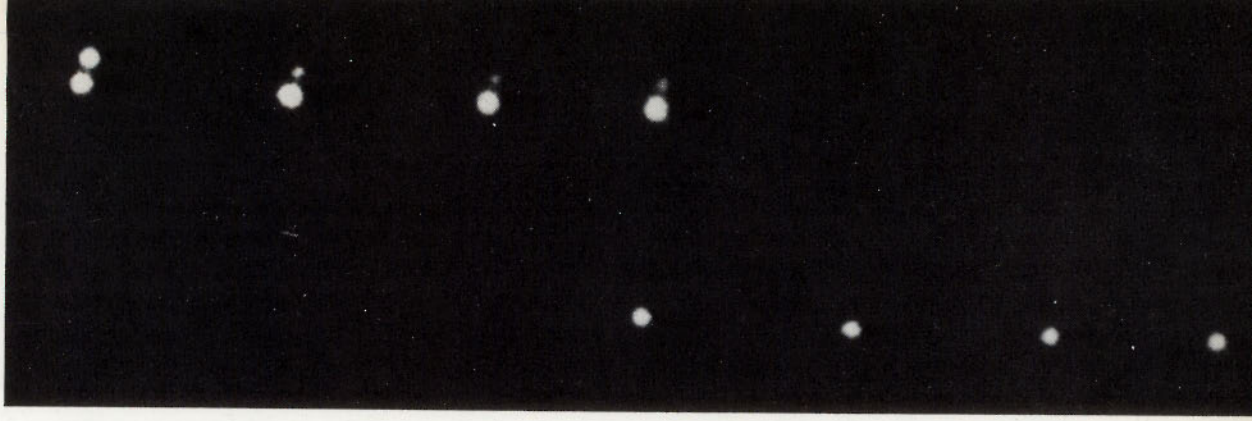


Figure 13.15
Flareup of Krüger 60B, July 26, 1939. The last of four successive exposures (left) on the binary and its distant optical companion, showing the brightening of the fainter star of the pair. *(Sproul Observatory photograph)*

the star. At very high levels the gas may condense into patchy clouds, which disappear by dilution and are replaced by others. Their partial veiling of the red star is believed to contribute to its variability.

Other irregular variable stars include the T Tauri variables (16.15) and the flare stars.

13.15 Flare Stars

These are red main-sequence stars that are subject to intense outbursts of very short durations remindful of the solar flares (10.28). They are designated in the *General Catalogue* as UV Ceti-type variables after a typical representative. This star is the fainter component of the binary Luyten 726-8. The main outbursts of UV Ceti occur at average intervals of 1½ days, when the rise in brightness of the star is generally from 1 to 2 magnitudes. On one occasion in 1952, however, an increase of 6 magnitudes was observed, the greatest flare on record for any star. Between the outbursts the light of the star varies continuously and irregularly in smaller amplitude.

F. Whipple and B. Lovell have observed UV Ceti simultaneously at radio and optical wavelengths and have found a correlation. If so, the radio energy is orders of magnitude stronger than that associated with solar flares. Similar correlation with other flare stars is lacking because of the extended observing periods required due to the randomness of the events.

About two dozen flare stars are recognized. Their spectra, which normally contain emission lines of hydrogen and ionized calcium, show some bright lines of helium as well during the flares. Another example of a flare was the sudden and brief increase of 1½ magnitudes in the light of the normally fainter star of the visual binary Krüger 60B (Fig. 13.15) reported by P. van de Kamp and Sarah L. Lippincott.

Although the outbursts of flare stars are believed to be confined to small areas of their surfaces, the greater brightness of the flares than that of their dimmer surroundings much increases the total light of these stars. Similar flares in other than faint red stars would be more likely to escape detection.

ERUPTIVE STARS

The word nova *(Latin for* new; *plural:* novae*) was used by astronomers in the past to identify stars that increase rapidly in brightness from previous relative obscurity which on occasions had kept them invisible before the time of their eruption. Current knowledge, however, shows that in spite of their traditional adjective, these stars are not new, but rather old ones making violent adjustments in their evolution. They are often designated by the word Nova followed by the possessive of the constellation name and the year of the outburst. Nova Aquilae 1918 is an example. More recently they have been designated by letters along with other variable stars. Thus Nova Herculis 1934 is also Nova DQ Herculis. In addition to typical novae, recurrent novae, and dwarf novae, there are also the rare and even more spectacular supernovae. The account in this chapter is restricted to novae in our galaxy.*

13.16 Typical Novae

More than 150 typical, or "ordinary," novae have been recognized in our galaxy, a number that is increasing by one or more in a year. Four novae were observed in the first half of 1970. Only one of them, *Nova Serpentis 1970,* became visible to the naked eye, and then only barely so. This star could not be seen on plates taken on February 12 with a limiting magnitude of 9. On the following day it had become a star of the seventh magnitude and on the 15th of the same month, it attained its peak brightness of the fourth magnitude. Many novae escape detection; it is estimated that a total of 25 appear yearly in the Milky Way. Five typical novae in

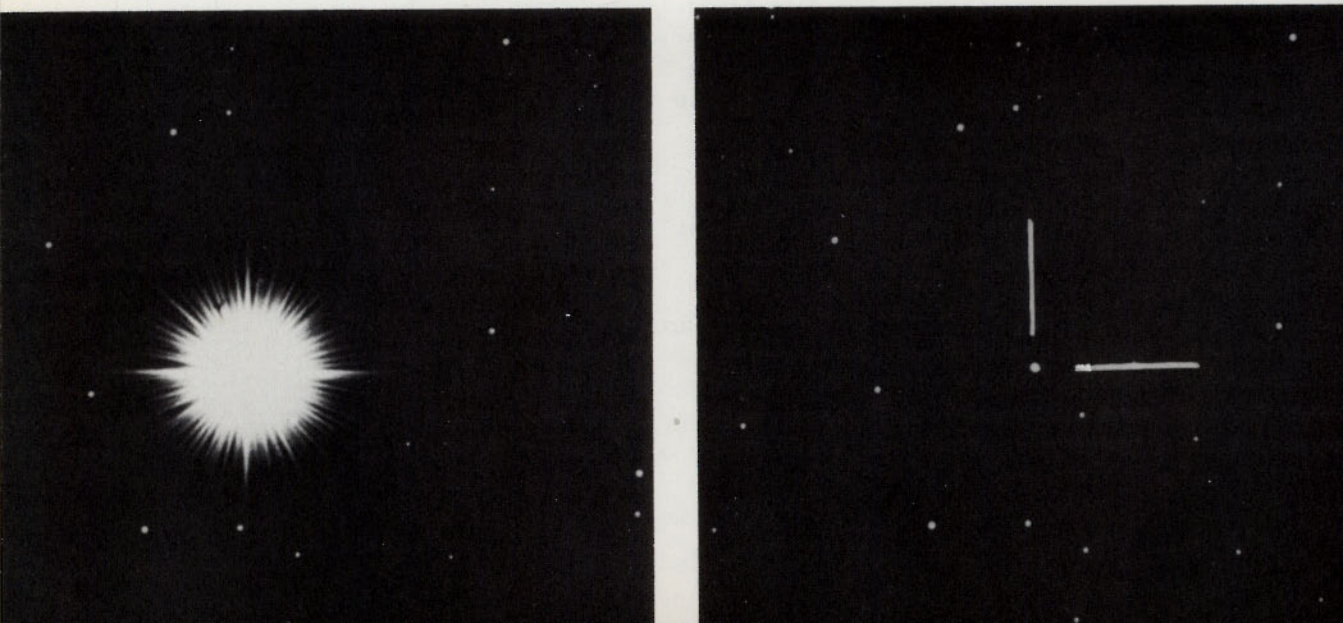


Figure 13.16A Change in brightness by Nova Herculis 1934 from March 10 to May 6, 1935. *(Lick Observatory photograph)*

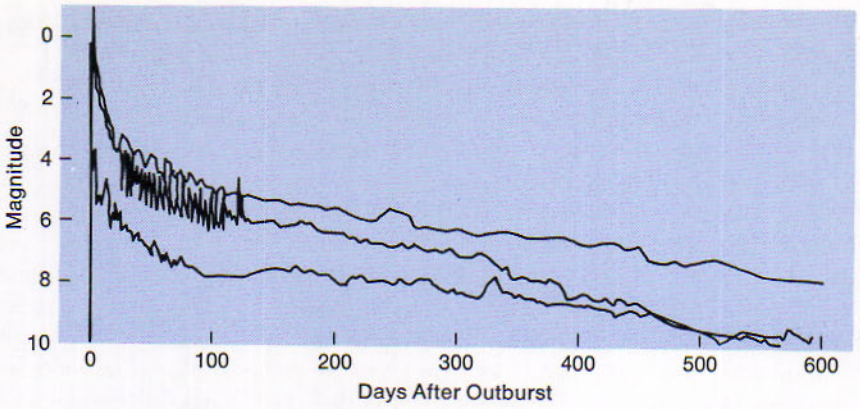


Figure 13.16B
Light curves of Nova Aquilae 1918, Nova Persei 1901, and Nova Geminorum 1912. They are designated in order of decreasing brightness at the maxima. *(Harvard Observatory diagram)*

the present century became stars of the first magnitude or brighter. Nova Persei 1901 rose to apparent magnitude +0.1, as bright as Capella. Nova Aquilae 1918 reached magnitude −1.4, as bright as Sirius. Nova Pictoris 1925, Nova DQ Herculis (1934), and Nova CP Puppis (1942) became, respectively, as bright as Spica, Deneb, and Rigel.

Typical novae are subdwarf stars smaller and less massive than the sun. Characteristic of their light variations is a single abrupt rise to maximum brightness and a much slower decline, interrupted by partial recoveries. The rise may exceed 12 magnitudes, representing an increase in brightness of more than 60,000 times, and usually requires only a day or two. The novae return eventually to about the same faint magnitudes they had before the outbursts occurred, suggesting that the effect of the eruptions on these stars is superficial. On their returns they are likely to fluctuate moderately for some time. As many as 20 to 40 years may elapse before they settle down to comparative stability.

13.17 Recurrent Novae

The novae already described have been observed to erupt only once. *Recurrent novae* have two or more recorded outbursts; otherwise they seem to differ from typical novae only in their less extreme rise in brightness. Six examples are recognized. Their names and the dates of their recorded outbursts are as follows:

Nova T Coronae Borealis rose in 1866 to visual magnitude 2, and in two months thereafter declined to the 9th magnitude; it flared out again in 1946, then to the 3rd magnitude. Nova RS Ophiuchi, normally around the 12th magnitude, rose to the 4th magnitude in 1898 and 1933, and to the 5th magnitude in 1958. Nova T Pyxidis rose from magnitude 13 to nearly naked-eye visibility in 1890, 1902, 1920, and 1944. Outbursts of Nova U Scorpii occurred in 1863, 1906, and 1936; of Nova WZ Sagittae in 1913 and 1946; of Nova Sagittarii in 1901 and 1919.

There is some indication that the amount of the rise in brightness for all recurrent and dwarf novae varies directly as the logarithm of the average interval of time between outbursts. If the relation holds for typical novae, the intervals between their outbursts may be many thousand years. Thus typical novae such as Nova Aquilae 1918 would be expected to flare out again eventually.

13.18 Spectral Changes of Novae

During the initial increase in brightness the spectrum of a typical nova usually contains a pattern of dark lines somewhat like that of a type A star. These lines are much displaced to the violet, showing that the gases in front of the star are rapidly approaching us.

Soon after the nova attains its maximum brightness, broad undisplaced emission lines suddenly appear, having the dark lines at their violet edges. As the light of the nova fades, the bright lines become stronger, and three sets of absorption lines successively make their appearance, each set more displaced to the violet than the preceding one. Radial velocities occasionally exceeding 3000 km/sec are represented by the displacements of the dark lines and the half-widths of the bright ones. With the further decline of the nova the bright lines persist until they resemble the spectrum of an emission nebula, except that the nova lines are wider.

13.19 The Nature of Novae

Novae are currently regarded as stars that are collapsing to become white dwarfs but are too infrequent, according to R. P. Kraft, to constitute a typical stage in stellar evolution. These stars become unstable at times, releasing more internal energy than their small photospheres can radiate. The following account of the nature of novae is taken from the interpretation of their spectral changes as described by D. B. McLaughlin.

A hot subdwarf star in pre-nova state may have remained constant in light or have varied through only a small range for many years. Quite suddenly an excessive amount of energy is liberated below the surface. A superficial layer is then violently ejected, but the main body of the star does not expand. The ejection is not instantaneous; it occurs during an appreciable fraction of the star's rise to maximum brightness. The total mass released in the whole explosion is about 1/10,000 of the star's mass and the energy released is 10^{45} ergs.

The enormous increase in the radiating surface of the ejected shell gives the effect of a rapidly expanding photosphere, causing the star to become much brighter. The dark lines strongly displaced to the violet in the nova's spectrum are produced by absorption in the part of the spreading shell that is in front of the star and is therefore approaching us. The broad emission lines appear soon after maximum brightness of the star when the gaseous shell becomes transparent enough by dilution so that the light begins to come through from all parts of it, some approaching and others receding from us.

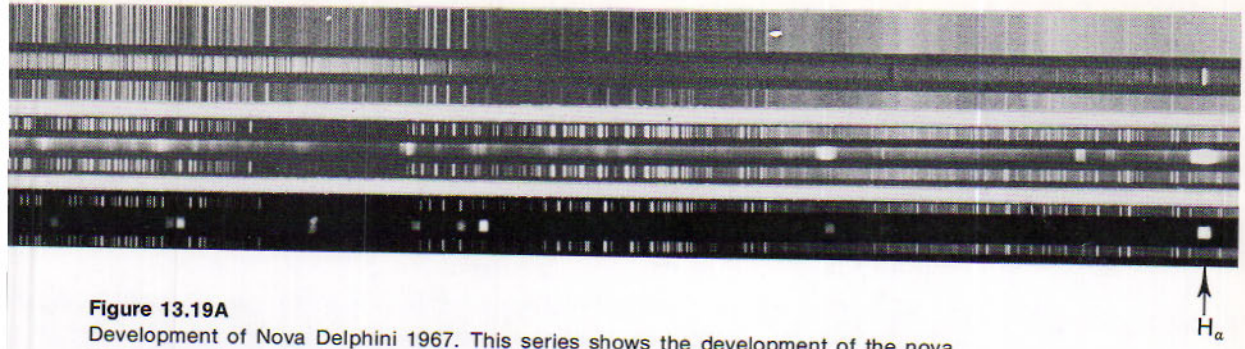


Figure 13.19A
Development of Nova Delphini 1967. This series shows the development of the nova from premaximum (top) through the early nebular stage (middle) to the post nebular stage (bottom) when the observations were terminated. The prints are at slightly different scales and have been aligned at $H\alpha$ for convenience. *(Courtesy of J. Grygar, Ondrejov Observatory)*

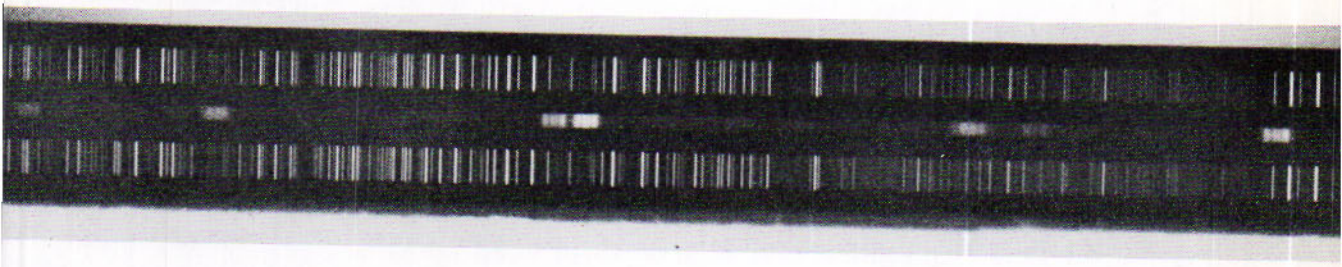


Figure 13.19B
Nova Delphini 1967 prior to the nebular stage. Notice blue shifted absorption components at −200, −400, and −1000 km/sec in this spectrogram taken on March 2, 1968. *(Courtesy of J. Grygar, Ondrejov Observatory)*

The different layers of the shell that have been successively ejected cause the successive patterns of dark lines to appear. The inner layers, having higher speeds of expansion, catch up with the outer one, forming with it the principal shell. This shell produces later the bright nebular lines in the spectrum, when the gases have become sufficiently rarefied by the expansion. The shell of a nearer nova has sometimes grown large enough to be visible directly with the telescope.

D. B. McLaughlin divides the typical history of a nova into two parts, *pre and post-maximum.* In the pre maximum part there are four stages; pre nova *flickering, initial rise,* a *pause (stillstand)* and a *final rise* of two magnitudes to maximum. After maximum, there are four more stages; the *nebular shell,* the *post maximum decline,* the *transition* and the *final decline.* All ordinary novae follow a very similar light curve except for the transition stage.

A nova entering the transition stage is on a general exponential decline in light. The nova may suddenly fluctuate wildly around a continuation of this line or it may drop several magnitudes and then recover to the normal declining line at the end of the stage. It then gradually fades in

the final decline until months or years later it returns to its original pre-nova brightness.

The duration of the various stages is proportionately the same in all normal novae except for the final rise to maximum. Here there are two distinct classes called slow and fast novae: in *slow novae* the final rise takes 25 days while in the *fast novae* it takes only 1 to 2 days. Nova Aquilae 970 was a slow nova. Nova Aquilae was a fast nova. It is interesting that the transition phase takes up about 20% of the light curve whether the lifetime of the nova event is six months or two years.

13.20 Expanding Envelopes of Novae

The envelope around Nova Aquilae 1918, for example, became visible with the telescope about four months after the star began to brighten. Thereafter, the envelope was observed to increase in radius about 1″ a year. The linear rate of the expansion, as determined from the shortward displacement of the dark lines in the spectrum, or by the halfwidth of the bright lines, was around 1700 km/sec. Because the tangential velocity, T, of the expansion was presumably the same as the radial velocity and because the angular velocity, μ, was also known, the distance, r, in parsecs of the nova, could be found by the relation (11.11): $r = 0.211\ T/\mu$. The distance of Nova Aquilae proved to be 360 parsecs, or about 1200 light years.

The envelope around Nova Aquilae attained a diameter exceeding 10,000 times the earth's distance from the sun, and disappeared soon afterward. The envelopes around other typical novae have generally vanished only a few years after their appearance. Their rapid expansions and short durations contrast sharply with the slow expansions and relatively long lives of planetary nebulae, such as the Ring nebula in Lyra. The envelopes of supernovae, however, expand rapidly, persist for centuries, and are discrete radio sources.

There are other ways to determine the distances to novae: one is similar in principle to the visible expanding shell; the other combines temperature and brightness measurements.

If the nova is located in a gaseous and dust region of the galaxy the visible radiation will illuminate the surrounding region. The illuminated region grows over an interval of time at the velocity of light (c) which

Figure 13.20A
Nova Delphini 1967 in the nebular stage. This plate taken on October 16, 1968 shows the symmetrically structured emission lines. *(Courtesy of J. Grygar, Ondrejov Observatory)*

Figure 13.20B
Expanding nebulosity around Nova Persei 1901. *(Photograph from the Hale Observatories)*

replaces the velocity, T, used in the above discussion and the calculation can then be made. This method was applied to Nova Persei which occurred in 1901.

Another method makes use of the observed temperatures, increase in radius, and apparent magnitudes. From spectrograms taken at times t_1 and t_2 ($t_2 > t_1$) and measurements of the expansion velocity, v, we can derive the difference in radii, ΔR.

$$\Delta R = R_2 - R_1 = v(t_2 - t_1)$$

From (11.24) we know the relation between the absolute magnitude, M, its radius, R, and its temperature, T. Since we can measure the apparent magnitude, m, at the different times we obtain another relation for the radii and hence the radii at the two times of observation. These values can then be substituted into our formulae to determine the absolute magnitude and hence the distance.

$$\log \frac{R_2}{R_1} = 5900\left(\frac{1}{T_2} - \frac{1}{T_1}\right) - \frac{m_2 - m_1}{5}$$

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

This last method does not depend upon measurements of the angular velocity and thus can be applied in cases where the object is so far away that no shell appears. While the method is not as reliable as those involving the angular velocity, it is often the only applicable one.

13.21 Supernovae in Our Galaxy

Supernovae are stars considerably more massive than the sun, which explode once in the course of their lifetimes. When they explode, they become many million to perhaps billions of times more luminous than the sun and blow into space gaseous material amounting to at least one solar mass. Six or seven such outbursts were recorded in the Galaxy during the past 2000 years. Allowing for the many that are too remote to be conspicuous, J. S. Shklovsky estimates that supernovae are exploding in the system of the Milky Way at an average rate of one in 30 to 60 years. F. Zwicky has proposed a more conservative number of one per century, based on extensive studies of various galaxies. It is even possible that the actual rate might be closer to that of one supernova every four centuries. Novae and supernovae in other galaxies are discussed in Chapter 18.

The apparently brightest supernova on record flared out in Cassiopeia in November, 1572, and was observed by Tycho Brahe. It became at least as bright as Venus and thereafter gradually faded until it disappeared to the unaided eye in the spring of 1574. "Kepler's star" in Ophiuchus in 1604 rivaled Jupiter in brightness. The explosion of a supernova in Taurus in the year 1054 produced the expanding Crab nebula.

Supernova seem to divide into two classes: type I and type II. Type I exhibit the presence of heavy elements in their spectra and relatively little hydrogen while type II exhibits primarily hydrogen. The differentiation is real and probably arises from differences in the masses of the original stars. In either case, the energy released in the supernova event is on the order of 10^{50} ergs. The matter is ejected explosively outward at velocities ranging from 1000 to 7000 km/sec. Supernova remnants such as Tycho's nova continue to emitt energy at the rate of 10^{36} erg/sec.

13.22 Expansion of the Crab Nebula

The Crab nebula (Messier 1) in Taurus has a radius of about 180″, which is increasing at the rate of 0″.2 a year. At the distance of 3500 light years, the present linear diameter is 6 light years. The nebula is expanding around the site of a supernova recorded in Chinese annals as having appeared southeast of Zeta Tauri on July 4, 1054. The supernova became as bright as Jupiter and remained visible for two years.

See Plate IX

The Crab nebula consists of a homogeneous central structure surrounded by an intricate system of filaments (Fig. 13.22A). The spectrum of the amorphous central region is continuous. Strong polarization of this light suggests that it is synchrotron radiation (4.27), like that produced by

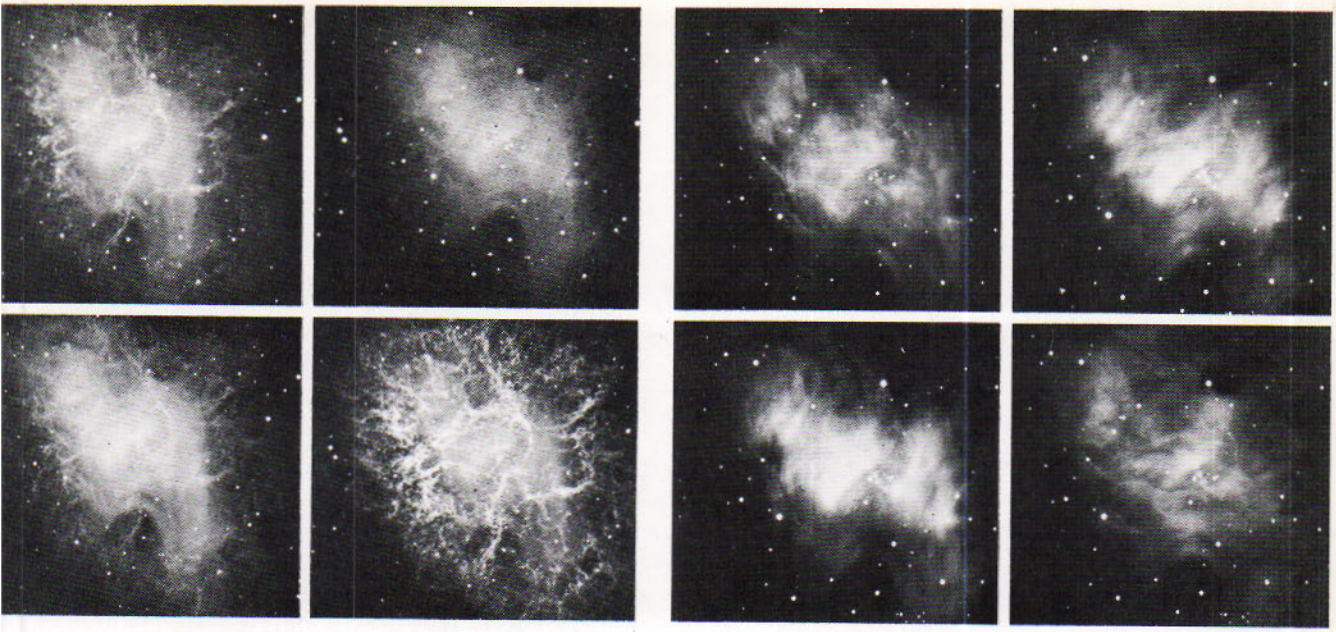


Figure 13.22A (Top left)
NGC 1952 "Crab" nebula in *Taurus.* M1. Four photographs in blue light (top left), yellow (top right), red (bottom left), and infrared (bottom right). *(Photographs from the Hale Observatories)*

13.22B (Top right)
The Crab nebula in polarized light. The polarized light comes from the hazy blue light region (Plate XI) and arises from synchrotron emission. (Photograph from the Hale Observatories)

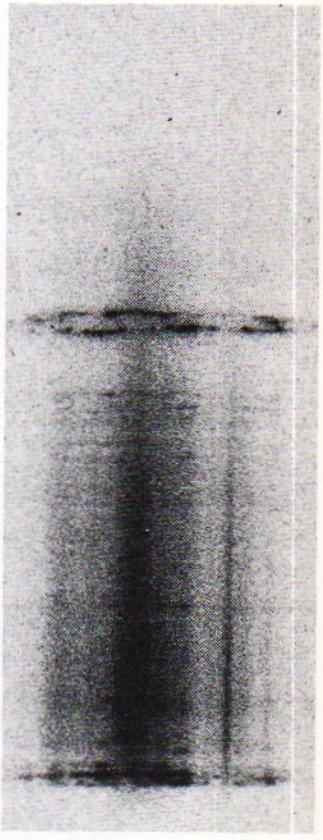
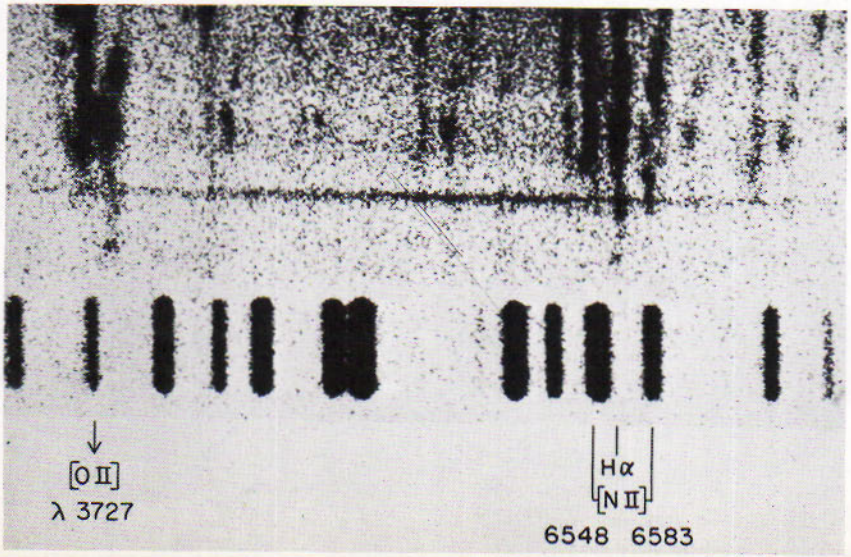


Figure 13.22C (Right)
Spectrum of the central region Crab nebula. Doubling of the lines is apparent. The slit spectrograph was oriented in such a way that NP 0532 was included on this plate taken on 29 Nov. 1956. However, the exposure was not quite long enough to record the pulsar. *(Courtesy of N. U. Mayall)*

Figure 13.22D (Bottom)
Spectrum of the south following edge of the Crab nebula. The splitting of the knots and complex structure of the filaments are shown. *(Courtesy of R. Minkowski)*

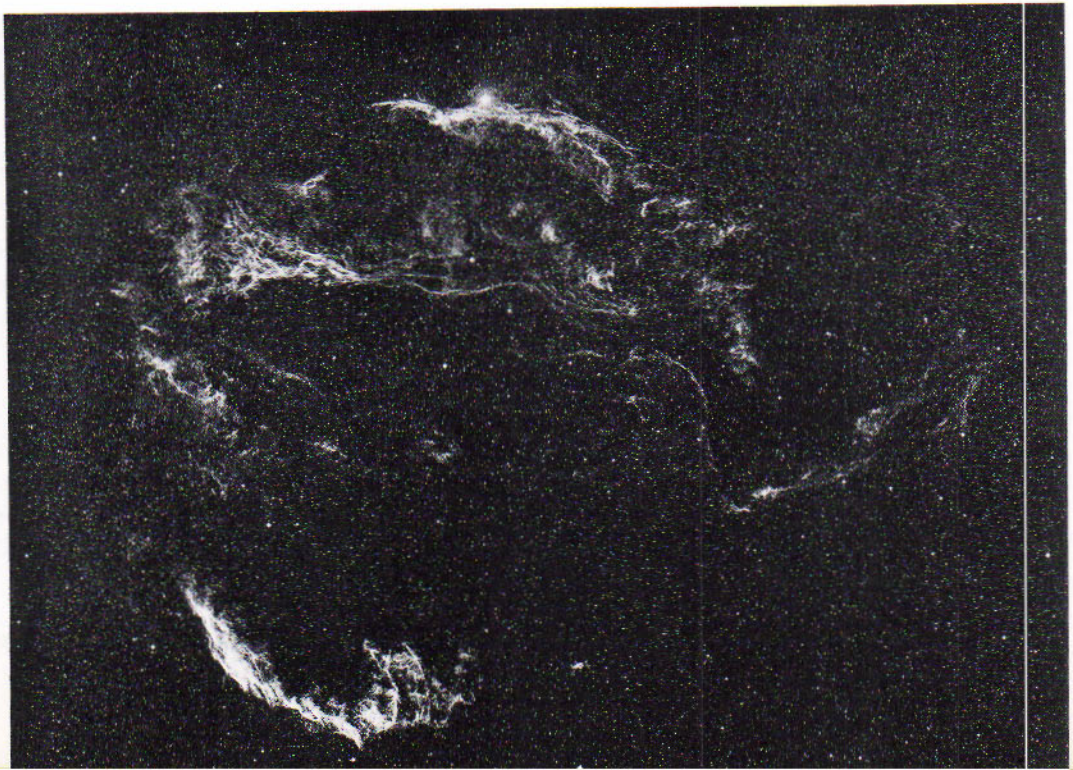


fast-moving electrons revolving in the magnetic field of a laboratory accelerator. The spectrum of the filaments shows emission lines of hydrogen, helium, and other elements observed in the spectra of planetary nebulae. Doubling of these lines (Fig. 13.22 C&D) reveals the expansion of the nebula; the rate of increase of its radius is 1100 km/sec, or about 60 million miles a day. The central star of the Crab is a pulsar (13.25).

13.23 Remnants of Other Galactic Supernova Envelopes

Fragments of the expanding envelope around Tycho's supernova, B Cassiopeia, 1572, have been photographed by Minkowski with the Hale telescope. They are described by him as a faint arc and two inconspicuous filaments of nebulosity moving toward the outside. Remnants of Kepler's supernova of 1604 in Ophiuchus similarly photographed are a relatively inconspicuous fan-shaped mass of filaments and several faint wisps of nebulosity. The motions are away from the center of the area. The radio source Cassiopeia A (17.20) is identified in the photographs by many fragments of an envelope spreading from an otherwise unidentified supernova explosion of about the year 1700. The familiar Loop of nebulosity in Cygnus and other partial wreaths of bright nebulosity in this vicinity (Fig. 16.8) are believed to be expanding around the sites of supernova outbursts.

Figure 13.23
Filamentary nebula in *Cygnus*, showing entire area of nebulosity. Photographed in red light. *(Photograph from the Hale Observatories)*



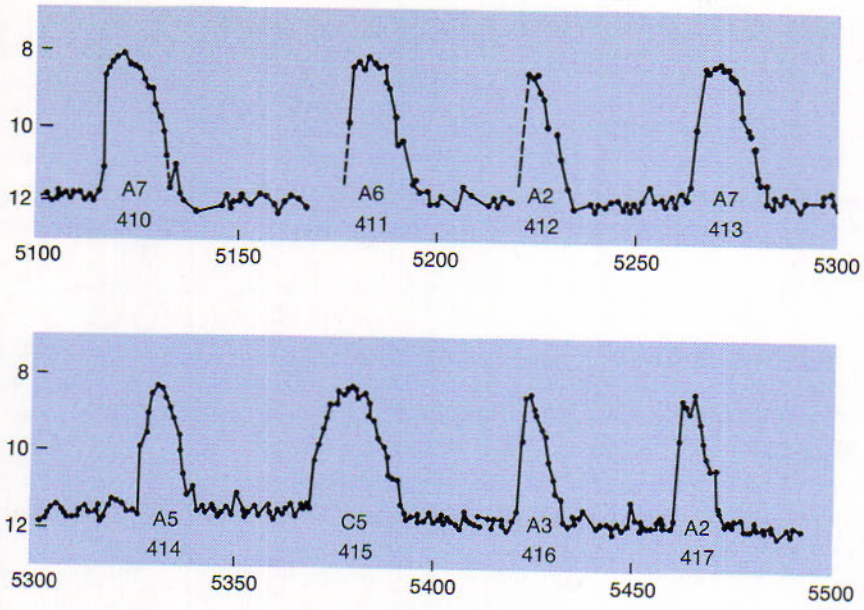


Figure 13.24
Light curve of SS Cygni in 1955. Numbers at the top are Julian days 2,435,100 to 2,435,500. Apparent magnitudes are indicated at the right and left. Determined by the American Association of Variable Star Observers. *(Diagram by Margaret W. Mayall)*

13.24 Dwarf Novae

These are a group of hot subdwarf stars somewhat smaller and fainter than other novae. Typical examples are SS Cygni and U Geminorum. The former star is normally around apparent magnitude 12; it brightens abruptly about 4 magnitudes at irregular intervals and returns to normal brightness in a few days. Over 10-year periods the average interval between outbursts for any one of these stars is about the same, ranging from 13 to 100 days for the group. The stars of a subgroup having Z Camelopardalis as prototype are more erratic in behavior. No evidence is available that dwarf novae eject gases into space at the outbursts.

A. H. Joy's discovery at Mount Wilson Observatory in 1943 that SS Cygni is a spectroscopic binary star may have prepared the way for the eventual understanding of all novae. The revolution period of $6^h\,38^m$ shows that the two stars of this binary system are very close together. One component is a larger red star; its smaller white companion is the nova. M. F. Walker observed in 1954 that the typical nova DQ Herculis is also one component of a very close double star. Other dwarf and typical novae have proved to be members of binary systems. R. P. Kraft has pointed out the possibility that membership in a certain kind of double system may be a necessary condition for a star to become a nova, and has discussed the evolution of such a system.

13.25 Pulsars

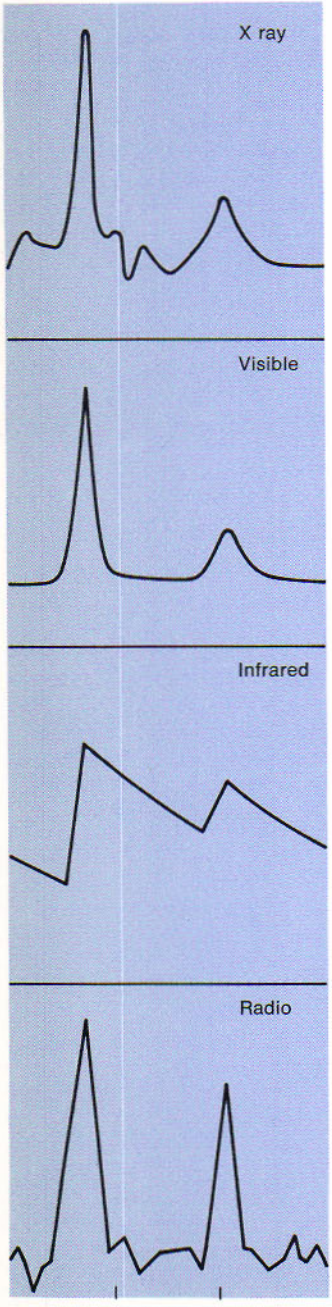


Figure 13.25A "Light" curves of Pulsar NP 0532 at four different wavelengths.

The newest group of variable stars is constituted by the *pulsars*, sources of pulses of radiation emitted at a rapid rate and discovered by A. Hewish and his colleagues in the fall of 1967. Although the best observed of all variable stars, since they have been observed over the entire electromagnetic spectrum, the pulsars are so far the least understood, with nothing but pure theory as the basis of an explanation of these objects as *neutron stars*. They are discussed here because their light curves remotely resemble those of dwarf novae.

Pulsars are designated by two letters followed by four numbers. The first letter identifies the observatory where the discovery is made, while the second is always P, for "pulsar". Thus, a pulsar discovered at the observatory in Cambridge, England will bear the letters CP plus four digits, which give the right ascension of the object to the nearest minute. To give a complete example, NP 0532 is a pulsar at 5^h32^m discovered by astronomers at the National Radio Astronomy Observatory. Up to 1969, 46 pulsars had been discovered, with periods ranging from 30 milliseconds to a little more than 2 seconds.

A feature of their light curves is their apparent "turn-off" between pulses, although this may also be due to a continuum far below the level of the pulses. Their radio pulses are not as sharp as their visible pulses, which in turn are apparently sharper than their x-ray pulses, as seen in Fig. 13.25A; but having only observed one pulsar in the visible region we should not place too much weight on the shapes of the light curves.

The small number of these objects makes it difficult to assign them to a population type although there is a hint of a population I distribution (Fig. 13.25B). One pulsar has been found in the Crab Nebula, NP 0532, (13.22). Since neutron stars may result from supernovae events and neutron stars can explain the pulse phenomenon, an association between the two can be made. The theory predicts that the pulses should slowly lengthen in period and eventually die out. Lengthening periods have been observed, but abrupt shortening of periods have been observed also and are not predicted by the theory. According to classical physics, a shortening of the period would involve a shrinking of the radius of a spinning sphere, assuming a single object is involved. It is interesting to note that in binary stars we find shortening and lengthening periods and we have already pointed out that dwarf novae appear to be in binary systems.

Nevertheless, astronomers agree that we are probably dealing with a single spinning body. If we are in fact dealing with neutron stars then pulsars must be classed as a completely different type of star. They must have very small radii (10 to 100 kilometers), must be very dense (on the order of 10^{15} ρ.) and must have fantastically large magnetic fields (more than 10^{10} gauss). Such conditions allow us to treat these stars as if they were material in the solid state. For example, disturbances would be propagated through the star much as seismic waves are through the earth.

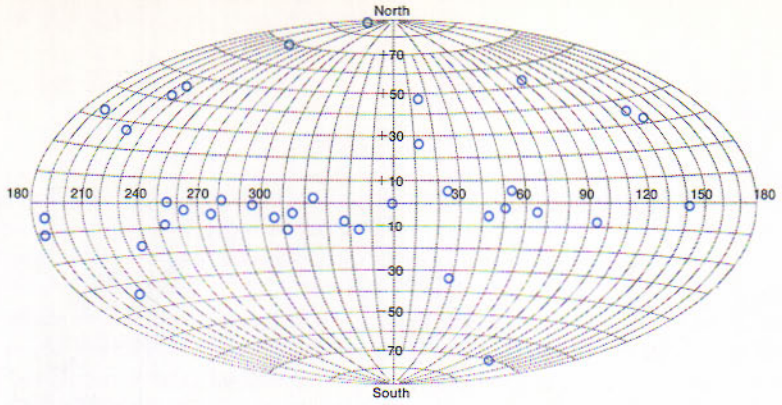
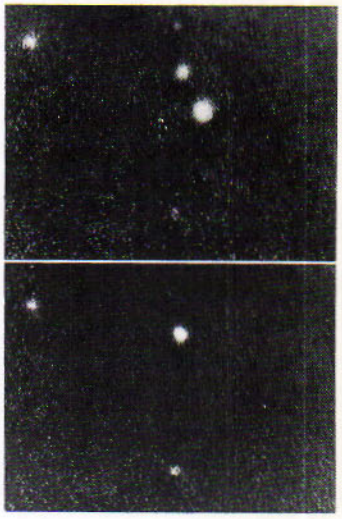


Figure 13.25B
Distribution in galactic coordinates of pulsars.

Figure 13.25C
Pulsar NP 0532. The pulsar is "on" in the top picture and "off" in the bottom one. *(Photograph courtesy of E. J. Wampler, Lick Observatory)*



The surface of these objects would have to be a metallic crust with a melting point on the order of $(10^9)°K$ and the density would increase rapidly toward the center of the star. As the spinning star slows down, stresses in the crust build up and a rapid adjustment in the radius takes place. G. Baym and his colleagues refer to the adjustment that takes place as a *starquake.*

J. Ostriker believes that the pulse emitting stage is short lived, perhaps only 10^7 years and after the pulses cease pulsars are only detectable gravitationally.

It is useful to summarize the various variable stars in a tabular way. Table 13.I summarizes the classes of intrinsically variable stars by the two major population types.

13.26 Classification of Variable Stars

TABLE 13.I
Classification of Variable Stars

PULSATING VARIABLES

POPULATION I		POPULATION II	
Classical Cepheids	$1-50^d$	RR Lyrae	1^d
Dwarf Cepheids	1^d	W Virginis	$7-30^d$
β Canis Majoris	1^d	RV Tau	70^d
β Ceph.	1^d	μ Cep,	70^d
Long Period	200^d		

NOT CLASSIFIED BY POPULATION

δ Scuti Stars
α^2C Vn Stars
SX Centauri
Ultra short period stars
SS Cyg
Pulsars

ERUPTIVE VARIABLES

POPULATION I	POPULATION II
Supernova, type II	Supernova, type I
T Tauri	Nova
RW Aurigae	
R Cor Bor	

NOT CLASSIFIED BY POPULATION

Recurrent Nova
UV Ceti Stars
Z Camelopardalis

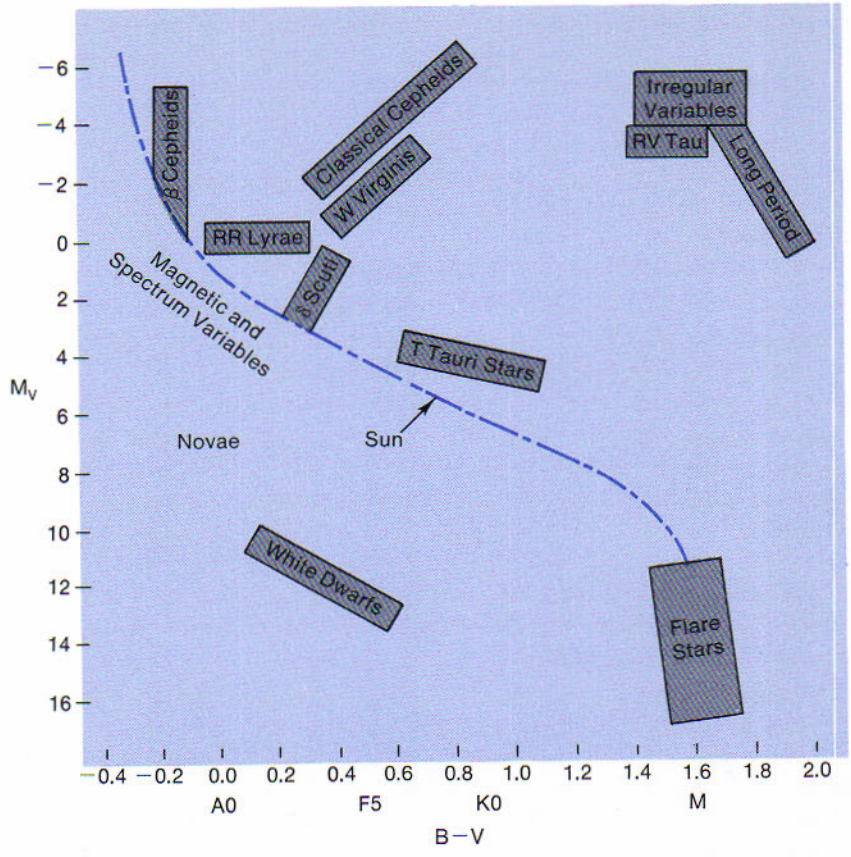


Figure 13.26
H-R diagram showing general regions occupied by various groups of variable stars.

REVIEW QUESTIONS

1. What are the light elements of variable stars?
2. What characteristic of the regular variable stars makes them so important in astronomy?
3. What spectroscopic observations assure us that the classical cepheids are not binary systems?
4. What would be the consequences if the present absolute magnitudes of classical cepheids turns out to be underestimated by one magnitude?
5. If a long-period variable is five magnitudes brighter at maximum brightness than at minimum, what is implied about the star's size? Assume that temperature effects are negligible.
6. Variable stars, according to their definition as used here, are relatively rare yet they account for a good 35% of the literature of stellar astronomy (excluding the Sun). Why?

7. List the nine stages in the history of a normal novae.
8. A nova is observed to have a shell expanding in radius at a rate of 0".2/yr. Absorption lines associated with the shell exhibit a velocity of 300 km/sec. How far away is the nova?
9. Why are we reasonably certain that the Sun is not a flare star in the current usage of the term?
10. If all typical novae are binary systems can you speculate on at least two possible explanations of this? What observations are required to decide between your hypotheses?
11. What is the present theory for the origin of the pulsars?

REFERENCES

Glasstone, S., *Sourcebook on Space Sciences,* Princeton: D. Van Nostrand Co., 1965.

Page, Thorton and L. W. Page, eds., *The Evolution of Stars,* New York: The Macmillan Co., 1968.

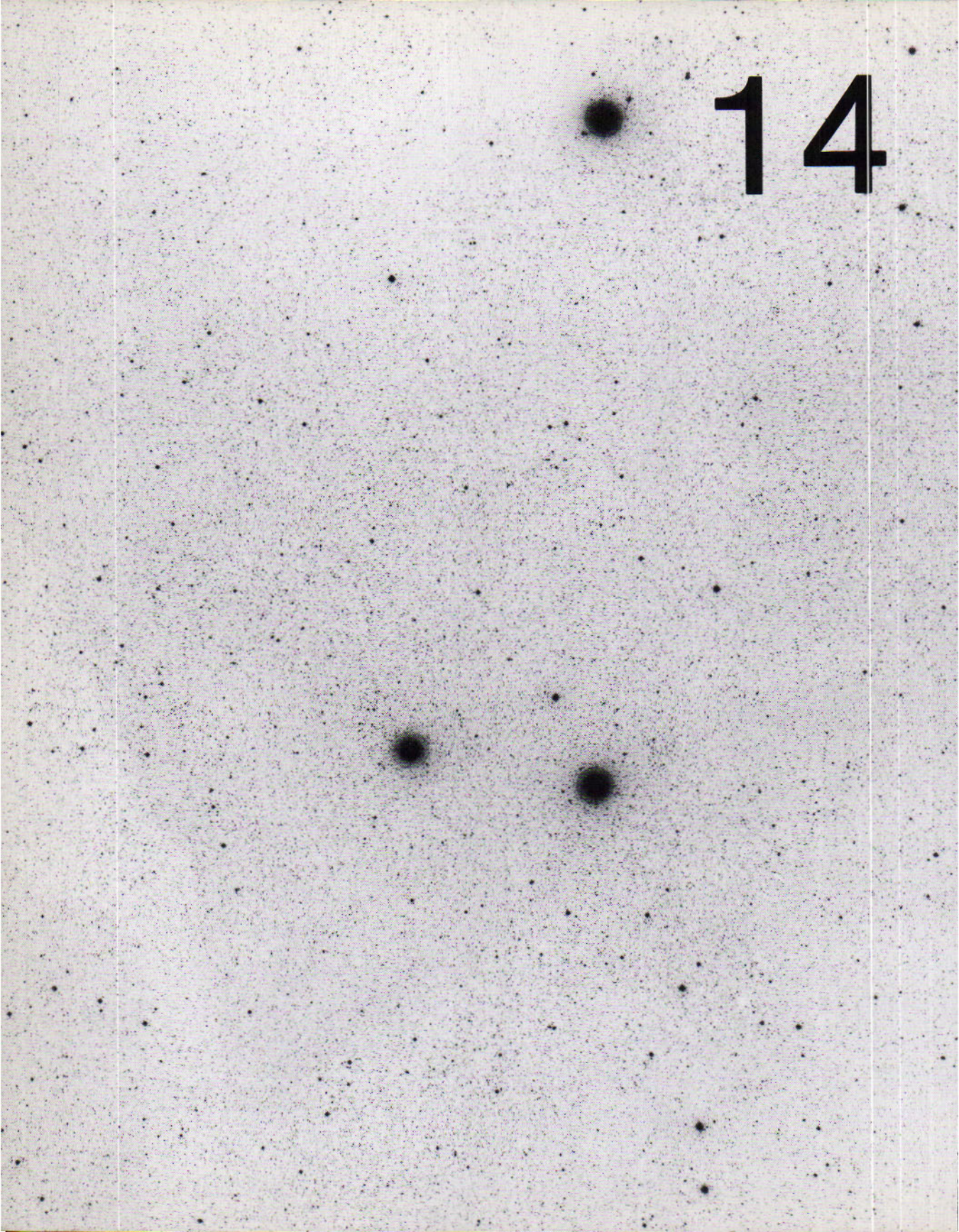
Payne-Gaposchkin, Cecilia, *Variable Stars and Galactic Structure,* London: Athlone Press, 1954.

Payne-Gaposchkin, Cecilia, *The Galactic Novae,* Interscience Publishers, New York, 1957

FOR FURTHER STUDY

Swihart, T. L., *Astrophysics and Stellar Astronomy,* New York: John Wiley and Sons, 1968

Installing the Dome on Louisiana State University's New Observatory. *(Louisiana State University Photograph)*


14

BINARY STARS

VISUAL BINARIES — SPECTROSCOPIC BINARIES — ECLIPSING BINARIES— ROTATIONS OF THE STARS

Binary stars are pairs of stars bound by their mutual gravitational fields. The connection is sometimes shown decisively by the mutual revolutions of the two stars, but is more often indicated only by their common proper motion. In the latter case the periods are presumably so long that the revolutions have not progressed far enough to be detected since the pairs were first observed.

Visual binaries are pairs of stars that can be separated with the telescope. *Spectroscopic binaries* appear as single stars with the telescope; their binary character is shown by periodic oscillations of the lines in their spectra. Many spectroscopic binaries have orbits so nearly edgewise to the earth that the revolving pairs undergo mutual eclipses and are *eclipsing binaries* as well.

The rotations of the stars are also described in this chapter because they were first detected in stars of eclipsing binaries.

Region of Zeta and Epsilon Aurigae. *(National Geographic Society and Mt. Palomar Sky Survey)*

VISUAL BINARIES

Optical double stars that are physically related are called visual binaries. Their study provided the first material for the masses of stars and the mass luminosity relation. Considerations as to the frequency of such stars leads to the belief that double and multiple star systems are the rule rather than the exception.

14.1 Optical and Physical Double Stars

The fact that certain stars which appear single to the unaided eye are resolved with the aid of the telescope into double stars was recorded casually by early observers, beginning with the discovery, in 1650, that Mizar in the handle of the Great Dipper is a double star. The reference is not to the naked-eye companion Alcor. The members of such pairs were generally but not always believed to appear close together only by the accident of their having nearly the same direction, until W. Herschel, searching for parallax, in 1803, reported that the components of Castor were in mutual revolution. He then made the distinction between "optical double stars" and "real double stars," the latter designation referring to two stars actually close together and united by the bond of their mutual gravitation.

Micrometric measuring of double stars was systematized by F. G. W. Struve at Dorpat, Russia, who in 1837 listed 3000 pairs. More recent surveys for the discovery of double stars have been made, particularly at Lick Observatory and the U.S. Naval Observatory for the northern hemisphere and at the University of Michigan southern station for the southern hemisphere. A total of 40,000 double stars, most of which are real double stars, are known. The pairs having the smaller separations, which are likely to progress more rapidly in their revolutions, can be observed satisfactorily only with the larger telescopes.

14.2 Measurements of Visual Binaries

In the past most visual binaries have been measured mainly with the position micrometer at the eye-end of the telescope. In this form of micrometer the thread is moved parallel to itself to measure angular distances and can also be rotated to measure directions in the field. The position of the *companion,* or fainter star of the pair, with respect to the *primary star* is obtained by measuring its position angle and distance. Such positions are employed to determine the apparent orbit of the companion relative to the primary.

The *position angle* ρ is the angle at the primary star between the directions of the companion and the north celestial pole; it is reckoned in degrees

from the north around through the east. The *distance* α is the angular separation of the two stars.

The smallest separation of a binary star that can be accurately measured with the 40-inch Yerkes refractor is 0″.2 and with the 82-inch McDonald reflector is about 0″.1, according to G. Van Biesbroeck. Photography with long-focus telescopes has completely supplanted the visual techniques, except for a few special cases, where the separation is greater than 2″ and for smaller separations by use of special methods such as image intensifier techniques and interferometers. From the photographs the similar orbits of the two components can be determined by referring the motions of the two stars to other stars called reference stars nearby in the field.

14.3 The Apparent and True Orbits

The *apparent orbit* of the companion relative to the primary star is the projection of the true orbit on the plane at right angles to the line of sight. This observed orbit is an ellipse and the law of areas is fulfilled by the line joining the two stars, but the primary star is not likely to be at the focus of the ellipse. The *true orbit* is calculated from the apparent orbit by one of several methods. It has the primary star at one focus and may be in any plane at all.

The *elements of the relative orbit* resemble the elements of a planetary orbit (7.26). They are: a, the semi-major axis of the orbit, or the mean distance between the stars, expressed in seconds of arc; T, the time of *periastron* passage, that is, when the stars are nearest; e, the eccentricity; i, the inclination of the orbit plane to the plane through the primary star at right angles to the line of sight; Ω, the position angle of the node that lies between 0° and 180°; ω, the angle in the plane of the true orbit between that node and the periastron point, in the direction of motion. P denotes the period of revolution in years.

When the parallax of the binary is known, the linear scale of the orbit can be found by the relation: a (in astronomical units) = α (in seconds of arc)/parallax. Everything is then known about the orbit, except which end is tipped toward us; it remains to be decided by the spectroscope whether the companion is approaching or receding from us when it passes the node.

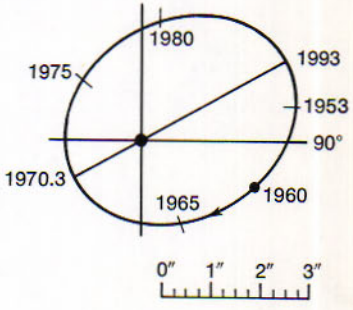


Figure 14.3
Orbit of Krüger 60. The relative apparent orbit of the fainter star as determined by R. G. Aitken.

14.4 Examples of Visual Binaries

More than 2500 visual binaries have already shown evidence of orbital motion. More than 10 per cent of these have revolved far enough since their discoveries to permit definitive determinations of their orbits. Some characteristics of a few binaries are given in Table 14.I, which is taken mainly from a more extended table by P. van de Kamp.

BD −8° 4352. This binary has the exceptionally short period of 1.7 years. The average separation of the two stars does not greatly exceed the earth's distance from the sun.

Delta Equulei. Also revolving in a very short period, the components have a separation less than Jupiter's distance from the sun.

TABLE 14.I
Orbits of Visual Binaries

NAME	VISUAL MAGNITUDES		PERIOD	SEMI-MAJOR AXIS	ECCENTRICITY	PARALLAX	MASSES	
			P	a	e	p	m_1	m_2
BD −8° 4352	9.7	9.8	1.7	0″.22		0″.157	0.6	0.3
Delta Equulei	5.2	5.3	5.7	0 .26	0.39	.056	2.0	1.1
42 Comae	5.0	5.1	25.8	0 .67	.52	.054	1.9	.9
Procyon	0.4	10.7	40.6	4 .55	.31	.287	1.9	.5
Krüger 60	9.8	11.4	44.5	2 .31	.41	.254	0.3	0.1
Sirius	−1.4	8.5	50.1	7 .62	.59	.376	2.4	0.7
Alpha Centauri	0.0	1.2	80.1	17 .66	.52	.760	1.3	0.6
Castor	2.0	2.8	420	6 .30	.37	.074	(3.4)	

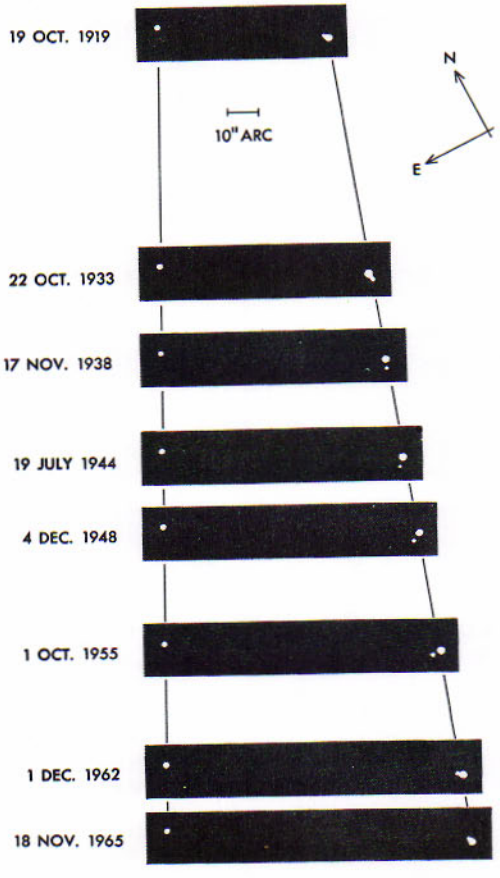


Figure 14.4
Composite of photographs of the binary star Krüger 60, referenced to the optical companion on the left. Note the effects of proper motion. Also note the parallactic effect on July 19, 1944 positions. *(Leander McCormick Observatory and Sproul Observatory composite diagram)*

42 Comae. The orbit is almost edgewise to the sun. It must not be precisely so, because eclipses have not been observed.

Krüger 60. This binary (Fig. 14.4) has an unusually small mass. The fainter component is a flare star (Fig. 13.15) and is probably degenerate.

Alpha Centauri. The nearest system to the sun, it was one of the first double stars to be discovered. The separation at periastron is a little more than Saturn's distance from the sun, and at apastron is midway between the mean distances of Neptune and Pluto from the sun.

Castor. This familiar double star was the first to be observed in revolution, although its period is nearly 4 centuries. The average separation (Fig. 14.7) is more than twice the mean distance of Pluto from the sun. At periastron in 1968 the two stars were 55 astronomical units apart.

14.5 Companions of Sirius and Procyon

The discoveries of the faint companions of Sirius and Procyon constitute the first chapter of what has been called the "astronomy of the invisible," namely, the detection of unseen celestial bodies by their gravitational effects on the motions of visible bodies. The discovery of Neptune (8.34) is another famous example. As in the case of Neptune, the companions of both stars were subsequently observed with the telescope.

F. W. Bessel, at Königsberg in 1844, announced that Sirius did not have the uniform proper motion that characterizes single stars, but was pursuing a wavy course among its neighbors in the sky. Having also found a similar fluctuation in the proper motion of Procyon, he concluded that both stars were attended by unseen companions and that the mutual revolutions of the pairs were causing waves in the proper motions of the

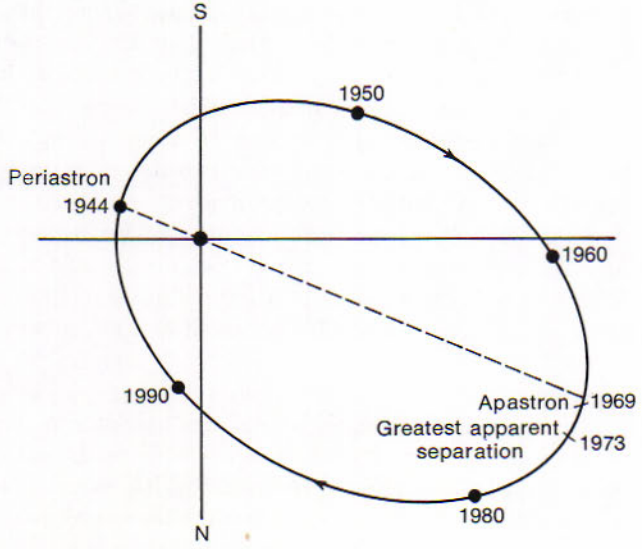


Figure 14.5
Apparent relative orbit of the companion of Sirius.

primary stars. Later, the orbits of both systems were calculated, although the companion stars had not yet been detected.

The companion of Sirius was first observed in 1862, by Alvan Clark, a telescope maker who was using the bright star for testing the 18½-inch refractor whose lens is now at Dearborn Observatory. Despite the brilliance of its primary, the 8th-magnitude companion is not difficult to see with a large telescope, except near its periastron. It reaches maximum apparent separation in 1973 and is easily visible. This star was among the first-known examples of the very dense white dwarf stars. The companion of Procyon proved to be more elusive; it was finally seen at Lick Observatory in 1896.

14.6 Unseen Companions

Variable proper motions of other apparently single stars have been discovered in recent times by photographic means. Such stars are referred to as *astrometric binaries.* In one case the fainter companion has since been observed, that of the 11th-magnitude red dwarf star Ross 614 at the distance from us of 13 light years. Calculation of the orbit by S. L. Lippincott showed that the pair would be most widely separated in 1955. The 15th-magnitude companion was then seen and photographed by W. Baade with the 200-inch telescope. Both stars have very small masses.

There are frequent cases where periodic perturbations of the revolutions of visual binary stars reveal the presence of third bodies in the systems. An example is the binary 61 Cygni, which has been studied by K. Aa. Strand. Here the two visible members revolve in a period of 720 years, while an invisible companion revolves around one of them in a period of nearly 5 years. There are also cases where the orbits of spectroscopic binaries were found to be disturbed and where the spectrum of a third star of the system was subsequently detected. An example is the eclipsing star Algol (14.21). Another almost classic case is that of the discovery of a perturbation in the motion of Barnard's Star (the second nearest star system) announced by P. van de Kamp in 1961. He used almost 700 plates taken over a 40 year interval to detect an oscillation caused by an object no larger than twice the mass of Jupiter. He has even suggested that the periodic motion may be caused by two objects instead of just one. We will consider later whether or not an object having twice the mass of Jupiter is really a planet or a degenerate star. Indeed, there are astronomers who hold that even Jupiter itself is a degenerate star.

14.7 Multiple Systems

The presence of more than two stars in what were originally thought to be binary systems is not exceptional. An estimate that 5 per cent of visual binaries are at least triple systems is considered conservative. A common type of triple system is represented by Alpha Centauri, where the binary is attended by a remote companion, Proxima. A similar pattern is found in the system of Castor, except that each of the three stars is a spectroscopic binary.

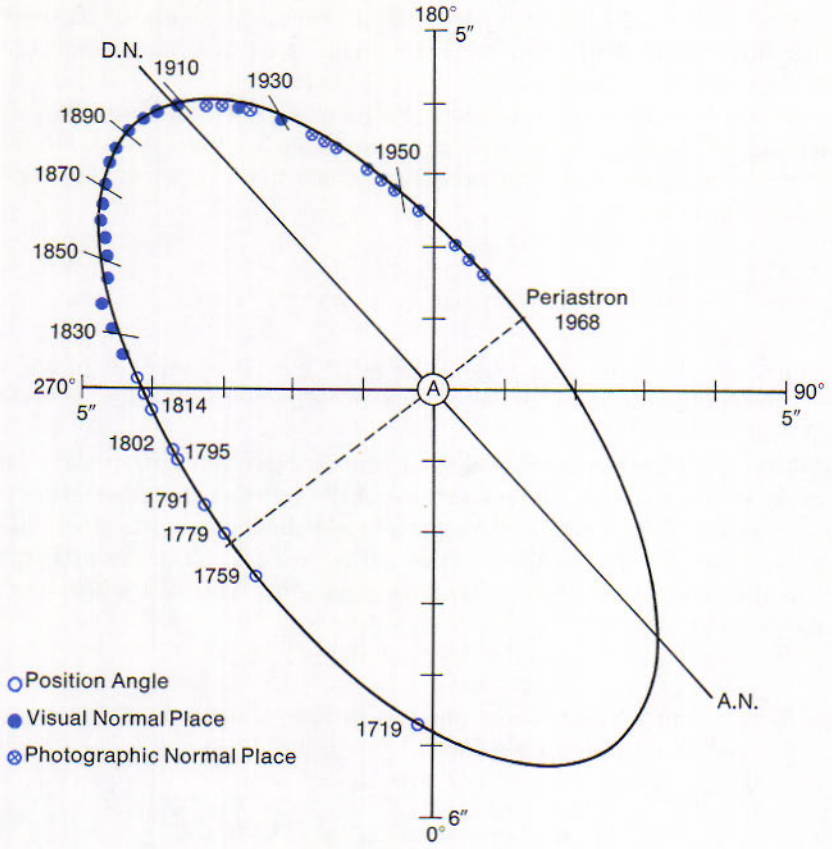


Figure 14.7
Apparent orbit of Castor. *(Determined by K. Aa. Strand)*

The familiar double-double Epsilon Lyrae is typical of quadruple systems. Two moderately wide pairs of stars mutually revolve in a period of several hundred thousand years. Another interesting example is Zeta Cancri, recently studied by C. E. Gasteyer at Dearborn Observatory. Two pairs of stars revolve in periods of 59.7 and 17.5 years, respectively, and move around a common center once in 1150 years. One star of the latter pair, not observed with the telescope, is recognized by the 17.5-year revolution of its companion. Since the pair is detected by its perturbed motion noticeable in the photographic plates, it is referred to as an astrometric binary (14.6).

14.8 Masses of Visual Binaries

The mass of a celestial body can be determined whenever its gravitational effect on the motion of another body a known distance away is appreciable. This method does not apply to single stars, which are too far removed from other stars to have their motions affected noticeably; nor to the majority of visual binary stars, which have thus far given no evidence of mutual revolution. It can be used to evaluate the combined masses of

binary systems that have progressed far enough in their revolutions so that their orbits have been calculated. This is an important result of the studies of visual binaries.

By the restatement of Kepler's harmonic law (7.22) the sum of the masses, m_1 and m_2, of the two components of a binary system, in terms of the sun's mass (the earth's relatively small mass is neglected), is given by the relation:

$$m_1 + m_2 = \frac{a^3}{P^2 p^3},$$

where a is the semimajor axis of the relative orbit in seconds of arc, P is the period of revolution in sidereal years, and p is the parallax in seconds of arc.

The sum of the masses is all that can be determined from the relative orbit. When, however, the revolutions of the two stars have been observed with reference to neighboring stars in the field, the individual masses become known. The center of mass of the system is thereby established, and the ratio of the masses is inversely as the ratio of the distances of the two stars from this point.

$m_1 a_1 = m_2 a_2$
$a_1 + a_2 = a$

As an example of the use of the above relation, the sum of the masses of Sirius and its companion is calculated from the data in Table 14.I as follows:

$$m_1 + m_2 = \frac{(7.62)^3}{(49.9)^2(0.379)^3} = \frac{443}{136} = 3.3$$

The combined mass of Sirius and its companion is 3.3 times the sun's mass.

14.9 Mass-Luminosity Relation

Studies of binary stars led to the discovery by A. S. Eddington, in 1924, of a simple and useful relation between the masses and luminosities of stars in general. The more massive the star, the greater is its absolute brightness. In Fig. 14.9, the logarithms of the masses of a number of stars in binary systems in terms of the sun's mass are plotted against their absolute bolometric magnitudes. *Bolometric magnitude* refers to the radiation of a star in all wavelengths, which can be derived from the visual absolute magnitude and spectral type with allowance for the absorption of the starlight by the earth's atmosphere. The mass of a single star may be read from this curve if the absolute bolometric magnitude is known and if the star is of the type that conforms to this relation.

Giant and main-sequence stars show in general a close agreement with the mass-luminosity relation. Special groups of stars do not conform.

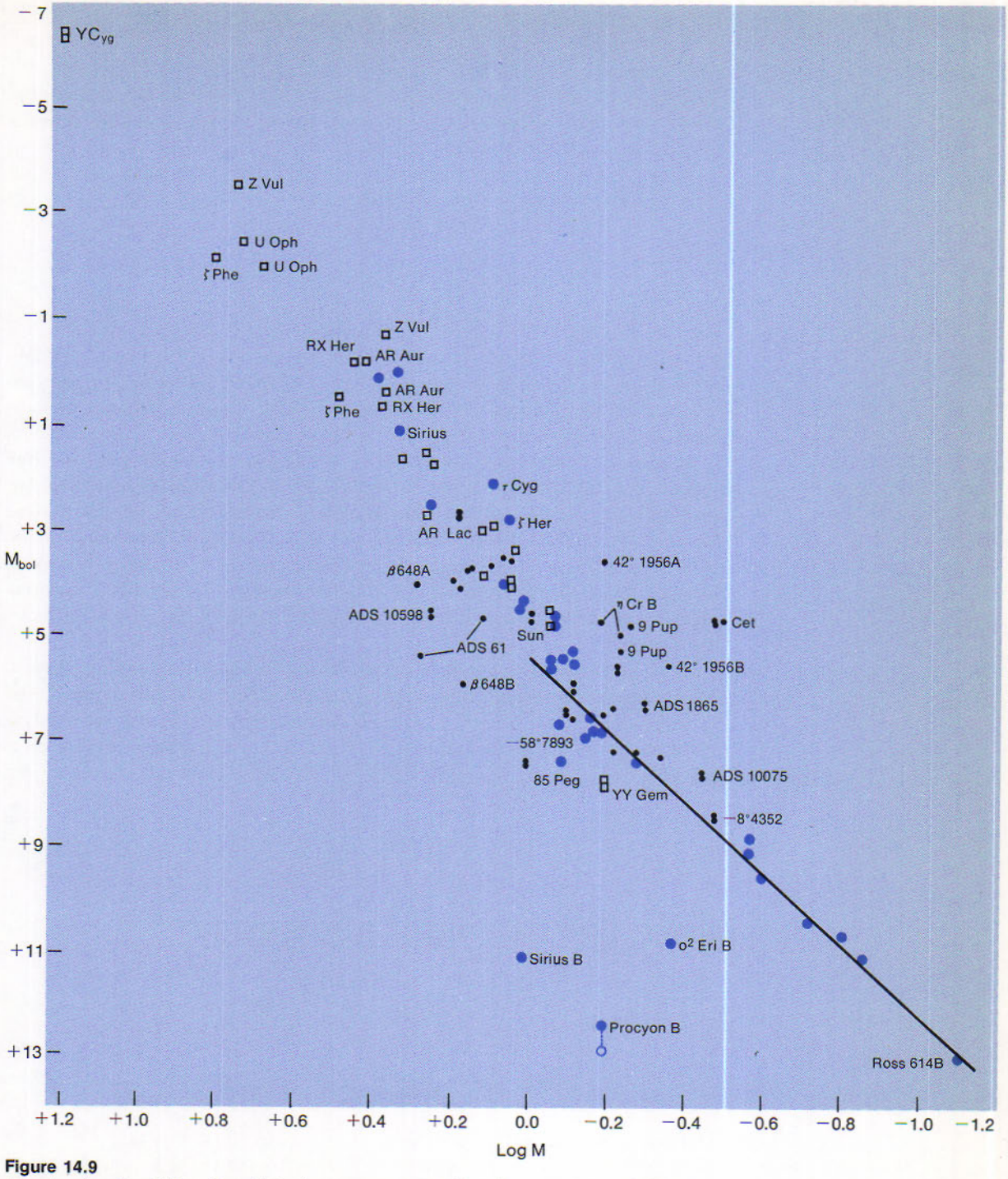


Figure 14.9
Mass-luminosity relation. Visual binaries are represented by dots, spectroscopic binaries by squares. *(After D. L. Harris III, K. Aa. Strand, and C. E. Worley in Stars and Stellar Systems, Basic Astronomical Data, by permission)*

Among these are a number of supergiants and white dwarf stars. Compared with main-sequence stars of equal mass, the white dwarfs are the fainter by several magnitudes.

14.10 Dynamical Parallaxes

The formula of Section 14.8 in the form:

$$p^3 = \frac{a^3}{P^2(m_1 + m_2)}$$

may be employed to determine the parallaxes of binary systems when the combined masses of the component are already known. An approximate value of the sum of the masses can be used, because the parallax is inversely proportional to the cube root of this sum. Parallaxes of systems so determined are known as *dynamical parallaxes.*

Because the combined mass of a binary system is usually not far from twice the sun's mass, a preliminary value of the parallax is found by putting $m_1 + m_2 = 2$ in the preceding formula, when a and P are known. Given the apparent magnitudes of the two stars and their preliminary distances from us, their absolute magnitudes are calculated, and more nearly correct masses are obtained from the curve of Fig. 14.9. The required parallax of the system is finally found by substituting the new masses in the formula.

By a procedure that depends primarily on the mass-luminosity relation, H. N. Russell and Charlotte Moore calculated the dynamical parallaxes of more than 2000 visual binary systems. Such parallaxes for appropriate systems having well-defined orbits have statistical probable errors of 5 per cent.

SPECTROSCOPIC BINARIES

Binary stars having their components so close together that they appear as single stars with the telescope are discovered and studied in photographs of their spectra. Unless their orbits are at right angles to the line of sight, the revolving stars alternately approach and recede from the earth. The lines in the spectra are displaced by the Doppler effect (4.9) to the violet in the first case and to the red in the second, so that they oscillate in the periods of the

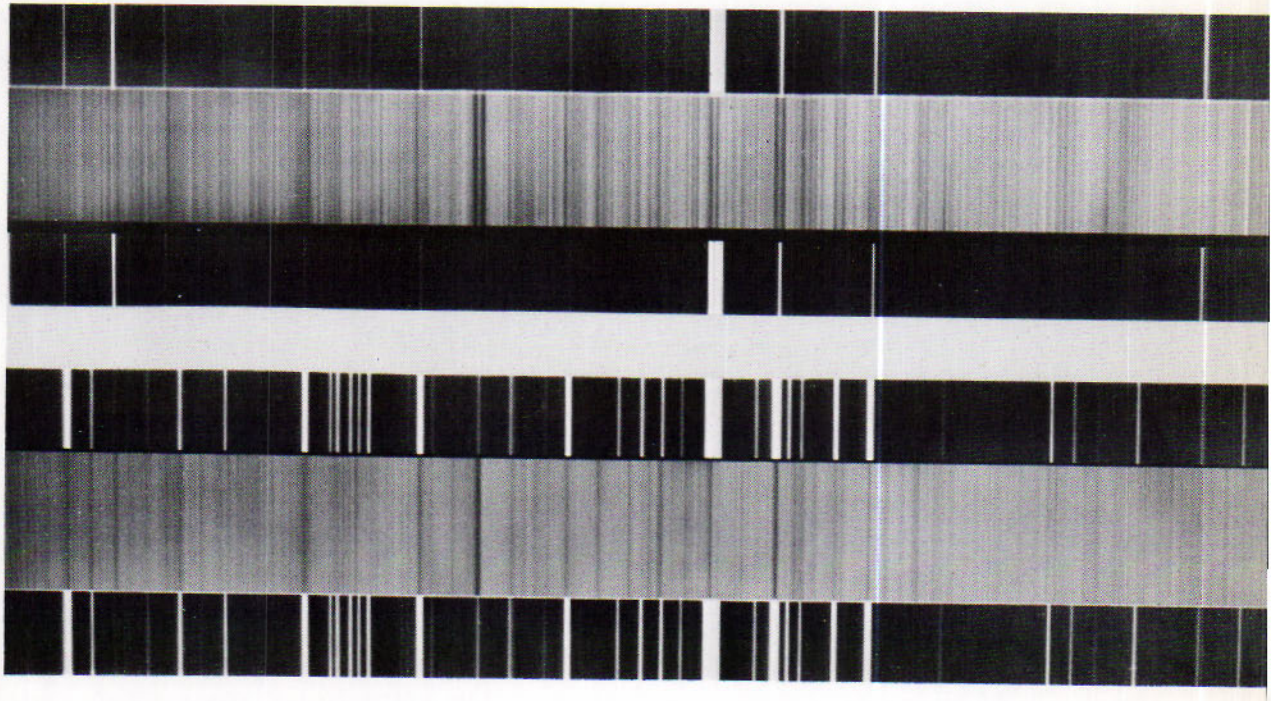
revolutions. Spectroscopic binaries are not to be confused with pulsating stars (13.7), where the spectrum lines also oscillate.

14.11 Oscillations of Spectrum Lines

The brighter component of Mizar, the first visual double star to be reported, was the first spectroscopic binary to become known, in 1889. The lines in the spectrum of this star were found to be double in some objective prism photographs at Harvard Observatory, whereas they were single in others.

Mizar is an example of spectroscopic binaries having components about equally bright and of the same spectral type. The spectrum shows two sets of lines that oscillate in opposite phase. When one star is approaching us in its revolution, the other star is receding from us. The lines in the spectrum of the first star are displaced to the violet, while those of the second star are displaced to the red, and the lines appear double. About a quarter of the period later, when both stars are moving across the line of sight, the lines of the two spectra have no Doppler displacements caused by the revolutions and are accordingly superposed.

Figure 14.11A
Spectrograms of Mizar. The lines of the two components are separated in the upper photograph and superposed in the lower one. *(Yerkes Observatory Photograph)*



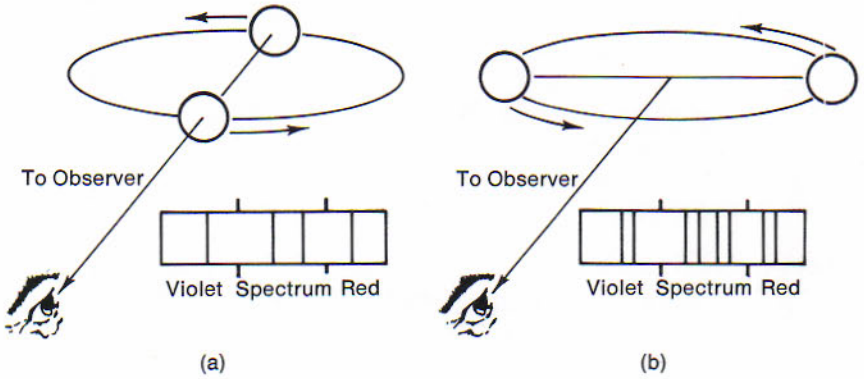


Figure 14.11B
Doppler displacements in the spectrum of Mizar. (a) When one star approaches and the other recedes from us, the lines are double. (b) When both stars are moving across the line of sight, the two sets of lines are undisplaced by the revolution and are superposed.

If one member of a pair is as much as 1 magnitude brighter than the other, which is true of most of these binaries, only the spectrum of the brighter star is likely to be visible. The periodic oscillation of the lines in the spectrum shows that the star is a binary. Examples of spectroscopic binaries among the brighter stars are Spica, Capella, Castor, and Algol.

14.12 Velocity Curve

Once the effect of the earth's motion is removed from the observations, the plot of the resulting data is called the *velocity curve* of one of the stars of a spectroscopic binary, and it shows how the velocity of the star in the line of sight varies during a complete revolution. The smooth curve

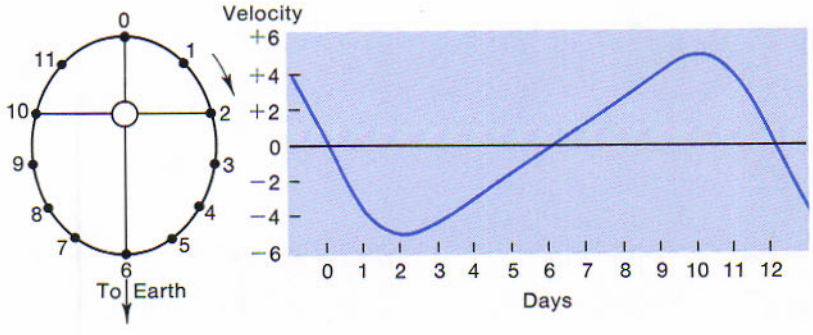


Figure 14.12A
Relation between orbit and velocity curve of a spectroscopic binary. The period of the revolution is 12 days. Only one spectrum appears. The star approaches the earth from phase 0 to 6 days. Maximum radial velocities occur at 2 and 10 days, when the star is crossing the "plane of the sky."

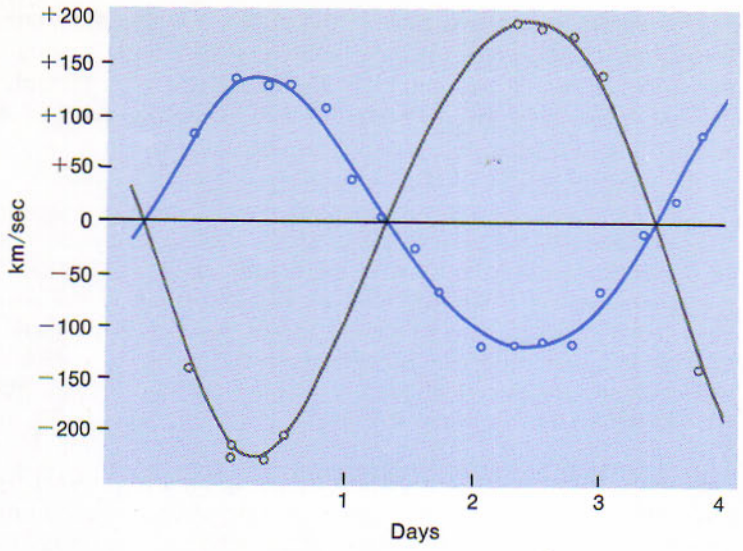


Figure 14.12B
Velocity curves of Spica. Both spectra were observed, but the lines of the fainter component were clearly seen only near the times of greatest separation of the lines. *(Publications of the Allegheny Observatory)*

plotted represents the observed radial velocities of a given star during different phases of its revolution. These radial velocities are calculated from the displacements of the spectrum lines measured in photographs of the different phases.

Sine curves represent the radial velocities of circular orbits. In the case of elliptical orbits the form of the velocity curve depends on the eccentricity of the ellipse and also on its orientation when projected on a plane passing through the line of sight. Fig. 14.12A shows the form of the velocity curve for an ellipse of moderate eccentricity, having its projected major axis directed toward the earth. Velocity of recession is denoted by the plus sign and of approach by the minus sign.

Conversely, when the velocity curve is known, the projected orbit of the star can be calculated by an appropriate method. Whenever the lines of both spectra are visible in the photographs, it is possible to determine the velocity curves (Fig. 14.12B) and projected orbits of both components of the binary.

14.13 Orbits of Spectroscopic Binaries

It is the projection of the orbit on a plane through the line of sight that is calculated from the velocity curve. The inclination, i, of the orbit plane to the plane of the sky cannot be determined from the spectrum. The semi-major axis, a, of the orbit is derived in combination with the inclination in the quantity $a \sin i$. Thus the actual size of the orbit remains unknown unless it can be found separately in another way; it can be found if the double star is an eclipsing binary. The other elements of the orbit are determined uniquely from the radial velocities excepting Ω.

The masses of the two stars of the binary are also uncertain unless the inclination of the orbit plane is known. When both spectra are visible, the value of $(m_1 + m_2) \sin^3 i$ is found by use of a formula similar to the one previously given (14.8). In such cases the ratio of the masses, m_1/m_2, is taken from the velocity curves; it varies inversely as the ratio of the velocity ranges of the two stars.

As an example, the velocity range of the brighter component of Spica (Fig. 14.12B) is 252 km/sec, and that of the fainter star is 416 km/sec. The ratio of the masses, m_1/m_2, is 416/252, or 1.6. The calculation of the projected orbits gives $(m_1 + m_2) \sin^3 i = 15.4$ times the sun's mass. If the inclination of the orbits to the plane of the sky is not far from 90°, $m_1 + m_2 = 15.4$; and the mass of the brighter component of Spica is 9.6 times the sun's mass, while the fainter star is 5.8 times as massive as the sun.

Studies of the spectra of certain eclipsing binaries (14.23) by O. Struve and others have shown that the revolving stars are surrounded by gas streams, which are going around at different speeds from those of the stars themselves. Dark lines abstracted from the starlight by the swirling gases blend with the dark lines in the spectra of those stars. Unless the effects of the gas streams around such binaries are allowed for, the radial velocities from the confused lines and the orbits derived from them are subject to considerable error.

14.14 Some Interesting Spectroscopic Binaries

Most eclipsing binaries, which we are about to consider, are spectroscopic binaries as well and are very interesting. Here we have selected three interesting systems exhibiting special effects.

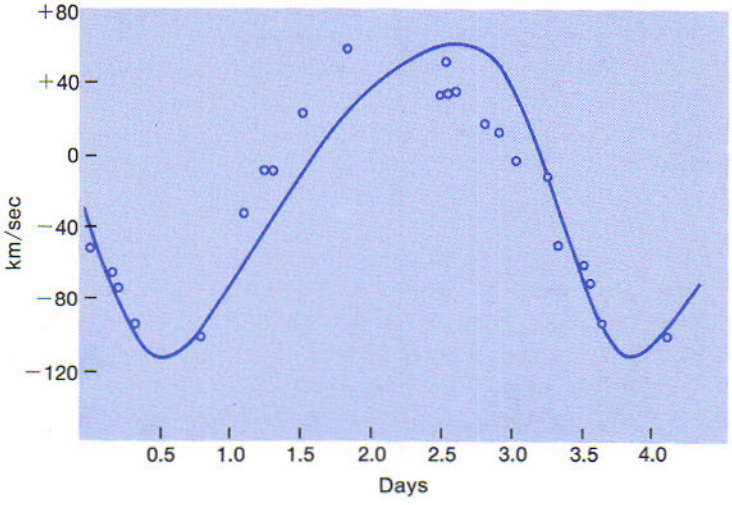


Figure 14.14A
Velocity curve of single-lined binary HR8800. Note change from 1919 (line) to 1957 (points) due to rotation of the line of apsides. *(Adapted from R. M. Petrie)*

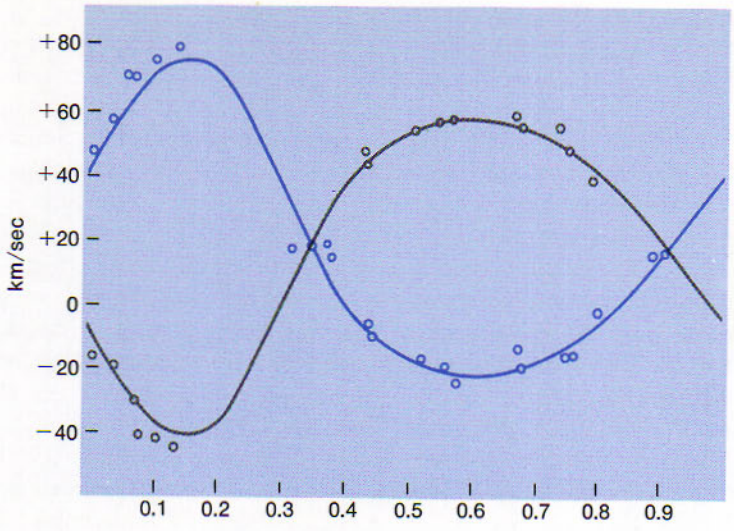
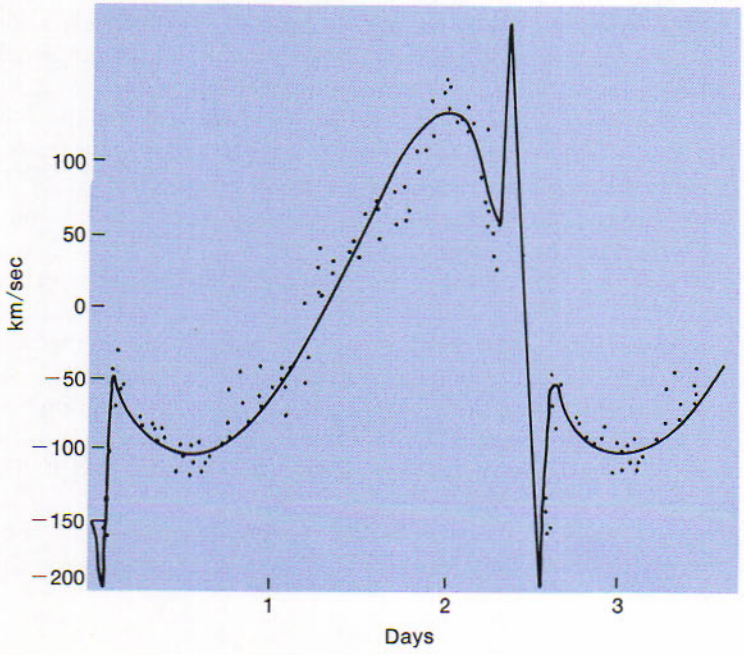


Figure 14.14B
Double-lined binary HD6619 having almost identical components. *(Adapted from D. S. Evans)*
Figure 14.14C
Velocity curve of u Cepheid showing the Rossiter effect.



HR8800. This binary is a classic single line binary clearly showing rotation of the line of the apsides. The system has a period of 3.3 days and in figure 14.14A we have shown the velocity curves obtained 38 years apart.

HD6619. This system is a double lined binary, much like Spica mentioned earlier, but demonstrates what the velocity curves look like when the two components are essentially identical. The orbits have excentricities somewhat larger than 0.2 (Fig. 14.14B).

U Cephei. This is the classic case displaying the Rossiter effect. When the visible component is about to be eclipsed, we see the rotational velocity of the trailing crescent of the star spinning away from us and then we see the approaching limb just after the leading edge appears from behind the second star. All of this is superposed on the normal velocity curve of the visible component (Fig. 14.14C).

14.15 Features of Binary Systems

Among the characteristics of binary systems that might seem to provide clues as to their origin and development, we note the following:

1. The great number of such systems informs us that the process which develops them is not unusual. As an example, very nearly half of the 59 stars within 16 light years from the sun are double or triple systems.
2. The great variety in their separations. There is a gradation from rapidly revolving pairs almost in contact to binaries having so widely separated components that the only observed connection between them is their common motion through space.
3. The correlation between eccentricity of orbit and length of period. There is a fairly steady statistical increase in eccentricity from nearly circular orbits of binaries having periods of a few hours or days to values around 0.7 for pairs with periods of revolution running into many hundreds of years.
4. The swirling gas streams surrounding many close spectroscopic binaries.
5. The systematic difference in spectral type of the components of visual binaries, first reported, in 1923, by F. C. Leonard. If both stars have the same brightness, their spectral types are the same. If they are main-sequence stars of different brightness, the brighter star is generally the bluer; if they are giants, the brighter star is likely to be the redder of the two.
6. The systematic difference in the speeds of rotation of the components of visual binaries (14.26).

14.16 The Problem of Their Origin

Their origin has long been debated. The *fission theory,* as explained by J. H. Jeans, G. H. Darwin, and others, was favored in former times, and some recent efforts have been made to revive it in amended form. The theory begins with a single star shrinking under the action of gravity, thus progressively rotating faster and increasing its equatorial bulge until the star approaches instability. We note presently that the rotations of some of the hotter stars are almost fast enough to threaten their stability.

When a critical stage in the process is reached, the theory based on a conventional model suggests that the star's equator might begin to be drawn out into an ellipse. Later, the star might assume the form of a dumbbell, or pear, and finally divide. The resulting pair of stars would at first rotate and revolve in the same period; but further shrinkage could put the two motions out of step, preparing the way for effects of tidal friction. The components might then increase their separation and period of revolution, but to a limited extent unless the binary is supplied in some way with many times its initial angular momentum.

The *separate nuclei theory* carries the problem back to early stages in the formations of stars from turbulent cosmic clouds. It supposes that binary-star patterns are set by mutually revolving proto-stars condensing from the clouds. Actually, the system that condenses from the interstellar medium is a cluster of multiple star systems. The binary stars that we see are the remnants of the earlier clusters. The multiple star system of three or more bodies has a very limited lifetime, cosmically speaking, unless its components have very special masses on very special orbits.

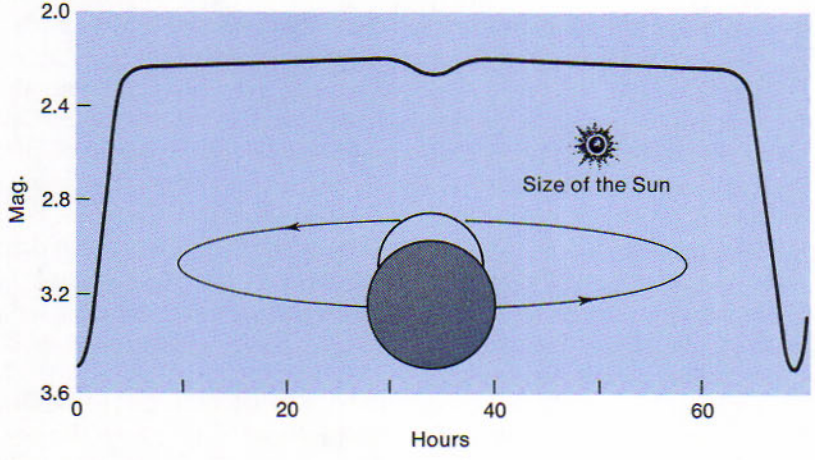


Figure 14.16
Light curve and system of Algol. The size of the sun is shown on the same scale. *(Light curve and orbit by J. Stebbins)*

ECLIPSING BINARIES

A spectroscopic binary is also an eclipsing binary when its orbit is so nearly edgewise to the earth that the revolving stars undergo mutual eclipses. Because the system appears as a single star with the telescope, what is observed is that the light of the star becomes fainter at regular intervals. Eclipsing binaries are variable stars only because the planes of their orbits happen to pass nearly through the earth's position. To an observer in another part of our galaxy their light might be practically constant, and another group of spectroscopic binaries, to us invariable in brightness, would exhibit eclipse phenomena.

14.17 Light Variations

The varying light from these stars is observed by means of a photoelectric photometer. Extremely accurate timing of the observations is essential, and all observations are reduced to heliocentric universal time, i.e. as if they were made at the center of the solar system, in order to remove the light time effect. This makes it easy for observers to use each others' material, facilitating the study of many systems being observed almost continuously around the world. These international cooperative efforts are encouraged by the International Astronomical Union. The light curve of an eclipsing binary shows how the magnitude of the system varies through a complete revolution. Twice during a revolution the curve drops to a minimum and rises again. The deeper minimum, or *primary minimum,* occurs when the star having the greater surface brightness is being eclipsed; the shallower one, or *secondary minimum,* occurs when that star is eclipsing its companion.

Even when the eclipses are not occurring, the light continues to vary appreciably in many of these systems, mainly because the stars are ellipsoids. They are elongated by mutual tidal action, each of the pair in the direction of the other, a relation that is maintained by the equality of their periods of rotation and revolution. During the eclipses the stars are seen end-on; halfway between the eclipses they are presented broadside, so that their disks are larger and the stars are accordingly brighter. Thus the light of the system rises to maxima midway between the minima (Fig. 14.17A). This effect becomes especially conspicuous when the two stars are almost in contact.

The hemispheres of the two stars that face each other are made brighter by the radiation of the other star. The difference is greater for the two hemispheres of the less luminous star, so that the light curve is higher near the secondary minimum (Fig. 14.17). In the more widely separated

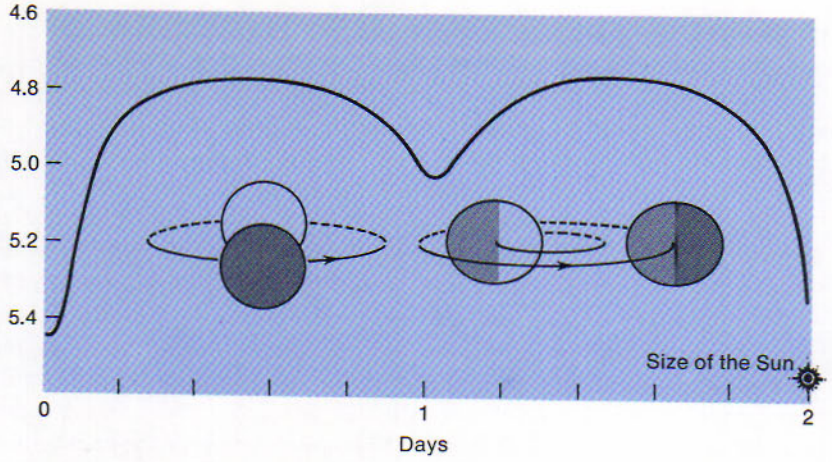


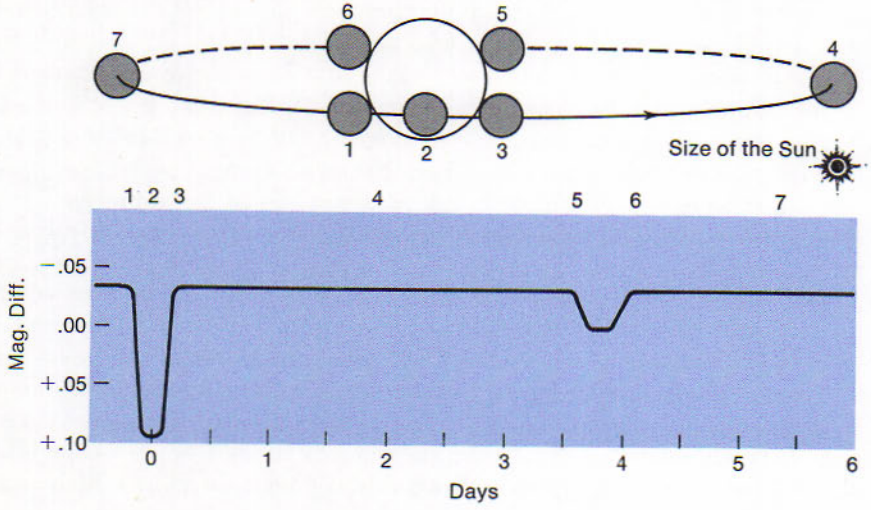
Figure 14.17
Light curve and system of the eclipsing binary u Herculis. The binary is shown at primary minimum and greatest elongation.

eclipsing binaries both ellipticity and radiation effects are so slight that the light curves outside the eclipses are practically horizontal (Fig. 14.18).

14.18 Light Variations During Eclipses

The eclipses of binary stars, like eclipses of the sun, may be total, annular, or partial. During a total eclipse the light remains constant at the minimum, and the duration of this phase is longest as compared with the whole eclipse when the orbit is edgewise to us and when the two stars differ greatly in size. That the light does not vanish during totality (the

Figure 14.18
Apparent orbit and light curve of the eclipsing binary 1 H. Cassiopeiae. The primary eclipse, of the bright star by its smaller companion, is annular. The secondary eclipse is total. Tidal and reflection effects are inconspicuous, because of the wide separation of the stars. *(Light curve and orbit by J. Stebbins)*



greatest observed decrease is 4 magnitudes) shows that the larger star is not dark, although it is usually the fainter of the two. When an eclipse is annular, the light may not remain quite constant because the eclipsed star, as in the case of the sun, is likely to be somewhat less bright near the edge than at the center. An *occultation* takes place when the larger star is in front of the smaller one. The opposite is called a *transit.*

When an eclipse is partial, or is total or annular only for an instant, the curve drops to its lowest point and begins to rise at once. In general, the depths and shapes of the light curve during eclipses depend on the relative size and brightness of the two stars and on the inclination of the orbit. The secondary minimum is scarcely discernible in some systems, whereas in others it may equal the primary minimum in depth. The fraction of the period in which the eclipses are occurring depends on the ratio between the sum of the radii of the two stars and the radius of the orbit. If the fraction is large, the stars are revolving almost or actually in contact.

14.19 Photometric Orbit

For any model of an eclipsing system, in which the inclination of the orbit is specified, it is possible to predict the form of the light curve. Conversely, when the light curve is determined by photometric observations, it is possible to calculate the elements of the orbit and the dimensions of the two stars in terms of the radius of the relative orbit. The analysis proceeds in stages. The observed light curve is first freed of the various effects such as the elliptricity of the stars mentioned earlier. This "rectified" light curve is then fit by a computed curve involving the various geometrical factors and the limb darkening, but not the inclination. After getting a good fit it is then possible to go back and determine the inclination. In reality, the whole operation can now be done in one grand step in a high speed computer.

If the spectroscopic orbit of each star is also known, we can return to it and supply the value of i in the expressions $a \sin i$, $m_1 \sin^3 i$, and $m_2 \sin^3 i$ (14.13), thus separately determining the radius of the relative orbit and the masses of the stars. Going back to the photometric orbit, in which the dimensions are derived in terms of the radius of the orbit, we have finally the absolute dimensions of the stars themselves. Thus the combination of the photometric and spectroscopic orbits permits the evaluation of the sizes and masses, and therefore the densities of the stars—data of great value in studies of the constitution of the stars.

In the study of eclipsing systems we have an example of the power of astronomical research. The largest telescope shows any one of these systems only as a point of light fluctuating periodically in brightness. Yet the observations of this light with the photometer and spectroscope and the judicious use of analysis lead to fairly complete specifications of the remote binary systems.

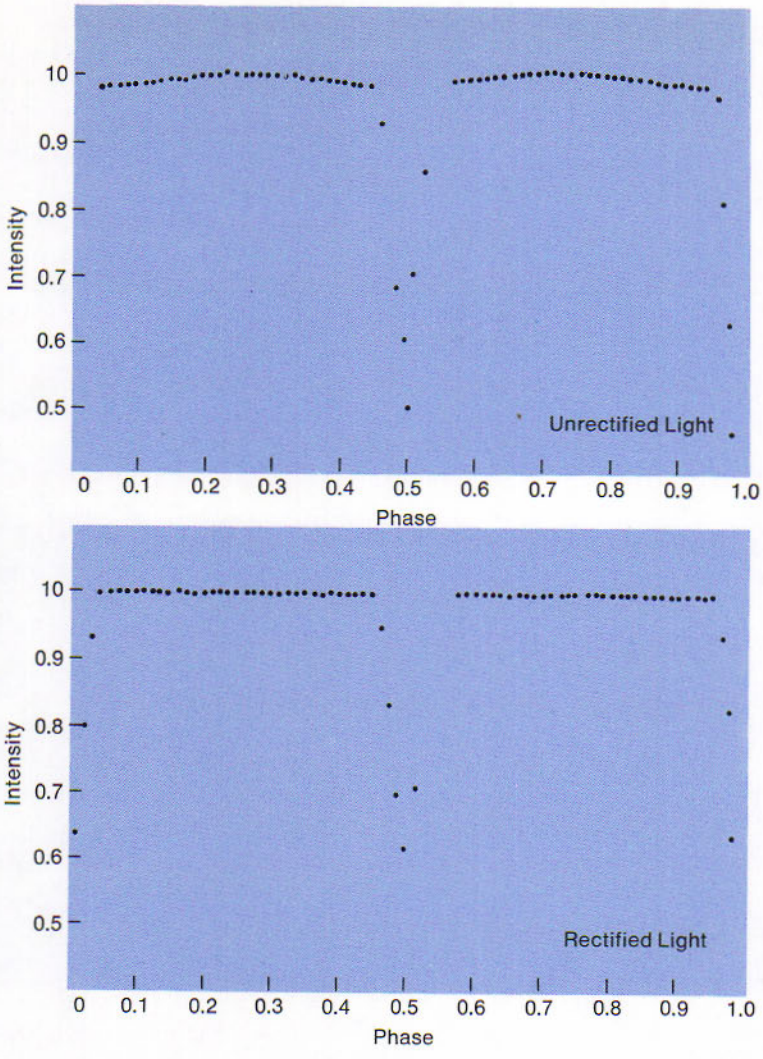


Figure 14.19
Observed and rectified light curve of WW Aurigae. Each point is the average of at least 25 observations.

14.20 Eclipsing Binaries with Elliptic Orbits

Most of these systems have nearly circular orbits. In cases where the orbits are considerably eccentric, additional effects are observed in the light curves, which follow from the law of equal areas (7.15) in its general form; the stars revolve faster near periastron. Two effects are as follows:

1. The two eclipses are of unequal durations. The eclipse that occurs nearer periastron is the shorter. The difference is greatest (b) when the major axis of the orbit is directed toward the earth.

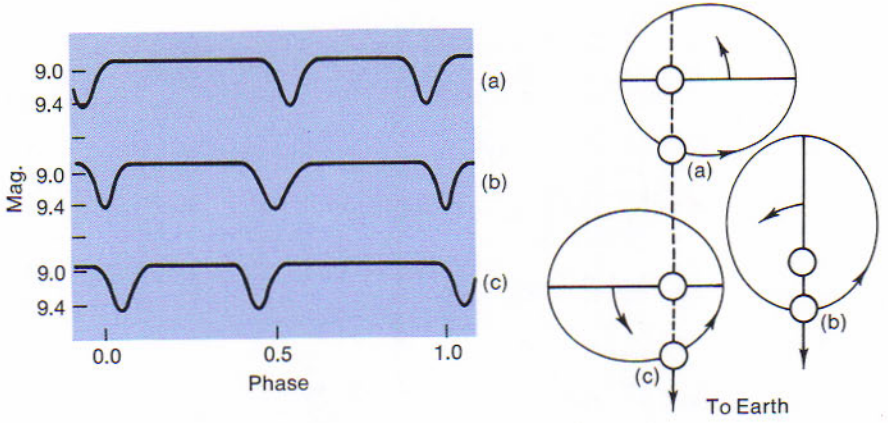


Figure 14.20A
Light curves and orbit of the eclipsing binary GL Carinae. The major axis of the orbit rotates in a period of 25 years. Note the difference in the light curves corresponding to three different directions of the major axis. *(From data by H. Swope, Harvard Observatory)*

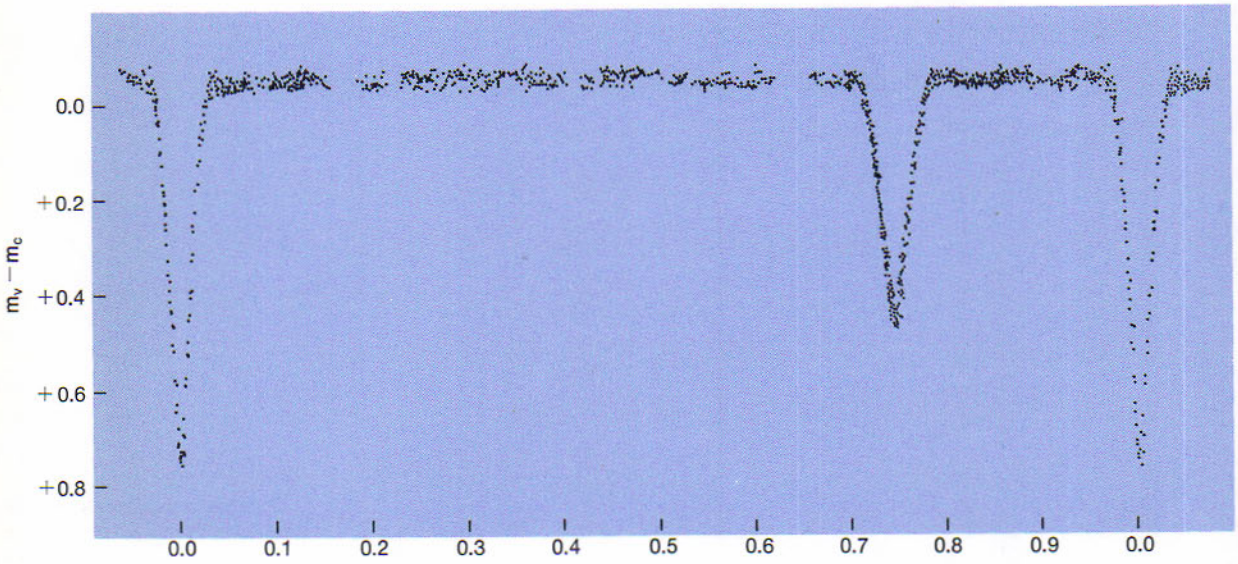


Figure 14.20B
The observed light curve of FT Orionis. The large displacement of secondary minimum leads to an orbital eccentricity of 0.4. The period of FT Orionis is 3.15 days. *(Diagram by S. Cristaldi as presented in Astronomy and Astrophysics, 1970)*

2. The intervals between the minima are unequal. The interval including periastron passage is the shorter. This difference is greatest, as for (*a*) and (*c*), when the major axis of the orbit is perpendicular to the line of sight.

In some of these systems the major axis of the orbit rotates rather rapidly in the direction the stars revolve, although in most systems the period of this rotation runs into hundreds of thousand years. The advance of periastron is caused by the oblateness of the stars. Thus in the period of rotation of the axis the light curve exhibits a cycle of the two effects that have been described. As an example, the major axis of the orbit of

GL Carinae rotates in a period of 25 years. The primary and secondary minima happen to have the same depth in the light curves of this system (Fig. 14.20A).

Other noteworthy features of certain eclipsing binaries are mentioned in a few sections that follow.

14.21 Some Interesting Eclipsing Binaries

The periods of eclipsing binaries range from 80 minutes in the case of WZ Sagittae to 27 years for Epsilon Aurigae, and are frequently around 2 or 3 days. Intensive study has made a number of these binaries known to all astronomers. The three discussed here are bright enough to be observed with very modest equipment.

Algol, the "Demon Star", is the most familiar of the eclipsing binaries and was the first of this type to become known. The discovery that its light diminishes at intervals of about 2 days and 21 hours was made as early as 1783, and the theory was then proposed that the bright star is partially eclipsed by a faint companion revolving around it in this period. The correctness of this view was established in 1889, when a spectroscopic study of Algol showed that it is a binary and that the radial velocity of the bright star changes from recession to approach at the time of the eclipse, as the theory required.

The brighter, B8V component of Algol has 3 times the diameter of the sun. The G8III companion is fainter by 3 magnitudes, but its diameter is 20 per cent greater than that of the brighter star. The centers of the two stars are 21 million km apart, or slightly more than a third of the average distance of Mercury from the sun. Their orbits are inclined 8° from the edgewise position relative to the earth. Once in each revolution the companion passes between us and the bright star, partially eclipsing it for nearly 10 hours and reducing the light of the system at the middle of the eclipse to a third its normal brightness. The slight decrease in the light midway between the primary eclipses occurs when the companion is partially eclipsed by the brighter star. A third star, revolving around the two in a period of 1.87 years, contributes a number of narrow lines to the spectrum of the system.

The eclipsing binary *Zeta Aurigae* is one that permits the study of a star's atmosphere at different levels. It consists of a B8 star considerably larger than the sun and a K5 supergiant star. The two stars have nearly the same photographic magnitude. They mutually revolve once in 972 days, or about 3 times in 8 years. The principal eclipse, of the blue star by the red one, is total for 37 days preceded and followed by partial phases, each lasting for 32 hours.

The most remarkable features are presented during the week before the eclipse begins and the week after it ends, while the smaller star is passing behind the atmosphere of the larger one. Fig. 14.21A shows spectra of Zeta Aurigae photographed by D. B. McLaughlin from January

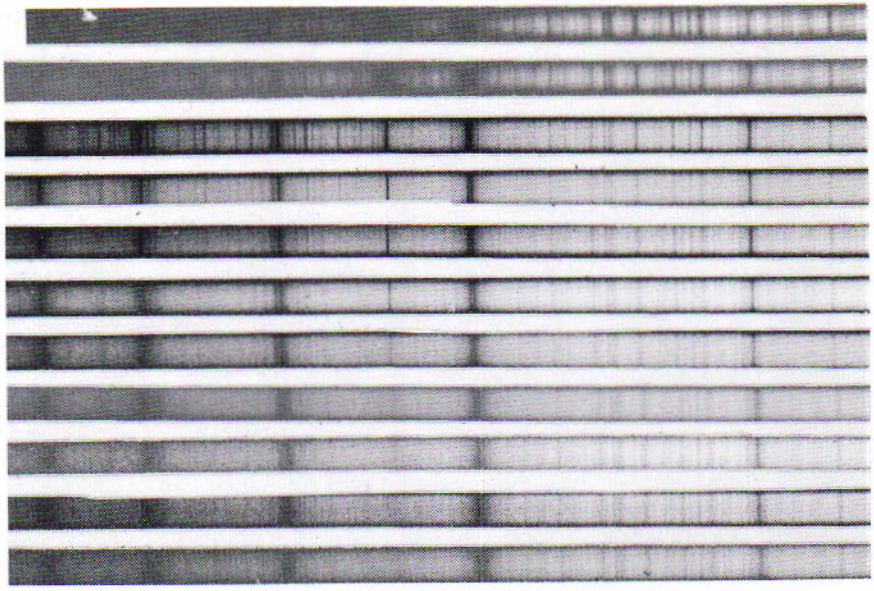
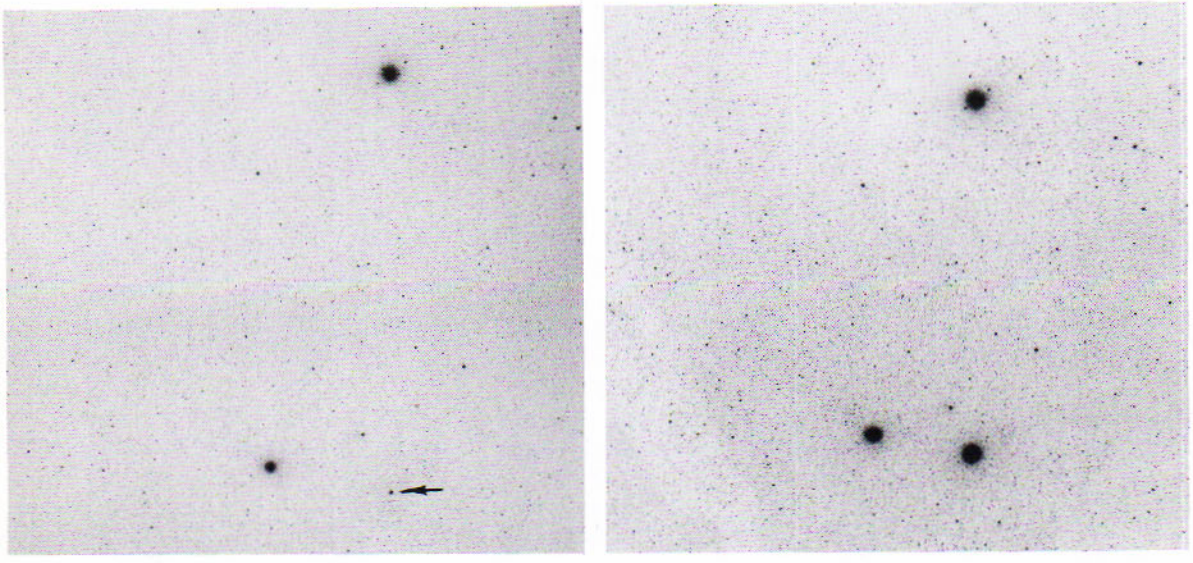


Figure 14.21A
Spectrum of Zeta Aurigae. The plates were taken between January 20 and February 4, 1948. The first two spectra show only the late type (cool) star. In the middle seven spectra the B star shines through the extended atmosphere of the cooler star producing additional sharp absorption lines. Note the enhancement of the K line of calcium in the middle of the series. The final two spectra show the composite spectrum of the two stars since the hot star is no longer shining through the atmosphere of the cooler star. *(Courtesy of A. Cowley, University of Michigan)*

Figure 14.21B
Region of Zeta and Epsilon Aurigae. The arrow points to Zeta in eclipse. In the view at the right the eclipse has ended. Epsilon is the bright star at the top. *(Photograph on right from National Geographic Society and Mt. Palomar Sky Survey)*



20 to 28, 1948. They begin shortly before the blue star has come out from total eclipse, and end when it has almost risen from behind the upper levels of the red star's atmosphere.

In the first photograph the spectrum of the large red star appears alone. In the third the blue star has completely emerged from behind its companion, adding its spectrum containing strong hydrogen lines. Here we also see dark lines abstracted from the light of the blue star by the red star's atmosphere, and in the later photographs we note how these lines fade as the blue star shines through successively higher levels of that atmosphere.

A detailed study of the spectra has shown that the extensive atmosphere of the red supergiant is arranged in strata. Atoms of neutral metals are prominent in the gases at the lowest levels. Atoms of ionized metals occupy the intermediate levels. Hydrogen and ionized calcium are abundant throughout and display their dark lines up to the highest levels, at distances above the star's surface equal to nearly half its radius.

Three other supergiant eclipsing stars are known that permit similar studies of the atmospheres before and after the eclipses. These are 31 Cygni, 32 Cygni, and VV Cephei.

The binary *Epsilon Aurigae* presents an interesting case where a periodic reduction in brightness is caused only by the passing of one star behind the cloudy atmosphere of the companion. This binary revolves in a period of 27.1 years. The whole eclipse lasts nearly two years, and the deepest phase, when the light is reduced 0.8 magnitude, has about half that duration. Struve in 1958 accounted tentatively for the eclipse as follows:

The binary consists of a type F supergiant and a relatively small companion, perhaps of type B, which is itself too faint to make an impression in the spectrum. The F star is not eclipsed directly by the companion, but by its enormous cloudy atmosphere. Gas clouds of this atmosphere are ionized by the radiations of both stars and are thereby made opaque enough to cause the observed dimming of the light of the system. The spectrum of the F star remains visible during the eclipse.

The latest eclipse of Epsilon Aurigae began in June, 1955, and ended in May, 1957. A doubling of many lines in the spectrum, observed at that eclipse by Otto Struve and Helen Pillans, is ascribed by them to clouds of gas around both stars, revolving generally in the direction of the orbital motion.

14.22 Gas Streams Around Close Binary Stars

Many close pairs of stars, where the separations are small compared with the diameters of the stars themselves, are involved in gas streams. The importance of such streams in the development of those binaries has been shown by F. B. Wood, by Z. Kopal, and by Struve and associates.

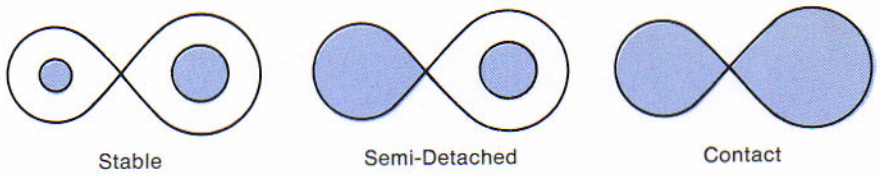


Figure 14.22
Types of close and contact binaries. The stars are shaded. Critical equipotential curves resemble the figure 8. *(Diagram by Z. Kopal, University of Manchester)*

If both stars are well within the critical equipotential surfaces between them (Fig. 14.22), the binary is *stable.* At the point where the surfaces join, the attractions toward the two stars are equal. When one star of a binary expands in its evolution from the main sequence until it fills its equipotential surface, the binary is *semi-detached.* Material expelled from this star, as by prominence action, may swirl around the binary and eventually fall into one star or the other. Kopal points out that the less massive component is the first to do so in every known case. Yet the current theory of stellar evolution would lead us to expect exactly the opposite. If both stars expand until they fill their equipotential surfaces, the result is a *contact binary* that is surrounded by gas from both. A numerous class of these is that of the W Ursae Majoris binaries, having periods of revolution around half a day.

14.23 Beta Lyrae Involved in Gas Streams

Beta Lyrae is a well-known example of an eclipsing binary surrounded by gas streams. It consists of a larger B8 star and a smaller F companion, which is in front at the primary eclipse and is enough fainter so that its spectrum is not observed. The two stars revolve in a period of $12^{d}\ 22^{h}\ 22^{m}$, which is increasing at the rate of about 10 seconds a year. Clouds of gas revolving around the binary add their dark and bright lines to the spectrum.

Spectra of Beta Lyrae taken by J. Sahade and O. Struve and discussed by them allow us to piece together a detailed picture of the system. In Fig. 14.23 we diagram the results of their studies. Around the outer edge of the figure is the phase or aspect of the system as we see it during any given period. The middle of primary eclipse occurs when the observer is in the direction of 0.000. The arrows indicate direction of motion and velocities are given in kilometers per second. The center of mass of the system is indicated by the X marked with a C.

Stationary interstellar lines appearing in the spectra allow very accurate relative measures to be made. The helium lines are split into many components and are the basis for the interpretation of a gas cloud surrounding the system as shown in the figure. In fact the gas cloud should really show discrete lanes of material spiralling outward and being lost from the system. It is this loss of mass that causes the lengthening of the period mentioned earlier.

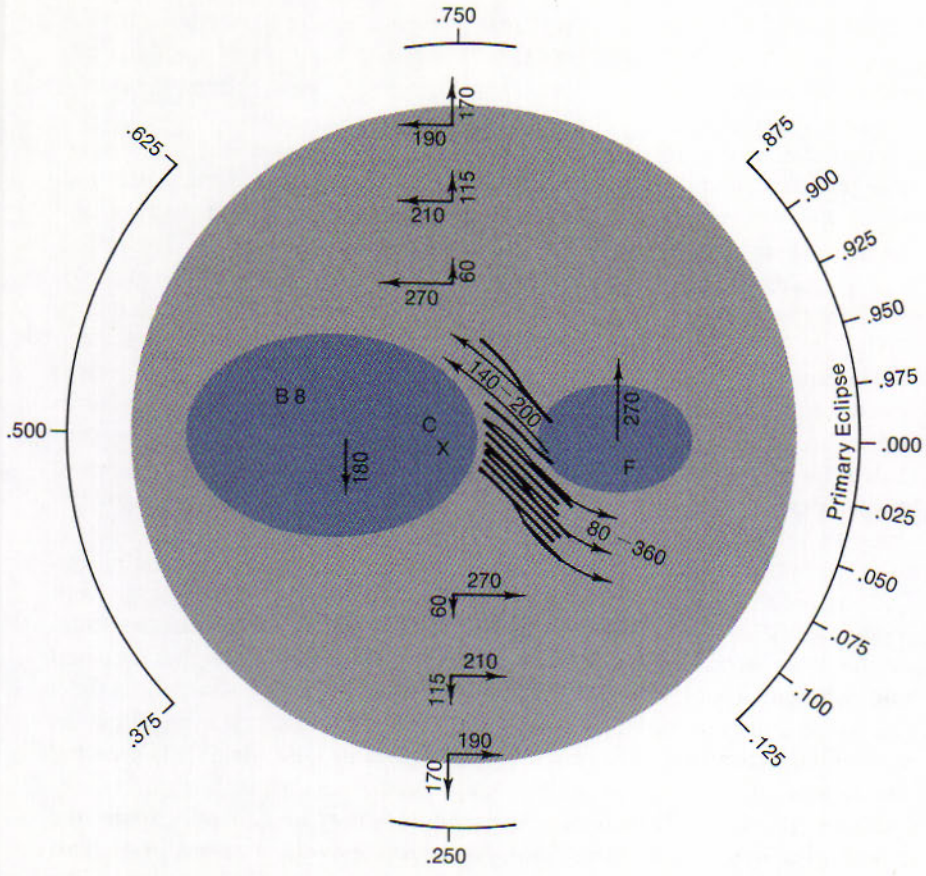


Figure 14.23
Gas streams around Beta Lyrae. The two stars are elongated by each other's attraction. Arrows for star and gas motions are marked in kilometers per second. *(Diagram by O. Struve)*

ROTATIONS OF THE STARS

Before concluding the account of eclipsing binaries we consider the evidence that these stars are rotating on their axes. This brings us to the rotations of single stars as well, as shown by the widening of their spectrum lines, and finally to the evidence of magnetic fields in certain stars.

14.24 Rotations of Eclipsing Stars

The spectroscopic method of studying the sun's rotation (10.2), by comparing the Doppler shifts of the lines at opposite edges of the disk, is not applicable ordinarily to stars that show no disks. The method can be applied, however, to some eclipsing binary stars. Preceding the middle of

the eclipse of a star by a much fainter companion, the light comes mainly from the part of the bright star that is rotating away from us. After the middle of eclipse the light comes mainly from the part of the star that is rotating toward us.

Thus during the initial phase of the eclipse the lines of the star's spectrum should be displaced farther toward the red, and during the final phase they should be displaced farther toward the violet end of the spectrum than can be ascribed to the revolution of the star. This effect was first detected by F. Schlesinger at Allegheny Observatory, in 1909, by the irregularity it caused in the velocity curve of the eclipsing binary Delta Librae. It was later observed in the curves of other eclipsing stars.

Beginning with these stars of known rotation and going on to spectroscopic binaries generally, assuming that each pair rotates and revolves in unison, it became possible to verify a Doppler effect that would be predicted for rotating stars. The spectrum lines are increasingly broadened as the speed of the rotation is greater. The way was now prepared for the study of the rotations of single stars.

14.25 Rotations of Single Stars

O. Struve and C. T. Elvey at Yerkes Observatory, about 1930, were pioneers in the study of the rotations of single stars from the contours of the lines in their spectra. The method was to compare the "washed out" lines of a star such as Altair (Fig. 14.26) with the sharp lines of a star of similar spectral type, such as Vega. The latter star is either in very slow rotation or, as seems more likely in this case, its axis is directed nearly toward us.

It was shown that such widening of the lines may be generally attributed to the rotations of the stars. Thus Altair rotates with a speed of at least 240 km/sec at its equator and in a period of 15 hours or less, depending on how nearly its axis is perpendicular to the line of sight.

14.26 Rotations of Main-Sequence and Giant Stars

If the rotation axes of stars have random directions, it is easy to calculate from the observed rotation speeds the average actual values for a particular class of stars. Single blue stars of the main sequence are likely to have high speeds of rotation. Some have equatorial speeds of 300 km/sec or more; such swiftly rotating stars may not be far from instability. The yellow and red stars of this sequence have more moderate speeds, except those that occur in close binary systems. Struve suggests that these yellow and red single stars may have imparted much of their original spins to the revolutions of their planets. If the sun were to absorb its planetary system, the sun would rotate 30 times as fast as its present equatorial rate of 2 km/sec.

The brighter star of a main-sequence visual binary is likely to rotate more rapidly than the fainter star, as Struve has observed from the

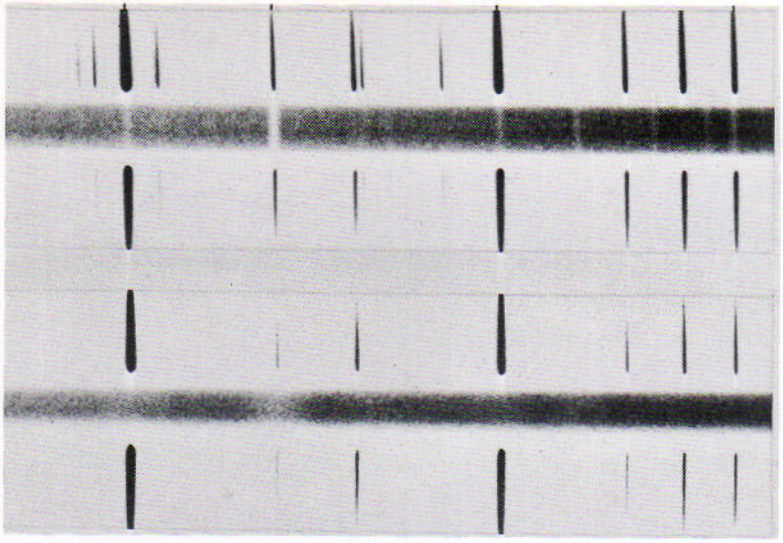


Figure 14.26
Negative prints of spectra of Vega and Altair. The broadened lines of Altair show the rapid rotation of this star. *(Yerkes Observatory photograph)*

widening of the spectrum lines. An extreme example is the binary ADS 8257. Although the two stars differ in brightness by a whole magnitude, they are both of type F. The lines in the spectrum of the brighter star (above) are made diffuse by a rotation as fast as 100 km/sec at the equator, whereas the lines of the fainter star (below) are practically unwidened, indicating a low speed of rotation.

Giant stars generally have slower rotations than corresponding main-sequence stars, as would be expected from the conservation of angular momentum (9.33) if they have expanded in their evolution from that sequence. Employing values of rotation speeds of giants determined by A. Slettebak, A. Sandage has calculated that when these giants were on the main sequence they were rotating in general as swiftly as do the stars that are now there.

14.27 Magnetic Fields of Stars

The Zeeman effect (10.17) in stellar spectra was discovered by H. W. Babcock at Mount Wilson Observatory in 1946 in the spectrum of the A-type star 78 Virginis. Employing a double polarizing analyzer, he observed a division of the lines corresponding to a polar field strength of 1500 gauss.

Babcock's catalog of 1958 lists 89 stars definitely showing magnetic fields and many others that probably show this effect. They are generally type A stars having sharp and unusually intense spectrum lines of metals such as chromium and strontium, and of rare earths such as europium. All the magnetic fields are variable in strength and some show reversals of polarity in cycles of a few days.

An example is α^2 Canum Ven. The polar magnetic field of this star varies in cycles of 5½ days between extremes of +5000 and −4000 gauss, which are comparable with the strongest fields observed in sunspots. At the first extreme the lines of metals in the spectrum have their greatest intensity, and at the second the lines of the rare earths are the most intense. An unusually strong magnetic field was reported in 1960 in the case of the 8.6-magnitude type A star HD 215441. The field is positive, varying irregularly from 12,000 to 34,400 gauss.

Extreme field strengths have been predicted for stars that evolve to small radii. The field strength is the number of flux lines passing through a square centimeter perpendicular to the flux lines. We can show for stars that the field strength should increase inversely proportional to the square of the radius. In the case of the sun, halving its radius would raise its average magnetic field to about 4 gauss. When the sun becomes a white dwarf its magnetic field will be on the order of 14,000 gauss. Recently a white dwarf having a field strength of 10^7 gauss has been observed by J. C. Kemp and his associates. From these arguments it is easy to see why field strengths of 10^{10} gauss and larger are quoted for pulsars (13.25) and are invoked to explain the beamed pulse phenomenon believed to be operating in those stars.

The proposal is made by Babcock that hydromagnetic fluctuations analogous to the 22-year solar magnetic cycle are occurring in the surface layers of these stars. Strong magnetic fields, as he explains, are the property of all rapidly rotating stars having convective outer zones; but they are observable only in the small proportion of such stars that have the rotation axes directed nearly toward the earth.

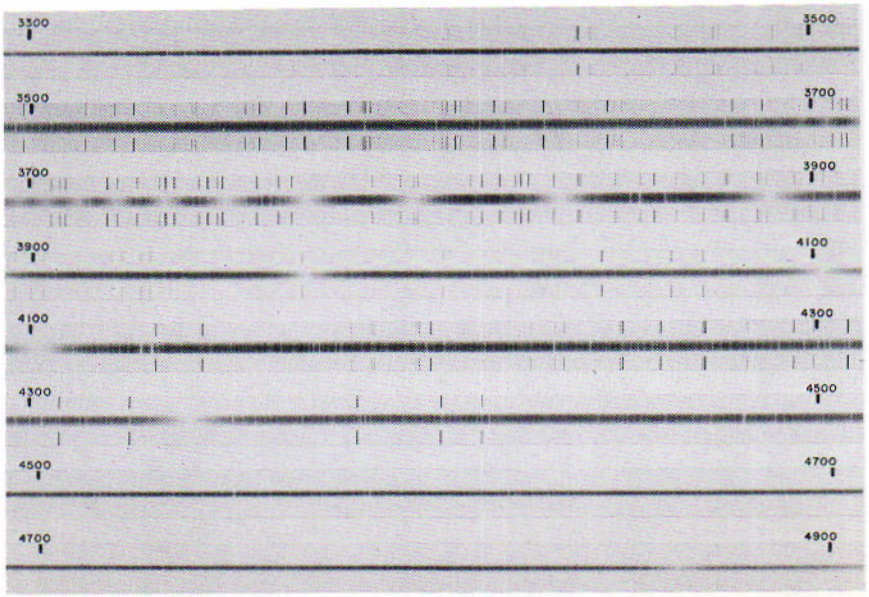


Figure 14.27
Spectra of α^2 Canum Venaticorum (AOp). Note the diffuse hydrogen lines. *(Photograph from the Hale Observatories)*

REVIEW QUESTIONS

1. The primary of a visual binary is also an eclipsing binary. Its absolute magnitude is 0.0. During total eclipse of the smaller star, its apparent magnitude is +5.0. The visual companion is 3″ distant. What do we expect the period of the visual system to be?
2. What is meant by the term position angle in visual binary observations?
3. The sum of the masses of 61 Cygni totals one solar mass. The period is 720 years with an average separation of 24″. What do you expect the parallax to be? (Check your answer in table 11.I.)
4. Why are double stars studied with care and devotion?
5. Enumerate the six features of binary systems.
6. What is the maximum effect of the earth's motion on the velocity curve? On the timing of a light curve?
7. What is the Rossiter effect?
8. What feature of the light variations of certain eclipsing binaries reveal elliptic orbits?
9. What is interesting and challenging about the system ε Aurigae? ζ Aurigae?
10. What observations suggest gas streams around β Lyrae?
11. Explain why the lines of a rapidly rotating star are sharp when the star is observed pole on.

REFERENCES

Binnendijk, Leendert, *Properties of Double Stars,* Philadelphia: University of Pennsylvania Press, 1960.

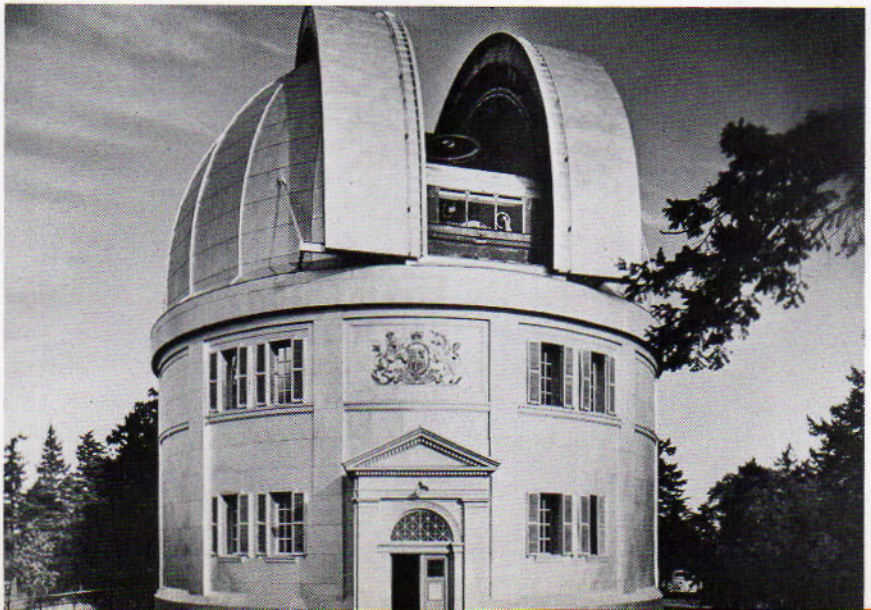
van de Kamp, Peter, *Basic Astronomy,* New York: The Macmillan Company, 1952.

Wood, F. B., ed., *Photoelectric Astronomy for Amateurs,* New York: The Macmillan Co.

FOR FURTHER STUDY

Kopal, Z., *Close Binary Systems,* John Wiley & Sons, New York, 1959.

Strand, K. A. ed., *Basic Astronomical Data,* Chicago: University of Chicago Press, 1963.



The Dominion Astrophysical Observatory's dome and building housing its 72-inch (182.8 mm) reflector at Victoria, B.C., Canada. *(Official photograph by the Dominion Astrophysical Observatory)*


15

STAR CLUSTERS

GALACTIC CLUSTERS — GLOBULAR CLUSTERS

Star clusters are physically related groups of stars having their members less widely separated than are the stars around them. The common motion of the members of a cluster through the star fields suggests their common origin by the condensing and fracturing of a large cosmic gas cloud. Because the stars of a particular cluster are practically at the same distance from us, they may be compared fairly one with another. Although their ages are about the same, the members have different masses, and the more massive ones have evolved the faster. Thus the cluster stars can inform us of the course of stellar evolution.

Star clusters are associated with other galaxies as well as with our own. They are of two types: galactic, or open clusters and globular clusters. An authoritative reference concerning them is Helen Sawyer Hogg's article in *Handbuch der Physik*, volume 53 (1959).

The double cluster in Perseus. *(Dyer Observatory photograph)*

GALACTIC CLUSTERS

Galactic clusters, *such as the double cluster in Perseus (Fig. 15.1A), are so named because those in our galaxy are near its principal plane. Thus they appear in or near the Milky Way, except some of the very nearest ones, notably the Coma Berenices cluster, which is not far from the pole of the Milky Way. Also known as* open clusters, *they are rather loosely assembled. Their separate stars are distinguished with the telescope, and in some cases the brighter ones are visible to the unaided eye. The nearest clusters, which are likely to have large proper motions, are sometimes called* moving clusters.

15.1 Cluster Catalogs and Designations

Star clusters, nebulae, and galaxies were formerly cataloged together. They are often designated by their numbers in one of those catalogs. The great cluster in Hercules, for example, is known as NGC 6205, or as M 13. The first designation is its running number in Dreyer's *New General Catalogue* (1887); this catalog and its extensions in 1894 and 1908, known as the *Index Catalogue* (IC), list over 13,000 objects. The second designation of the Hercules cluster is its number in the catalog of 103 bright objects that the comet hunter C. Messier prepared for the *Connaissance des Temps* of 1784. A useful list of these objects, with their positions in the sky, is given in *Sky and Telescope* for March, 1954.

Figure 15.1
Pair of clusters in Perseus: h and χ persei. *(Dyer Observatory photograph)*

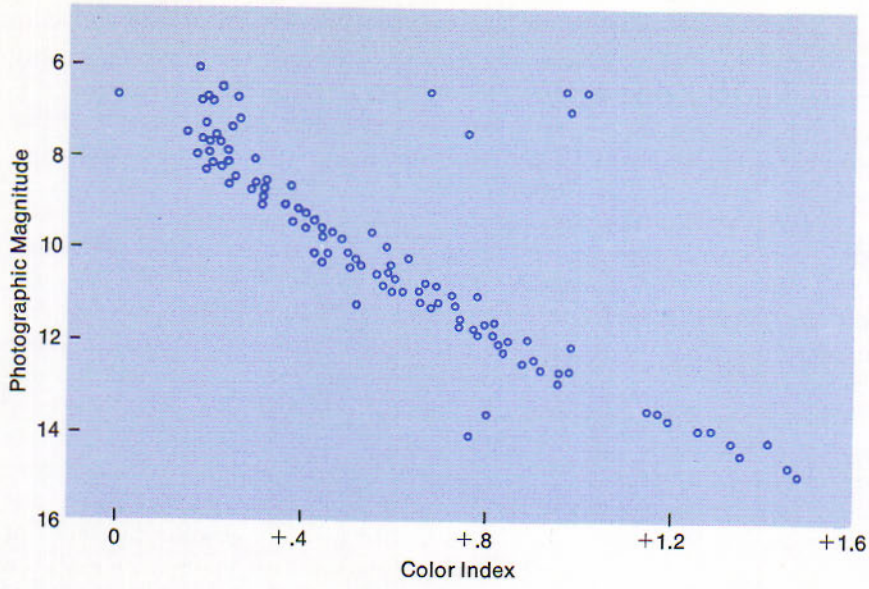


Figure 15.2A
Color-magnitude diagram of the Praesepe Cluster. *(Adapted from a diagram by H. L. Johnson)*

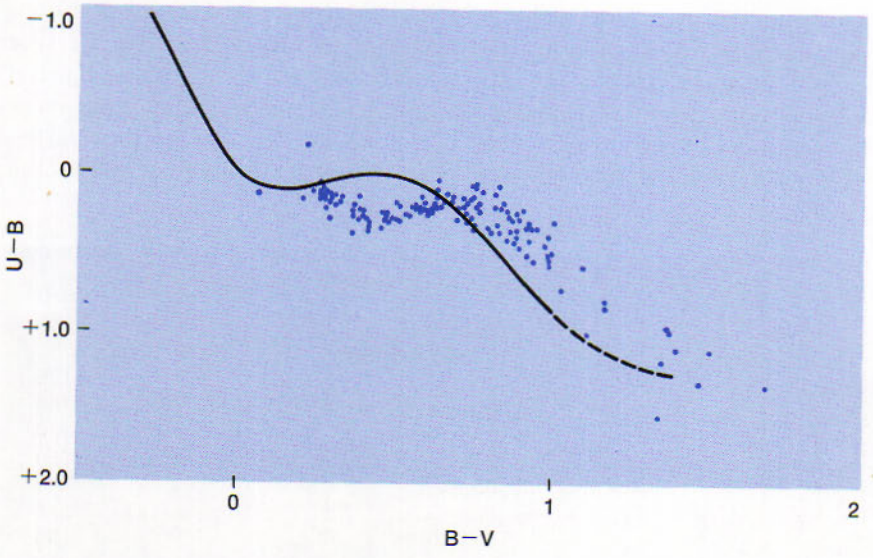


Figure 15.2B
Observed color-color plot of a galactic cluster. Unreddened main sequence is indicated by curve. *(Adapted from U. S. Naval Observatory Publications).*

15.2 Color-Magnitude Diagram

Except for the very nearest clusters, the dimensions of star clusters are small enough to be neglected for the present purpose compared with their distances; thus, it is assumed that the stars of a cluster are at the same distance from us. Thus the apparent magnitudes, m, of these stars differ from the absolute magnitudes, M, by the same amount: $m - M = 5 \log r - 5$, where r is the distance of the cluster in parsecs. The quantity $m - M$ is referred to as the *distance modulus.*

When the apparent magnitudes of the cluster stars are plotted with respect to their spectral types, the array of points has the same significance as the spectrum-absolute magnitude diagram except for the constant difference $m - M$. When the color indexes are employed instead of the spectral types, we have the equivalent of the *color-apparent magnitude diagram* (Fig. 15.2B).

Comparing this diagram for the Praesepe cluster with the standard spectrum-absolute magnitude diagram of Fig. 11.27 and remembering that color index has a known relation to spectral type, we see how the distance of the cluster may be determined. H-R diagrams may be used as well, but the color index of a star is much easier to obtain, especially in the case of faint ones, as we have seen. The method is to match comparable parts of the two arrays, to note the difference between the apparent and absolute magnitudes, and to calculate the distance by the formula given above. Allowance must be made for any dimming of the cluster stars by intervening cosmic dust, which would make the calculated distance greater than the actual distance, and also for any displacement of the cluster stars in their evolution from the standard main sequence. The standard procedure is to make a color-color plot as in Figure 15.2B and determine the reddening by the amount of shifting required to get it onto the color-color plot of the unreddened main sequence (Fig. 11.24B). Our measured colors and magnitudes are then corrected for reddening and then the differences in magnitude, corrected for neutral absorption mentioned above, is read off giving the distance modulus of the cluster. The distance of the Praesepe cluster so determined and corrected is 158 parsecs, or 515 light years.

TABLE 15.I
Distances of Galactic Clusters*

CLUSTER	CONSTELLATION	PARSECS
Hyades	Taurus	40
Coma	Coma Berenices	80
Pleiades	Taurus	126
Praesepe	Cancer	158
M 39	Cygnus	250
IC 4665	Ophiuchus	330
M 34	Perseus	440
M 25	Sagittarius	550
M 67	Cancer	830
NGC 2264	Monoceros	870
M 36	Auriga	1260
NGC 2362	Canis Major	1450
NGC 6530	Sagittarius	1580
NGC 2244	Monoceros	1660
M 11	Scutum	1740
Double cluster (h and χ)	Perseus	2250

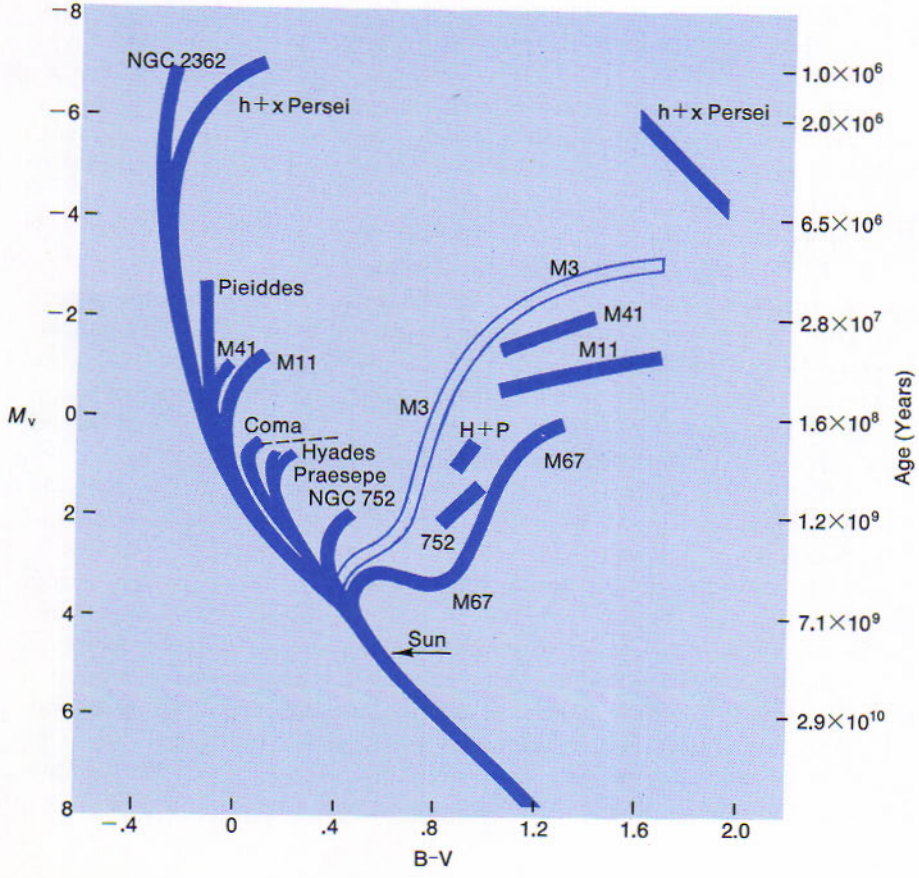
* Determined photoelectrically by H. L. Johnson

15.3 Star Clusters of Different Ages

In the color-magnitude diagram of the Praesepe cluster, the top of the main sequence bends to the right, and a few of the brightest stars are still farther to the right. Turning now to Fig. 15.3, we see that other clusters break off the sequence at other places. Main-sequence stars are absent in galactic clusters above the breaks. The opinion is that the differences between such diagrams are mainly effects of advancing age of the clusters and also of their chemical composition.

As the clusters grow older, according to a theory of stellar evolution that is examined further in the following chapter, its stars leave the main sequence and move into the giant sequences. The more luminous stars shift more rapidly, because they are converting their hydrogen to heavier elements at a faster rate. The less luminous stars have not yet had time to show conspicuous shifts.

Figure 15.3
Color-magnitude diagrams for 10 galactic clusters and the globular cluster M 3 (Fig. 15.15). Color indexes are plotted with respect to absolute visual magnitudes. *(Diagram by A. R. Sandage)*



NGC 2362 is the youngest cluster represented in the figure; its age is given as only 10^6 years. The double cluster in Perseus is also in its youth, according to this view, which seems to find support in the unusually large memberships of these clusters and the extent of the main sequence still intact. The Pleiades cluster is middle-aged, and the Hyades and Praesepe clusters are approaching old age. The cluster M 67 is the oldest galactic cluster represented in Sandage's diagram of 1957; its age is 5 billion years by the dating process then employed. Another cluster, NGC 188, was represented in a later diagram as about twice as old; and all the ages were increased by a revised dating process. We have not altered the original diagram because further revisions are expected.

15.4 Lifetimes of Clusters

The common motion of the stars of a cluster suggests their common origin, probably by the condensation of a very large cloud of interstellar gas and dust. The maintenance of the common motion shows that the cluster stars are not quickly affected by the field stars through which they pass. Yet at each encounter with a field star the cluster stars are attracted in slightly different amounts depending on how near the intruder they pass. Thus very gradually stripped of members by the field stars and even more effectively by collisions with interstellar clouds, as Lyman Spitzer has shown, the cluster should ultimately be dispersed, generally in a shorter time than the lifetime of the galaxy itself. The more compact clusters and especially those farther removed from the central plane of the star clouds should have the longer lives.

The galactic cluster M 67, where the stars break from the main sequence (Fig. 15.3) along with those of the durable globular cluster M 3, has excellent reason for its long life. This cluster near Alpha Cancri is about 10° from the Praesepe cluster. Both clusters are more than 30° from the central line of the Milky Way. M 67, however, is 5 times as remote as Praesepe and is at an unusually great distance above the principal plane of the Galaxy. There it is relatively immune to disturbances which hasten the disruption of clusters; it has retained at least 500 members.

A few galactic clusters are sufficiently close to the Sun for motions within the cluster to be discernible after a long period of observation. Studies of these motions yield results in agreement with the photometric ages. Such studies determine the total mass of the cluster which, along with an assumed luminosity function and the known bright members, allows us to extrapolate to the number of fainter and unseen members. Results of studies of Praesepe show that we see only about ⅓ of the members and the mean internal motion is such that the cluster is bound together rather strongly by the mutual gravitation of its components. The dissolution of this particular cluster must proceed very slowly and is due to field star and gas cloud encounters mentioned above.

Figure 15.4
Praesepe cluster in Cancer. A fine example of a galactic cluster. *(Harvard College Observatory Photograph)*

The dissolution of a cluster by loss of the more distant members should proceed more rapidly with time. If we assume that the energy of the cluster remains constant, then each time an outer member is lost the inner members become more tightly bound, until ultimately there is left a small, dense, almost spherical cluster as a remnant. This remnant cluster contains only a few members and should not be confused with the globular clusters soon to be discussed.

15.5 Examples of Galactic Clusters

The cluster of the Pleiades, or "Seven Sisters," in Taurus is known to many people. The V-shaped Hyades cluster in the same constellation is also conspicuous in our skies. The brighter stars of both clusters are plainly visible to the naked eye, and those of the Coma Berenices cluster are faintly visible. The Praesepe cluster in Cancer (Fig. 15.4), also known as the "Beehive," the double cluster in Perseus, and a few others appear as hazy spots to the unaided eye and are resolved into stars with slight optical aid. These nearer clusters are well observed with binoculars. Many others can be viewed with small telescopes. Galactic clusters are prominent in photographs of the Milky Way.

About 500 galactic clusters are cataloged in the galactic system. Their memberships range from around 20 stars to a few hundred, and to more than a thousand in the rich Perseus clusters. Their visibility is limited because of their small memberships and the scarcity of high-luminosity stars in them. Many clusters must be unnoticed against the bright background of the Milky Way or concealed by the heavy dust in these directions. The galactic clusters we observe around us are all within 20,000 light years from the sun, but they are doubtless as abundant in other parts of the spiral arms of our galaxy as they are in our own neighborhood.

15.6 The Ursa Major Cluster

This cluster contains the bright stars of the Great Dipper, except its end stars Alpha and Eta, and several fainter stars of the constellation. Although its members are considerably separated in the sky because they are near us, they form a compact group having about the size of the smaller galactic clusters. Surrounding the cluster is a large and sparsely populated stream having so nearly the same motion as the cluster itself that all these stars were formerly believed to comprise together a cluster of extraordinary size.

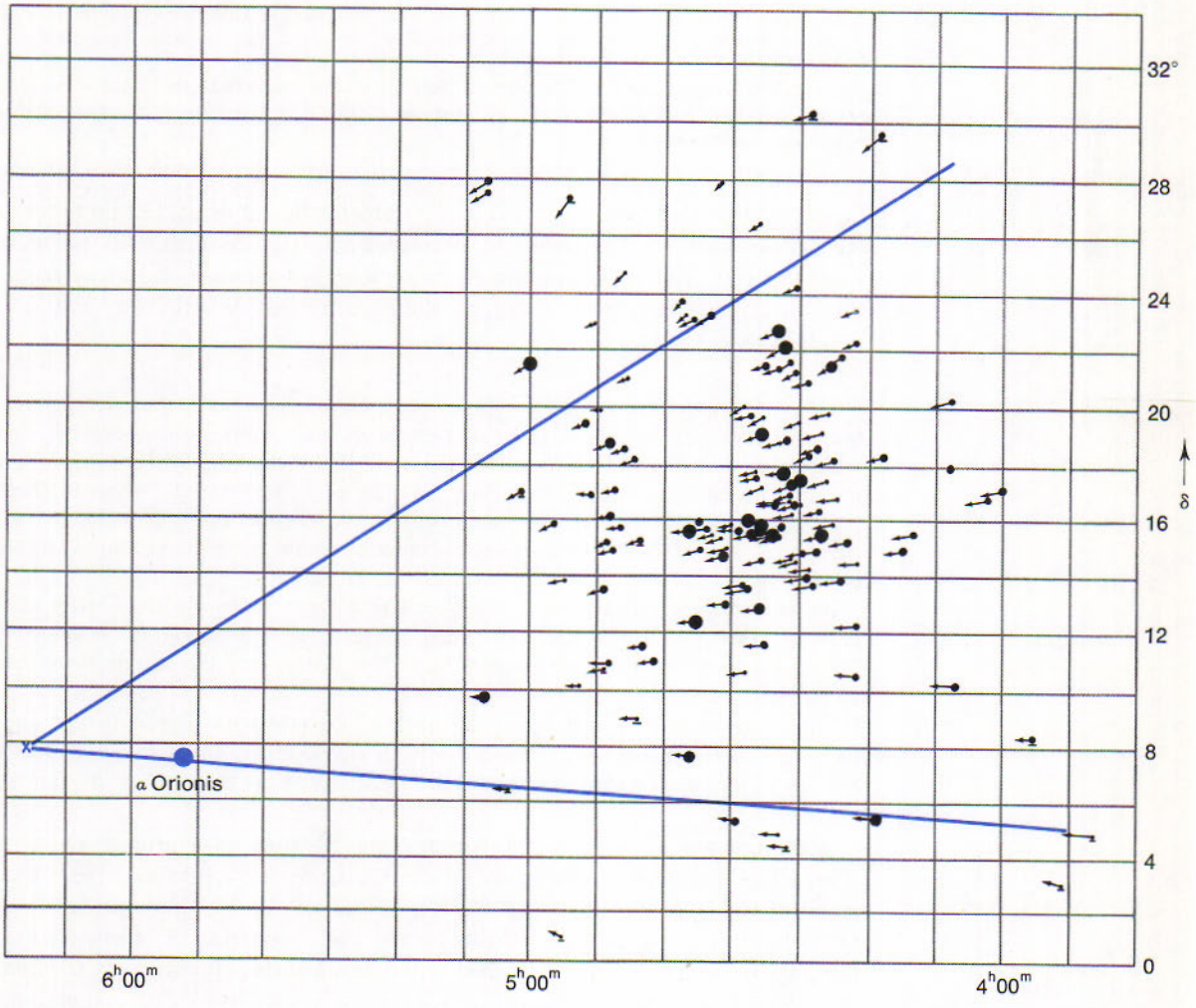
The stream around the Ursa Major cluster includes such widely scattered stars as Sirius, Alpha Coronae, and Beta Aurigae. Nancy G. Roman's spectroscopic parallaxes at Yerkes Observatory show that this stream envelops not only the sun, which is itself not a member, but two other galactic clusters as well.

15.7 The Hyades Cluster

Just as the rails of a track seem to converge in the distance, so the parallel paths of stars in a cluster are directed toward a convergent point if the cluster is receding from us. This effect of perspective is especially noticeable in the proper motions of the Hyades cluster (Fig. 15.7A), which is so near us that it covers an area of the sky 20° in diameter.

The Hyades cluster comprises the stars of the familiar group itself, except Aldebaran, which has an independent motion not shown in the

Figure 15.7A
Convergence of the Hyades. Alpha Orionis is not a member of this famous cluster. *(Diagram adapted from H. G. van Bueren by permission of BAN)*



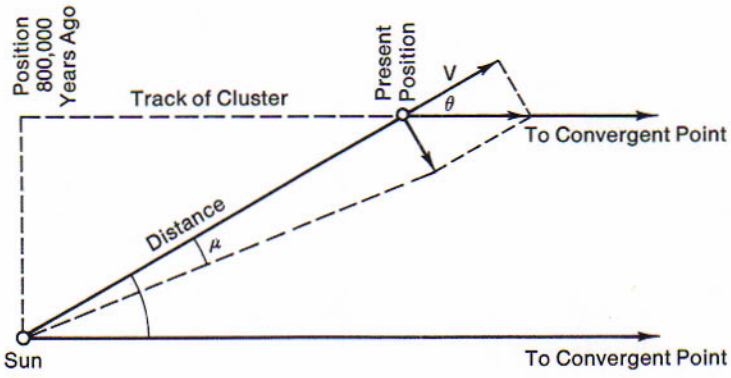


Figure 15.7B
Track of the Hyades Cluster. The present distance of the cluster can be found when the proper motion, μ, radial velocity, V, and angular distance, θ, from the convergent point have been determined.

figure, and of the region around it. The cluster has at least 150 members. Its denser part is 10 parsecs in diameter, and its center is 40 parsecs from the sun. The space motion of these stars is eastward and away from the sun. The convergent point of their paths lies a little way east of Betelgeuse in Orion.

When the convergent point of a moving cluster is known and the proper motion, μ, and radial velocity, V, of one of its stars have been observed, the distance of that star and of any other member of the cluster with known proper motion can be calculated. The space velocity is $v = V/\cos\theta$, where θ (Fig. 15.7B) is the star's angular distance from the convergent point. The tangential velocity is $T = v \sin\theta$. The parallax can now be found from the relation (11.11): $p = 4.74\mu/T$.

As an example, the observed values for Delta Tauri in the Hyades cluster are: $\mu = 0''.115$, $V = +38.6$ km/sec, and $\theta = 29°.1$. The resulting space velocity is 44.0 km/sec, and the parallax is $0''.025$. The distance of the star is therefore 40 parsecs.

The space velocity and θ define the track of the star with respect to the sun. By means of relations easily worked out from Fig. 15.7B it can be shown that the Hyades cluster was nearest the sun 800,000 years ago at the distance of 20 parsecs.

Although we have so far considered the Hyades as a unique cluster, it may well be what is left of an enormous association resulting from the star formation process. Spread out far ahead of the Hyades and trailing far behind it along an arm of the galaxy are thousands of stars sharing the galactic motion of the Hyades. If we assume the Hyades was formed as a very large, dense association with only a small velocity dispersion, then differential galactic rotation is sufficient to explain the elongated distribution we now observe.

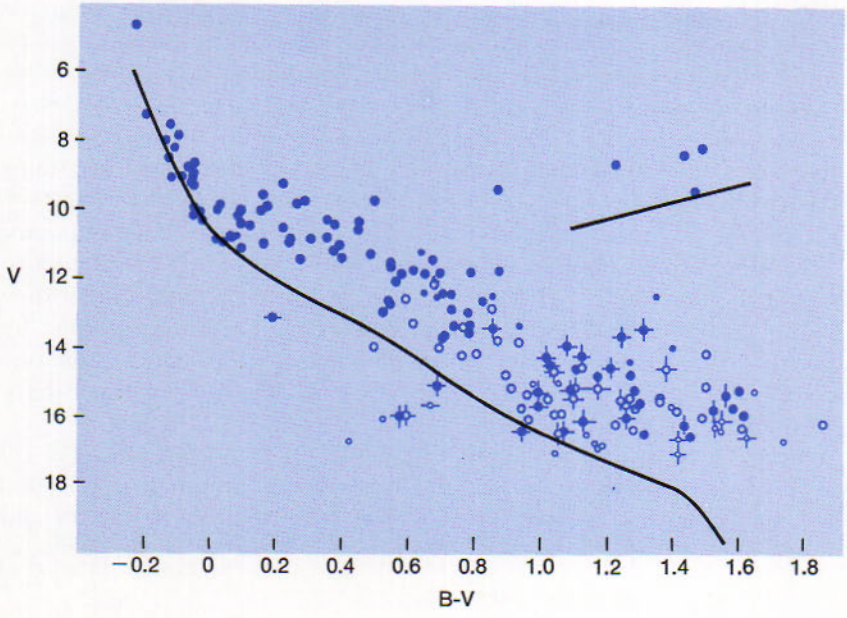
15.8 Very Young Clusters

Some galactic clusters, such as NGC 2362, are even more youthful than the double cluster in Perseus. The upper parts of their main sequences in the color-magnitude diagrams are intact, containing the highly luminous O and B stars, which have arrived there a million years or so after their births. The less luminous stars from the blue down to the red ones lie above the main sequence, presumably because they have not yet had time to arrive. G. H. Herbig has shown this to be the case for the redder stars in a heavily clouded region of Taurus, and P. P. Parenago at Moscow found for a group associated with the Orion nebula that stars redder than A5 had not yet reached the vicinity of this sequence.

An example of an extremely young galactic cluster is NGC 2264, about 15° east of Betelgeuse and 870 parsecs from us. M. F. Walker's color-magnitude diagram of this cluster shows the O and B stars on the main sequence. The redder stars depart abruptly from the sequence at A0 and tend to lie about 2 magnitudes above it. Many of these are variable in brightness and have been regarded as T Tauri stars (16.15), which are brightened irregularly by causes associated with their extreme youth.

While it is implied in this explanation that the stars lying above the main sequence have not yet reached the main sequence, it is possible

Figure 15.8A
Color-apparent magnitude diagram of cluster NGC 2264. The dots represent photoelectric and the circles photographic measures. Vertical lines denote known variable stars, and horizontal lines stars having bright Hα lines in their spectra. The curve is the standard main sequence. *(Diagram by Merle F. Walker)*



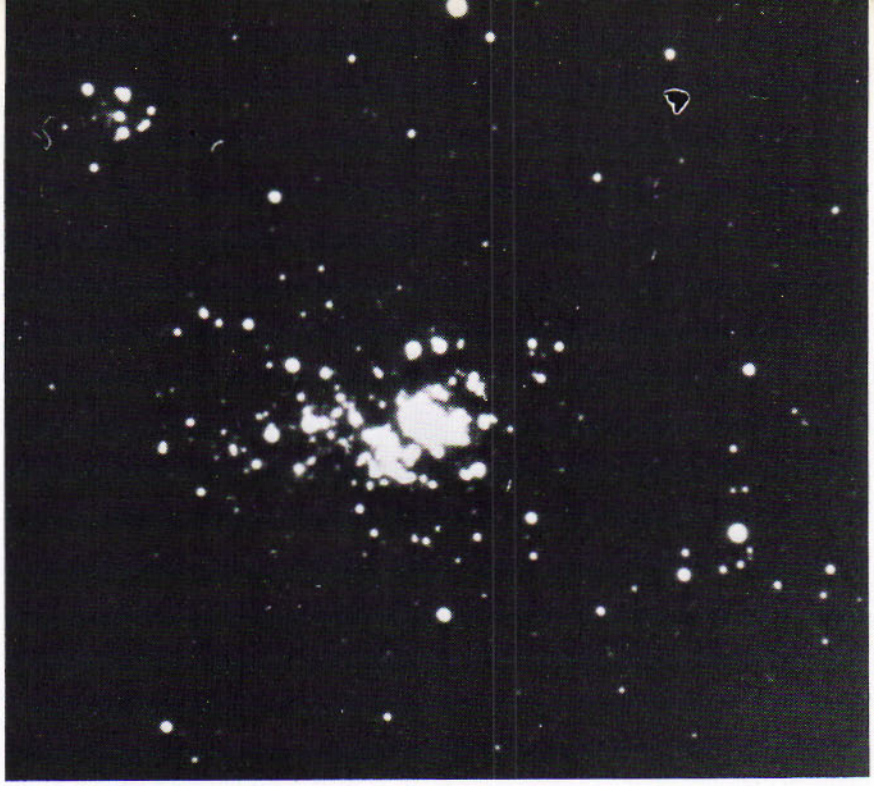


Figure 15.8B
30 Doradus cluster. An early open cluster in the Large Magellanic Cloud photographed with the 74-inch (197.7 cm) reflector at Mt. Stromlo Observatory. *(Photograph by B. Westerlund, courtesy of B. Bok)*

that we are seeing a "main sequence" of stars formed from an "enriched" interstellar medium (16.22). Such an explanation would allow for the giant branch which appears in the diagram, the stars of which are also about two magnitudes above the "normal" position. Even if the latter explanation is true we must still imply a very young age for the cluster.

15.9 Associations of Stars

It has long been known that O and early B stars tend to occur in groups in the Milky Way, which are less compact than the galactic clusters. In 1949 V. A. Ambartsumyan first called attention to the temporary existence of such *associations;* they are too feebly bound by gravitation to hold together very long, but their stars are evidently so youthful that they have not had time to disperse. Some associations have been studied by A. Blaauw, W. W. Morgan, and others, who find that the stars of each one are spreading rapidly from a common center where they must have originated.

Actually Ambartsumyan's work began with the T Tauri stars (16.16) and was later generalized to the O, B and A star groupings. He noted that most T Tauri stars were found in two small regions of the sky and that one of these regions was only 100 parsecs distant and had a diameter

of 25 parsecs. This led him to the belief that the grouping was not accidental and the stars must be associated in one way or another. The main conclusion to be drawn from these associations (T Tauri star, O star, B star and A star groupings being referred to as T, O, B and A Associations respectively) is that the stars must be young, otherwise the associations would have long since dissolved. This reinforces our belief that stars are still in the process of formation. A common feature for all associations is that they contain emission-line stars.

A group of 17 hot stars in Perseus is an excellent example of an O-B association. It is known as the Zeta Persei association from the name of its brightest member. The stars in its outskirts are moving at the rate of 14 km/sec from a center where they were born only 1.3 million years ago.

Figure 15.9
The Lagoon nebula, (M8, NGC 6523) in Sagittarius. Note the irregular black globules. *(Lick Observatory Photograph)*

Another association is centered near the Great Nebula in Orion and at about the same distance from us. Blue stars are withdrawing from the center mainly at the rate of 8 km/sec. But three of the stars, AE Aurigae, Mu Columbae, and 53 Arietis, have velocities of 106, 123, and 59 km/sec, respectively, away from this center. With such high speeds they seem to have escaped into other constellations only a few million years ago. Blaauw calls them "runaway stars." He has experimented with the hypothesis that these three single stars were originally the less massive components of binary systems. The primary stars may have exploded as supernovae, leaving their companions to continue on away from the center of the association with something like their former speeds of revolution in the binary systems. High-velocity stars are withdrawing from other O-B associations as well.

GLOBULAR CLUSTERS

Globular clusters *are more compact and are spheroidal in form. The cluster M 13 in Hercules (Fig. 15.11) is an example. They are larger, more populous, and more luminous than the galactic clusters, and those in our galaxy are less confined to the vicinity of the plane of the Milky Way.*

15.10 The Brightest Globular Clusters

About 120 globular clusters are recognized in the vicinity of our galaxy. These include a dozen faint ones reported in 1955 by A. G. Wilson and G. O. Abell from the Palomar Sky Survey and remote enough to be regarded as intergalactic objects. Only a few of the apparently brightest globular clusters are visible without the telescope.

The southern clusters Omega Centauri (Fig. 15.10) and 47 Tucanae are the brightest and also the nearest, at the distance of 6,700 parsecs. The former appears to the naked eye as a hazy star of the 4th magnitude; it was recorded as a star by Ptolemy and was later given a letter in the Bayer system. Because of their low declinations, these two clusters are not favorably placed for observers anywhere in the United States.

The globular cluster M 13 in Hercules is the brightest and best known to observers in middle northern latitudes, where it passes nearly overhead in the early evenings of summer. This cluster and M 22 in Sagit-

Figure 15.10
Globular cluster Omega Centauri. *(Yerkes Observatory Photograph)*

tarius are faintly visible to the naked eye. M 5 in Serpens, M 55 in Sagittarius, and M 3 in Canes Venatici can also be glimpsed in favorable conditions.

15.11 The Hercules Cluster

This cluster is spectacular as viewed with large telescopes, and especially so in long-exposure photographs with such telescopes (Fig. 15.11), where it appears not unlike a great celestial chrysanthemum. Its diameter in these photographs is about 18′, which at its distance of 9,200 parsecs corresponds to a linear diameter of 49 parsecs, or the distance of Spica from the sun. The cluster is estimated to contain 500,000 stars of average mass equal to that of the sun. The red stars of its main sequence have not

yet been observed, and the stars in its central region are too congested to be counted separately. Around the center the density of stars may be 100 times as great as the average for the cluster and 50,000 times that of the stars in the sun's neighborhood.

Like other globular clusters, M 13 has an outline that is not quite circular. It is an oblate spheroid presumably flattened at the poles by its slow rotation, although no other evidence of rotation has been detected either directly or with the spectroscope. The apparent oblateness of the cluster is 0.05, or less than that of the planet Jupiter.

Figure 15.11
Globular Cluster M 13 in Hercules. *(Official U.S. Navy photograph)*

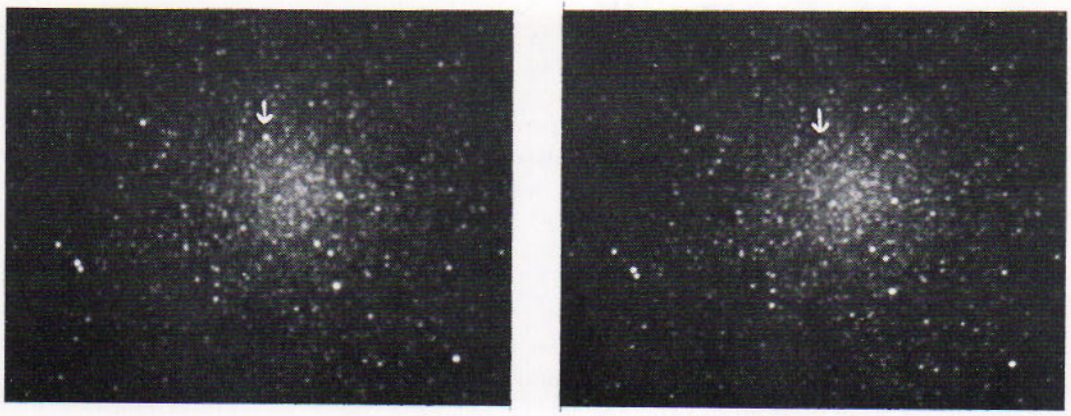


Figure 15.12
A Type II Cepheid in the Globular Cluster M14 (= NGC6402). The arrows point to the variable near maximum (left) and minimum (right) light. The apparent magnitude at maximum is 14.65 and at minimum 16.10. This Cepheid is unusual in that its period is shortening at a rate of 0.1 seconds per day. *(Photograph courtesy of Helen Hogg, David Dunlap Observatory)*

15.12 Variable Stars in Globular Clusters

A total of 1612 variable stars were reported, in 1961, by Helen Sawyer Hogg in more than 80 globular clusters that have been searched for such objects. The clusters M 3 and Omega Centauri are the richest in known variable stars, having 189 and 165 examples respectively. No variables at all have been found in five clusters searched. Ninety per cent of the variables in globular clusters are RR Lyrae stars. The remainder are of various kinds, most frequently cepheids of type II, Mira-type variables, and irregular variables.

Because the RR Lyrae stars were formerly believed to have the same median absolute magnitudes, the distances of globular clusters could be evaluated from the observed median apparent magnitudes of these variables observed in the clusters. By this means, H. Shapley, in 1917, determined the distances of the clusters. He then constructed a model of the cluster system with the idea that the array should be similar in extent and center to the system of the Milky Way.

Shapley's results brought out clearly for the first time the separate status of the Galaxy, and they prepared the way for the recognition of exterior galaxies. His original distances of the clusters and the dimensions of the Galaxy based on them were later reduced, when the importance of correcting for the effect of intervening cosmic dust came to be understood. A further revision of the cluster distances will be required when a new system of median absolute magnitudes of RR Lyrae variable stars (13.8) is firmly established.

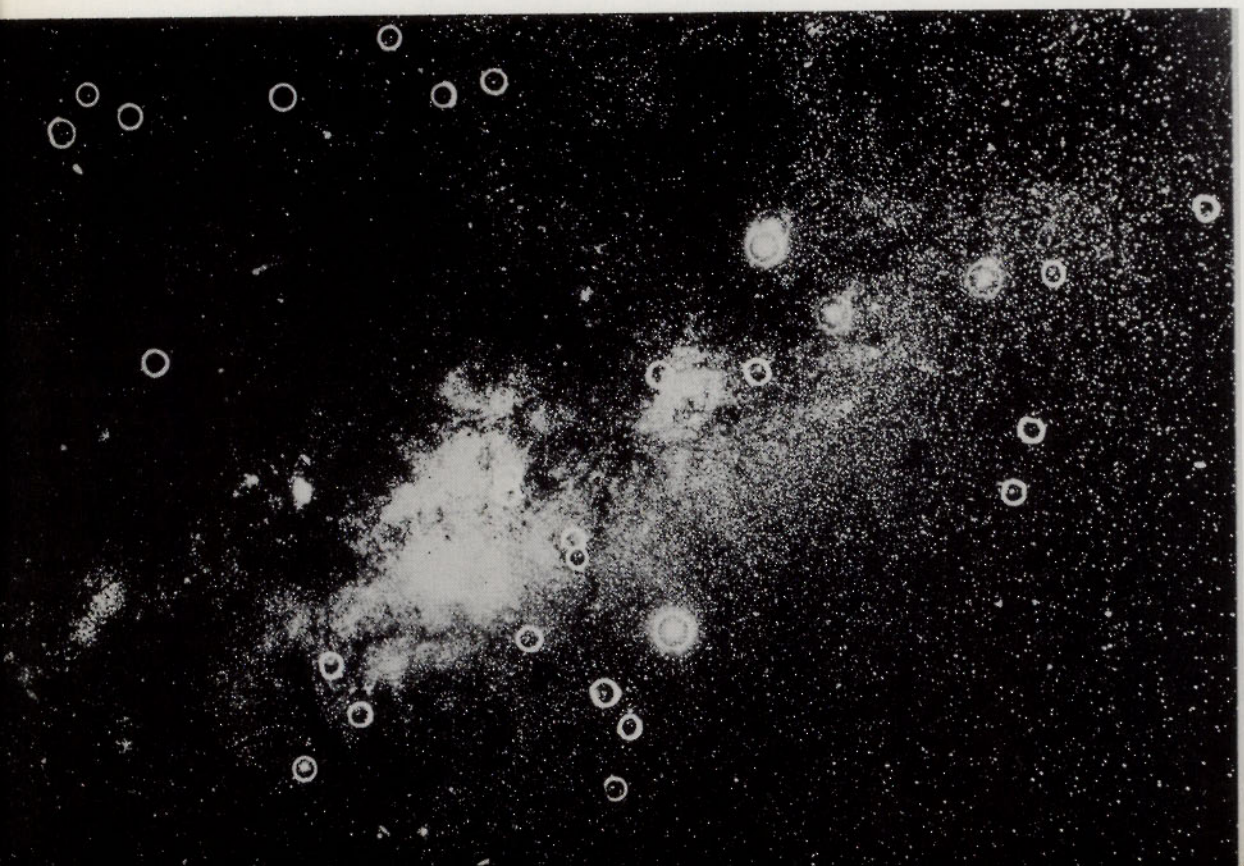
However, the assumption that the RR Lyrae stars have statistically the same magnitude, is a critical question that leads to a consistent picture

of the galaxy when certain corrections are made, and changing the median magnitude for the RR Lyrae stars only expands or contracts the dimensions of the Galaxy accordingly. A disconcerting factor appears when globular clusters are compared. If the RR Lyrae stars are matched, the various main sequences fall at different levels, but if the main sequences are matched instead, then the RR Lyrae regions do not match.

15.13 The Array of Globular Clusters

Starlight is dimmed by dust in interstellar space, particularly in the directions of the Milky Way, so that the photometrically determined distances of the stars are greater than the actual ones unless appropriate corrections are made. That the globular clusters of our galaxy are also dimmed and reddened by the dust is shown by the photoelectric measures of Stebbins and Whitford. The clusters far from the Milky Way are of uniform color corresponding to spectral type F6. They become redder toward the Milky Way, until they correspond in the extreme to the color of a type M star. In the direction of galactic center some of the clusters are totally obscured (see Fig. 15.13A).

Figure 15.13A
Distribution of Globular Clusters in the direction of Sagittarius. The clusters are encircled for easier recognition. Note their absence around the dark clouds due to obscuration. *(Yerkes Observatory Photograph)*



Statistical corrections for the effects of the dust have been obtained by supposing that the obscuring material is spread uniformly in a layer 1000 parsecs in thickness around the principal plane of the Milky Way. The measured distance of each cluster is reduced by an amount that depends on the galactic latitude (17.4) of the cluster and the assumed *optical thickness* of the dust layer, that is, the amount in magnitudes that the light would be dimmed if it should pass vertically through the layer.

Fig. 13.15B illustrates the results where the optical thickness is taken to be 0.46 magnitude. The places of some of the clusters are here projected on the plane through the sun and the center of the cluster system and perpendicular to the plane of the Milky Way. The corrections place the clusters mainly within a circle having a radius of 20,000 parsecs and its center about 9000 parsecs, from the sun. A few faint clusters are far outside these limits.

The distribution of globular clusters over the face of the sky shows that the center of the cluster system is in the direction of Sagittarius. Almost all recognized clusters are found in the hemisphere of the sky having its center in that constellation, and fully a third are in the vicinity of the great star cloud of Sagittarius.

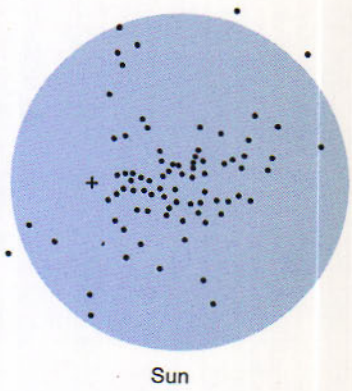


Figure 15.13B
Positions of Some Globular Clusters in the Galaxy. These are shown projected on a plane perpendicular to the principal plane of the Galaxy.

15.14 Two Stellar Populations

The spectrum-luminosity diagram for the stars in the sun's neighborhood differs from the diagram for the stars in globular clusters. Interest in the matter was not fully awakened until 1944, when W. Baade explained that populations of stars represented by the two diagrams are found in different parts of our galaxy and of other galaxies as well. Baade designated the two populations as types I and II respectively.

The *type I population* (Fig. 15.14) is represented by the sun's region of our galaxy, and was accordingly the first to be recognized. It frequents regions where gas and dust are abundant. Its brightest stars are blue stars of the main sequence, which is intact, and its red giants are around absolute visual magnitude zero. This is a young population, comprising stars of relatively high metal content (16.22).

The *type II population* (Fig. 15.14), represented by the stars of normal globular clusters, occurs in regions that are nearly free from interstellar gas and dust. Its giant sequence begins at the redder end with K stars of absolute visual magnitude −2.4. On the downward slope in the diagram this sequence divides into two branches, one of which runs horizontally to the left at about magnitude zero. The second branch continues on in nearly the same direction until it joins the type I main sequence. To the left of this junction there are no original main sequence stars of the cluster. This is an old population, comprising stars of low metal content.

The prototypes of Baade's two stellar populations are extreme cases. Intermediate types have been suggested. A possible choice of five different populations in our galaxy is noted later (17.7).

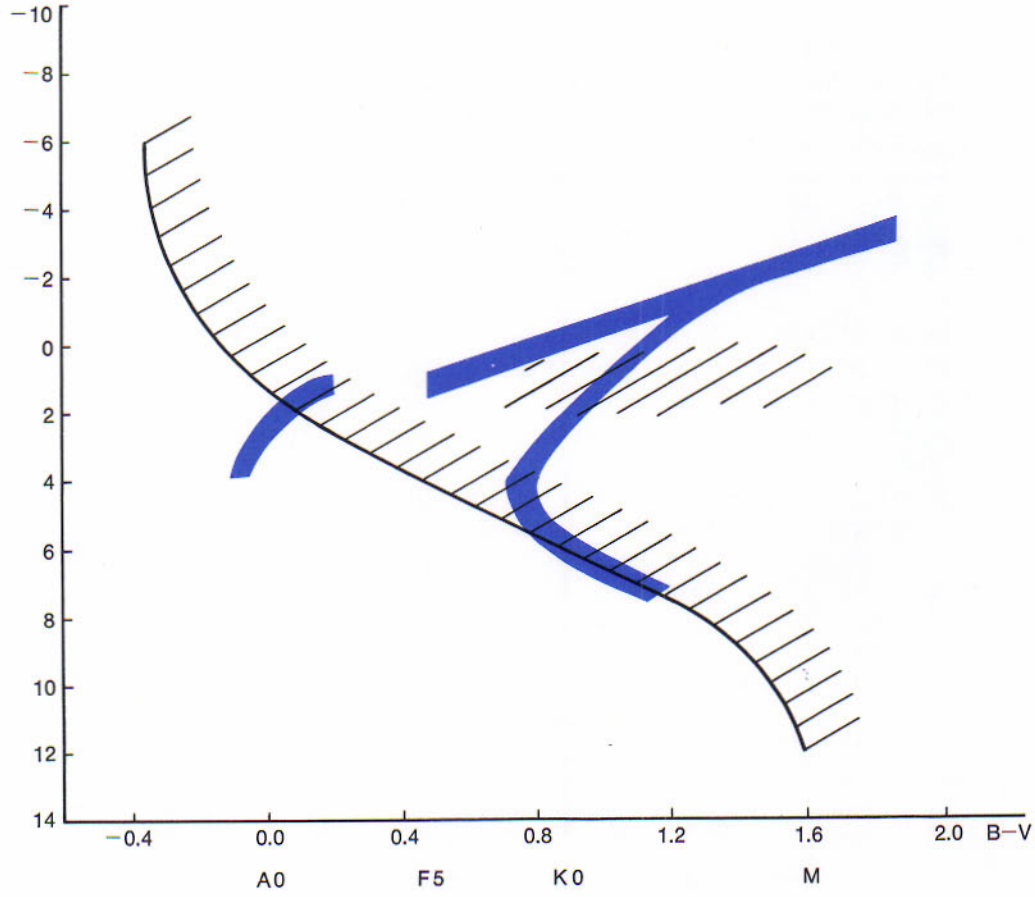


Figure 15.14
Spectrum-Luminosity Diagrams. Hatched area represents stars of Baade's type I population, as in Fig. 11.27. The shaded strips represent giants of his type II population. *(Adapted from the original diagram by Walter Baade)*

15.15 The Cluster Color-Magnitude Diagram Extended

Baade's original diagram for the type II population has been extended by the investigations of Arp, Baum, and Sandage. Their color-magnitude diagrams of the clusters M 3 and M 92 were derived from photographs with the 100-inch and 200-inch telescopes.

The diagram for M 3 is shown in Fig. 15.15. The nearly vertical giant branch joins the cluster main sequence at type F5, absolute magnitude +3.5. This sequence goes on downward from there, with slightly steeper slope than the normal main sequence, to magnitude +5.6, the limit of these photographs, and is entirely absent above the junction. The horizontal giant branch terminates at the left with extremely blue stars. The omission of the RR Lyrae variables leaves a gap in this branch, indicating that nonvariable stars are not generally present here.

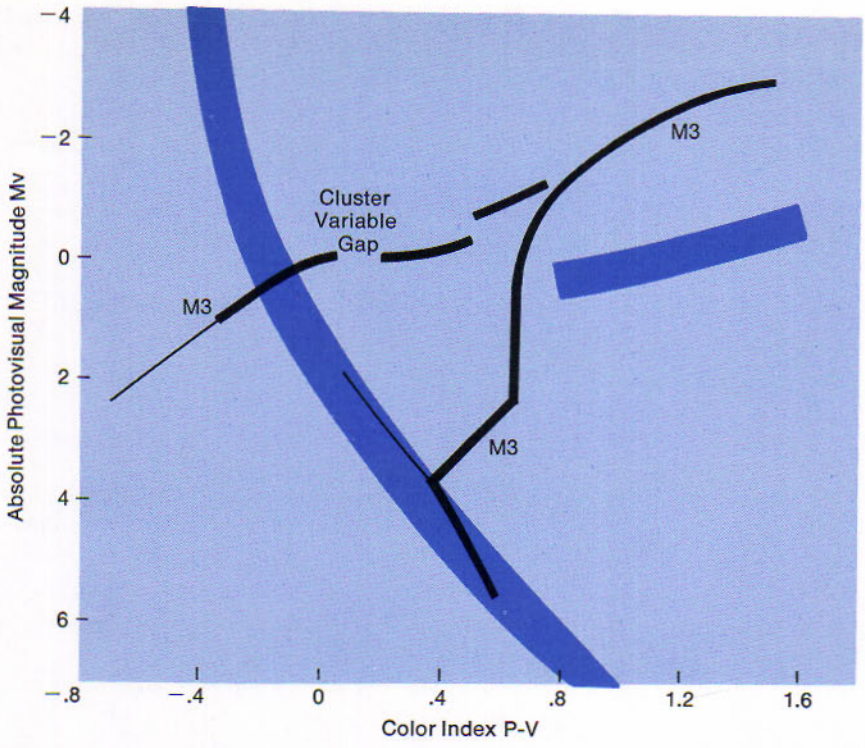


Figure 15.15
Color-Magnitude Diagram of the Globular Cluster M 3. The type II array of the cluster is shown by the solid lines. The shaded areas represent the main sequence and red giants of the type I population. *(Diagram by H. C. Arp, W. A. Baum, and A. R. Sandage)*

The point where the main sequence breaks away in the globular clusters indicates a very old age, 3 to 9 billion years, as suggested above. These clusters are so far away that we cannot measure their internal motions by standard astrometric techniques. An alternate method is to obtain the internal motion from the radial velocity component. To do this, high dispersion is required, together with the ability to look at single stars within the cluster. Since all the cluster stars are faint, the restrictions are severe, but the application of image tubes and space telescopes may allow us to use this approach.

15.16 Variety Among Star Clusters

Globular clusters of our galaxy have been distinguished from galactic clusters by their greater size, compactness, and luminosity. Although the impression long prevailed that globular clusters were very much alike, the present view is that the division of all star clusters into two types is too simple. This is particularly the case when the color-magnitude diagrams are available and when the clusters in other galaxies are included.

S. C. B. Gascoigne and G. E. Kron reported in 1952 that the integrated color indexes of objects known as globular clusters in the Magellanic Clouds (18.7) are about equally divided in two groups. The clusters of one group are red, colored by their brightest giant stars; they are more nearly like the normal globular clusters of our own galaxy. The clusters of the second group are blue, corresponding to the colors of their most luminous main sequence stars; they are unlike the galactic clusters in appearance and content. Brightest of the blue clusters in the Large Magellanic Cloud is NGC 1866, having a population exceeding 10,000 stars.

P. W. Hodge prefers to designate the two groups as "young populous clusters" and "old populous clusters," depending on whether their main sequences extend above absolute visual magnitude zero or fail to do so. H. C. Arp suggests that the term "globular cluster" be reserved for clusters of great age and low metal content, such as M 3.

As assemblages of older stars the red globular clusters of our galaxy have been regarded as free from interstellar gas and dust. Supplies of such material originally possessed by the clusters were believed to have condensed to form stars or to have been exhausted in other ways. Yet the dark regions observed in some of these clusters are caused most likely by clouds of gas and dust in the clusters themselves, as pointed out by M. S. Roberts. He concludes that the gas has been shed by stars evolving to the white dwarf stage. The gas may now be condensing into young stars. This, he believes, could account for the few blue main-sequence stars found above the turnoff points in color-magnitude diagrams of certain old clusters.

REVIEW QUESTIONS

1. What are the general characteristics of galactic clusters?
2. Discuss methods for obtaining distances to galactic clusters.
3. How far away is a cluster whose AO stars have an apparent magnitude of +10?
4. Sketch the life history of a galactic cluster.
5. How do we determine the ages of open clusters?
6. Why does the apparently old cluster M67 contain so many stars?
7. Where do stellar associations fit into the cluster scheme?
8. What are the general characteristics of globular clusters?
9. How did the globular clusters contribute to leading us to a galactocentric point of view?
10. The globular clusters led to the concept of stellar populations; what population makes up the globular clusters and what are its characteristics?
11. If the absolute magnitude of the RR Lyrae stars is actually one-half magnitude fainter than we presently believe, what does this do to the scale of the Galaxy?
12. What arguments can be given to show that the division of clusters into galactic (or open) clusters and globular clusters is too simple?

REFERENCES

Bok, Bart J., and Priscilla F. Bok, *The Milky Way*, 3rd ed., Cambridge, Massachusetts: 1957.

Shapley, Harlow, *Galaxies*, Rev. ed., Cambridge, Massachusetts: Harvard University Press, 1961.

FOR FURTHER STUDY

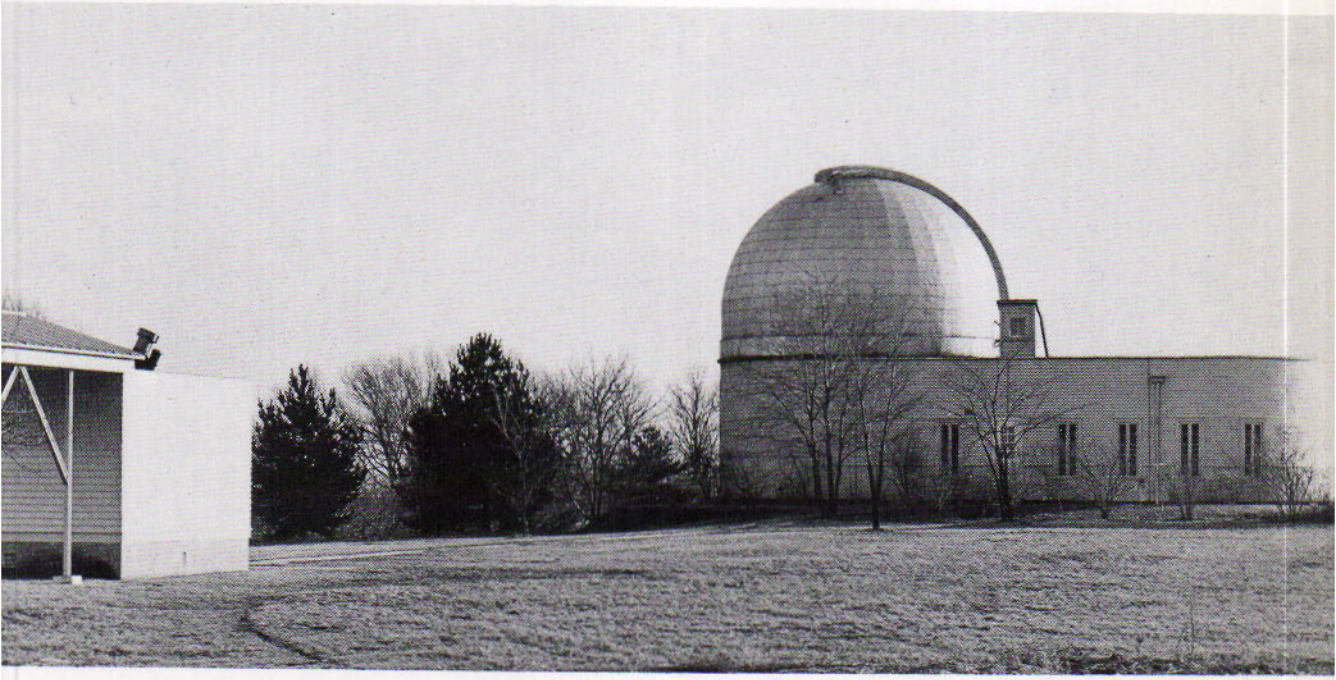
Blaauw, Adriaan A., and Maarten Schmidt, Eds., *Galactic Structure*, Chicago: University of Chicago Press, 1965.

Becker, Wilhelm, *Sterne and Sternsysteme* Dresden: Steinkopff-Verlag, 1942.

Mihailas, Dimitri and Paul M. Routly, *Galactic Astronomy*, San Francisco: Freeman, Cooper & Co., 1968.

Shapley, Harlow, *Galaxies*, rev. ed., Cambridge, Mass.: Harvard University Press, 1961.

Goethe Link Observatory of Indiana University. *(Indiana University Photograph)*




16

INTERSTELLAR GAS AND DUST

DIFFUSE NEBULAE — THE INTERSTELLAR MATERIAL — THE LIVES OF THE STARS

Faintly glowing spots in the heavens, excluding the comets, were called *nebulae* from early times. Some of these proved later to be remote star clusters. Others seemed to avoid the region of the Milky Way and came to be known as extragalactic nebulae. These are stellar systems outside our own galaxy, which are now simply called galaxies. The galaxies are described in Chapter 18.

The spaces between the stars in the sun's vicinity contain about 3 to 5% as much gas as there is in the stars themselves. The gas is accompanied by smaller amounts of dust. Concentrations of the interstellar material produce the more obvious bright and dark nebulae and are being effectively studied by radio techniques.

Eta Carinae nebula photographed in red light. *(Courtesy of B. Bok)*

DIFFUSE NEBULAE

The nebulae proper in our galaxy and in other galaxies as well are generally of the type known as diffuse nebulae. *Condensations in the interstellar material, they are clouds of gas and dust having irregular forms and often large angular dimensions. Some are made luminous by the radiations of neighboring stars and possibly by collisions between clouds, whereas others are practically dark. Nebulae of a different type (12.25) are gaseous envelopes around certain hot stars.*

Some diffuse nebulae resemble the cumulus clouds of our atmosphere. Others have a filamentary structure that is remindful of our high cirrus clouds. All are turbulent and are also moving as a whole in various directions. Where clouds come together and interpenetrate, the collision velocities are likely to exceed the speed of sound at these low temperatures. Shocks, compressions, and magnetic fields can account for the intricate structure of the nebulae.

16.1 The Great Nebula in Orion

This nebula is the brightest of the diffuse nebulae in the direct view. Scarcely visible to the naked eye, its place is marked by the middle star of the three in Orion's sword. Through the telescope it appears as a greenish cloud around the star, which itself is resolved into a group of type O stars.

In photographs with large telescopes the Orion nebula is spread over an area having twice the apparent diameter of the moon. At its distance of 500 parsecs, the nebula as shown in these photographs has a linear diameter of 7 parsecs, or about the distance of Vega from the sun. It is a prominent concentration of the nebulosity that is spread over much of the region of Orion. The nebula is accompanied by an aggregate of several hundred O and B stars in a giant O-B association.

See Plate XI

Many of the bright diffuse nebulae we see in photographs are invisible in the direct view with the telescope. These objects are intrinsically faint, and not made faint by great distance as a star may be. They would not be any easier to detect at a much closer range. Luminous areas have the same brightness per unit angular area regardless of distance. A change of a factor of two in the distance to a luminous area results in a four-fold change of the intensity of the light *and* of the area itself, consequently, the brightness per unit of angular area remains constant.

16.2 The Illumination of Nebulae

The presence of stars near or actually involved in the nebulae is mainly responsible for their shining. In the absence of such stars the nebulae are practically dark. This relation was demonstrated, in 1922, by E. Hubble at Mount Wilson Observatory. Particular stars can be selected that are

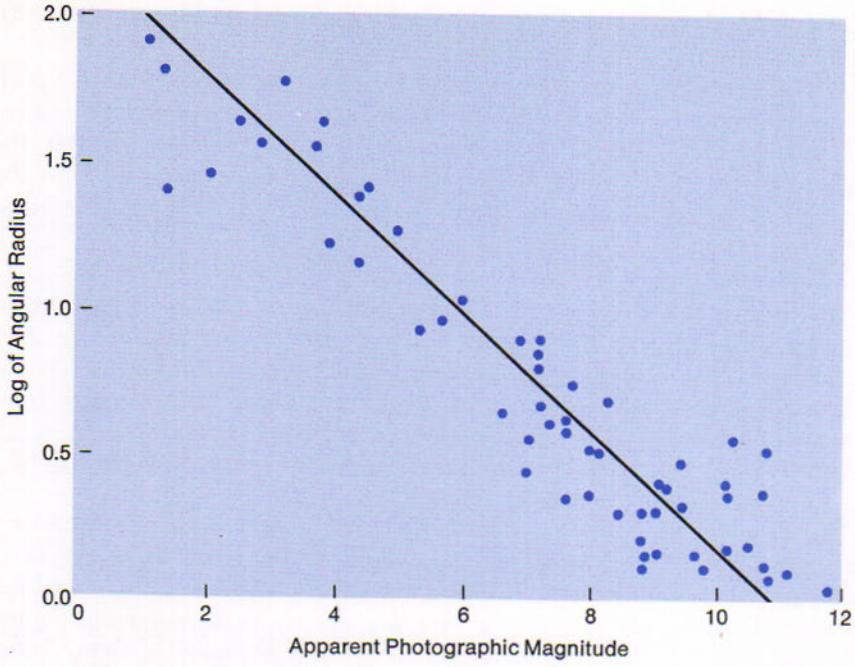


Figure 16.2
Increase in extent of nebular illumination with increasing brightness of involved stars. *(Adapted from a diagram by Edwin Hubble)*

associated with almost every known bright nebula, and in each case the radius of illumination of the nebula is roughly proportional to the square root of the brightness of the star (Fig. 16.2). A star of the 1st magnitude illuminates the nebula to an angular distance of 100′ in the photographs, whereas the effect of a 12th-magnitude star extends less than a minute of arc.

Another conclusion from Hubble's extensive investigations is the relation between the temperature of the associated star and the quality of the nebular light. If the star is as hot as type B1, the spectrum of the nebula differs from that of the star in having strong emission lines. If the star is cooler than B1, the nebular light resembles the starlight. Bright nebulae are accordingly of two types: emission nebulae and reflection nebulae and it is interesting to note that Fig. 16.2 holds for both types of nebulae.

16.3 Emission Nebulae

In the vicinity of a very hot star, where the radiation is largely in the ultraviolet, the hydrogen gas of an interstellar cloud is kept ionized. Radiation from the star of wavelength less than 912 Å can remove the electrons from normal hydrogen atoms. When substitute electrons are captured by the positively charged protons, the electrons may land in any one of the possible orbits and reach their lowest level by a series of transitions with the emission of light. Atoms of other elements in this

part of the cloud are also excited by such radiations or by collisions with other atoms and contribute to the illumination.

Thus the diffuse nebulae such as the great nebula in Orion, which have hot stars involved in them, are *emission nebulae.* Their light differs in quality from that of the stimulating stars. Other examples are the planetary nebulae, which have central stars of type O or W, and the envelopes of novae when their gases have become sufficiently attenuated by expansion.

Bengt Strömgren has shown that a single O star ionizes almost all hydrogen atoms in a gas of suitably low density to a distance of 50 parsecs. The Orion nebula has a much smaller radius than this because its density is unusually high, so that the effectiveness of the stars' radiations is more quickly diluted. In other, more common cases, electrons recombine with protons at the distance predicted by Strömgren and a very bright spherical nebula appears, referred to as the *Strömgren sphere.*

16.4 Bright-Line Spectra of Nebulae

The bright lines that characterize the spectra of emission nebulae are mainly lines of hydrogen, neutral helium, and oxygen and nitrogen in different stages of ionization. Prominent among them are a pair of lines at wavelengths 3726 Å and 3729 Å in the ultraviolet (combined at the left in Fig. 16.4), due to singly ionized oxygen, and a pair at 4959 Å and 5007 Å in the green, due to doubly ionized oxygen. The latter pair gives the greenish hue to such nebulae. Neither pair has as yet been observed in laboratory spectra of oxygen.

The identification of these two pairs of lines, and of additional lines of oxygen, nitrogen, and some other elements, rests on theoretical evidence given, in 1927, by I. S. Bowen. They are "forbidden lines," so called because the electron transitions producing them are much less likely to occur in a gas in ordinary conditions than are lines normally observed. The reverse is true in the rare and extended gas of the nebulae; the unusual lines are more likely to appear than the normal ones.

The relative strength of the bright lines does not at once show the relative abundance of the chemical elements in the gases of these nebulae. Oxygen and nitrogen are less abundant than hydrogen and helium, but in collisions with other atoms they are able to utilize greater quantities of energy provided by the exciting starlight.

With respect to the various lines, emission nebulae resemble planetary nebulae. The primary differences are in the ratio of the brightnesses of the oxygen lines which are very strong in planetary nebulae and the Hβ line which is relatively weaker than the nearby oxygen lines in the planetary nebulae than in the emission nebulae. This can be explained by the temperatures of the existing stars and the efficiency of oxygen under the conditions encountered.

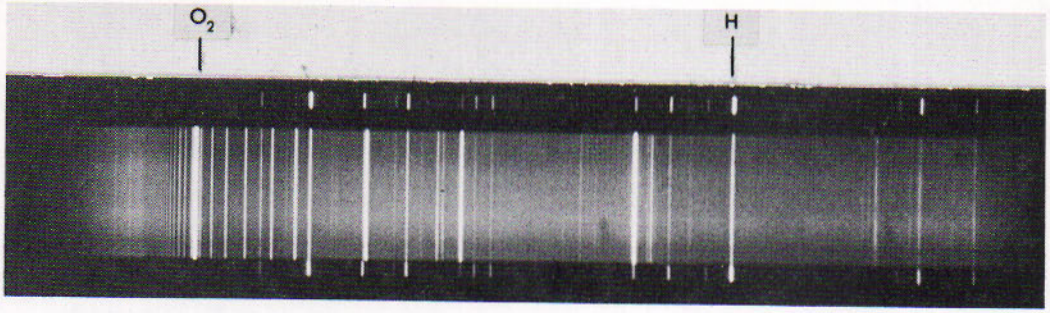


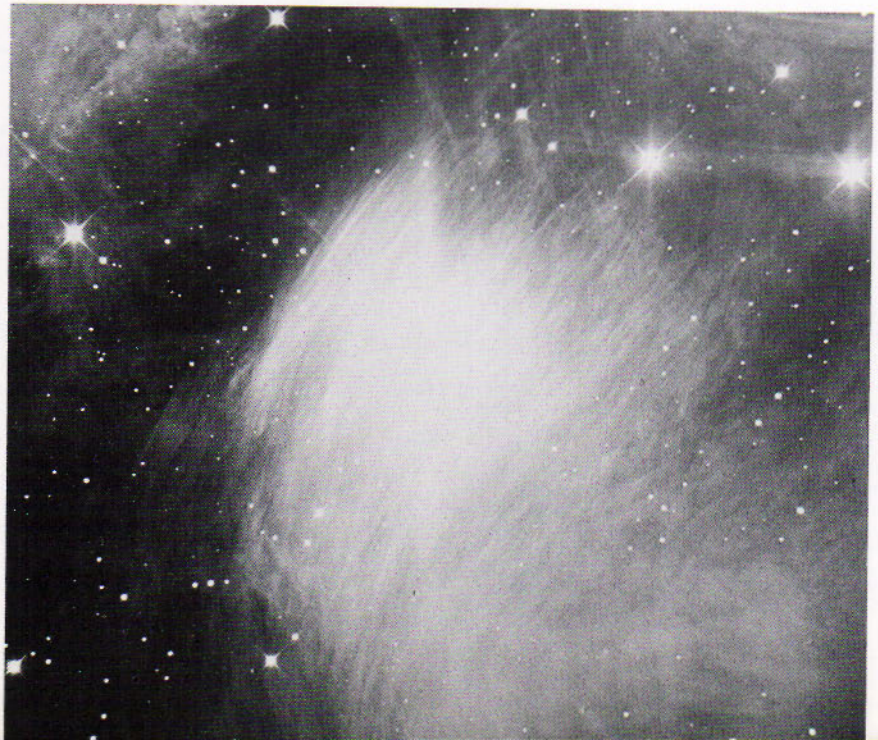
Figure 16.4
Spectrum of the Orion nebula. The forbidden oxygen doublet appears as a bright single line on the left. The bright line just to the right of center is Hβ. Comparison lines are hydrogen and helium. *(Photograph by D. E. Osterbrock, Washburn Observatory)*

16.5 Reflection Nebulae

Where the stars in the vicinities of the interstellar clouds are cooler than type B1, the nebulae are not noticeably self-luminous. Their spectra are the same as those of the associated stars as shown by V. M. Slipher. Examples are the nebulosities surrounding stars of the Pleiades. The light of such nebulae is starlight scattered by the dust particles that are present in the clouds along with the gas. Scattered starlight is, in fact, a constituent of all nebular light, but its presence in emission nebulae may be

See Plate V

Figure 16.5A
Nebulosity around Merope in the Pleiades. An excellent example of a reflection nebula showing organized structure. *(Photograph from the Hale Observatories)*



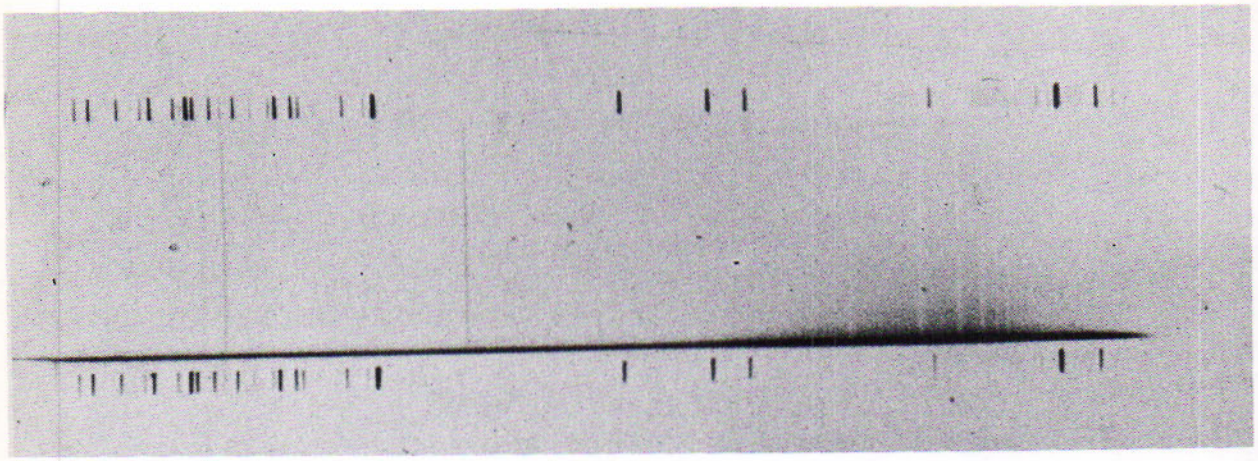


Figure 16.5B
Spectrum of a reflection nebula. The absorption lines are the same as those of stars. The emission lines are night sky lines. *(Yerkes Observatory Photograph)*

unnoticed. The emitted light concentrated in a few bright lines of the spectrum is much more conspicuous than the scattered light, which is spread over all wavelengths.

The colors of reflection nebulae are nearly the same as those of their associated stars. This is well demonstrated in the photographs of a cloudy region in Scorpius, where bright nebulae appear around the stars like the glow around street lamps on a foggy night. Thus the illumination around the red star Antares, which is scarcely noticeable in blue light, becomes conspicuous in yellow light. The opposite is true of the nebulae around the blue B stars.

16.6 Dark Nebulae

These nebulae have no stars nearby to illuminate them. They make their presence known by dimming the light of whatever lies behind. Some, such as the rift in Cygnus or the Coalsack in Crux, are visible to the naked eye. The majority are revealed in the photographs by their obscuration of bright regions of the Milky Way. The darkest clouds are relatively near us at distances of from 300 to 500 light years. Few dark nebulae are known in our galaxy that are more distant than 5000 light years. At greater distances their contrast with the bright background is diluted by stars in front of them.

The average dark nebula is not much more than 10 parsecs in diameter and contains material in excess of 50 solar masses. It reveals itself in star counts, and stars located behind it are often dimmed by three magnitudes (four or more magnitudes in a few extreme cases). Except for a few cases dark nebulae are irregular in shape. This reinforces the

Figure 16.6
"Horsehead" nebula in Orion, south of Zeta Orionis, IC 434, Barnard 33. *(Photograph from the Hale Observatories)*

idea that if a bright star or stars were located in or near one of these nebulae, it would be an emission or reflection nebula.

A great complex of cosmic clouds is centered in the region of Scorpius and Ophiuchus, and goes on northward to the Northern Cross, forming the Great Rift in the Milky Way. Another extends from Cassiopeia through Perseus and Auriga into Taurus.

THE INTERSTELLAR MATERIAL

In addition to its more conspicuous showing in the bright and dark nebulae, the interstellar material is recognized by the radiation of its neutral hydrogen as recorded with radio telescopes. The presence of intervening gas is also revealed by the dark lines it imprints in the spectra of stars, and that of dust is shown by its reddening and dimming of stars behind it.

16.7 Hydrogen in Interstellar Gas

Hydrogen is the most abundant element between the stars as it is in the stars themselves. It is accompanied by helium and by heavier gases in much smaller amounts. Where an interstellar cloud surrounds a hot star, the hydrogen is ionized and set glowing in a spherical region around the star, having a radius that is greater as the involved star is hotter. The part of the cloud outside the sphere is not ionized and is normally dark. Strömgren has called the two parts of the clouds the H-II and H-I regions.

The H-II Regions. The presence of clouds of ionized hydrogen over large areas of the Milky Way, in addition to the more obvious emission nebulae, was first observed by Struve and Elvey at McDonald Observatory, in 1937. They had devised a fast nebular spectrograph for detecting the faint emission lines and found them more often where large numbers of B and O stars are present in the vicinities. Other emission regions have since been found in photographs with wide-angle cameras, taken with plates sensitive to the part of the spectrum around the red hydrogen line. The emission nebulosities and associated blue stars are being employed with optical telescopes to trace the spiral arms of our galaxy (17.12). About 5 per cent of the gas in these arms is in the H-II form.

The H-I Regions. The clouds of neutral hydrogen are invisible by optical means. Their primary radiation is at a wavelength of 21 cm, which is

recorded with radio telescopes. Neutral (but excited) hydrogen has been observed in and near H-II regions by means of the recombination line between the 110 and 109 levels, whose emission has a wavelength of about 6 cm.

The lowest level of the hydrogen atom is really a pair of levels where the electron may be, depending on whether its magnetic spin is parallel or opposed to that of the nucleus of the atom. Transition of the electron from the upper to the lower level produces the radiation at 21 cm. Observable radiation of the clouds at this wavelength, predicted by van de Hulst in 1944, was first detected by the Harvard physicists H. I. Ewen and E. M. Purcell, in 1951. This radiation has effective use with radio telescopes in tracing spiral arms of the Galaxy (17.21) and in other investigations within and beyond the Galaxy.

The very existence of the 21 cm line attests to the enormous abundance of hydrogen. Once a hydrogen atom is in the upper level of the ground state it will remain there for about 10 million years; in other words, the transition probability is very low. Since the strength of a spectral line depends essentially upon the product of the number of atoms and the transition probability, it is obvious that there must be a great deal of hydrogen indeed.

16.8
Hydrogen Wreaths in Cygnus

See Plate XVI

The region of the Northern Cross exhibits a bewildering array of bright and dark clouds. We view this region along the arm of the Galaxy that includes the sun. Most remarkable in the display are the wreaths formed by filaments of faintly luminous nebulae. The smallest and brightest of these is the familiar Loop of nebulosity near Epsilon Cygni, of which the brightest parts are the filamentary nebulae NGC 6960 and 6992. Other wreaths appear only as arcs of circles that may be mostly concealed by other nebulae. The larger wreaths were unknown until they were revealed in a mosaic of photographs (Fig. 16.8) with the 48-inch Palomar Schmidt telescope. They are conspicuous in Sky Survey photographs with red-sensitive plates, which bring out the faint nebulae more clearly.

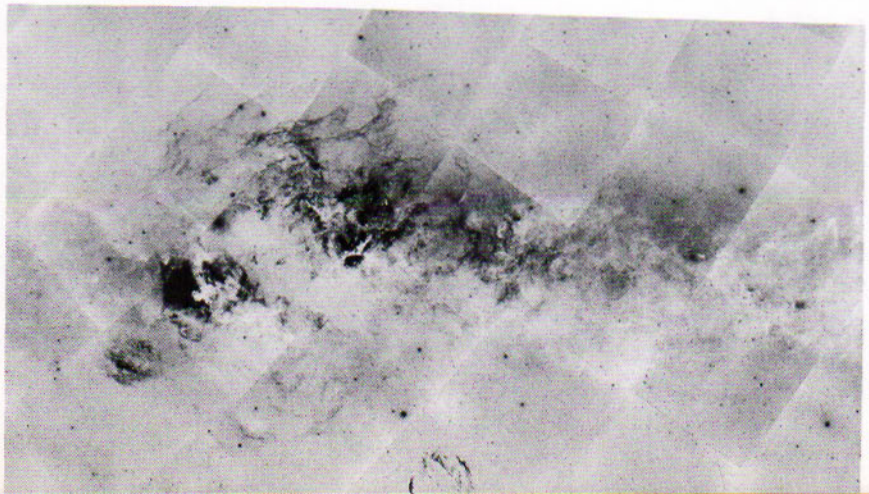


Figure 16.8
Hydrogen clouds and wreaths in Cygnus. In this mosaic of negative prints from the 48-inch Schmidt telescope emission nebulae and stars appear black. The familiar loop nebula is seen at the bottom of the plate. *(National Geographic Society and Mt Palomar Sky Survey, courtesy of J. L. Greenstein)*

All these formations suggest expanding shells. The Loop itself, at the distance of about 760 parsecs, is known to be increasing in radius, and it may have been slowed considerably from the original rate by resistance of dark material outside it. The expansion has been ascribed to the outburst of a supernova that may have occurred as much as 67,000 years ago, but the central star has not been identified. The illumination of this nebulosity is believed to be caused by collision of the expanding shell with the surrounding interstellar material, although its radio emission is typical of the remnants of a supernova.

16.9 Interstellar Lines

Many years ago, a "stationary" dark line was observed in the spectrum of the binary star Delta Orionis by J. Hamilton; this line is narrower than the lines in the spectrum of the star itself and does not oscillate with them in the period of its revolution. It was the first known example of *interstellar lines*, which are absorbed in the spectra of stars by intervening cosmic gas. The recognized constituents of the interstellar gas are neutral atoms of sodium, potassium, calcium, and iron, singly ionized atoms of calcium and titanium, water vapor and hydroxyl, and the cyanogen and hydrocarbon molecules, including singly ionized hydrocarbon and formaldehyde. The shape of the absorption line tells us how many atoms per square centimeter are involved. If we know the distance to the star or source in whose spectrum we observe the line we can calculate how many atoms per cubic centimeter there are. This way, an average of 1.7×10^{-24} grams per cubic centimeter of gas has been estimated for the Galaxy.

Interstellar atomic lines arise only in electron transitions from the lowest levels in the atoms. Such lines are likely to appear in the far ultraviolet where they are usually unobservable. This is true of hydrogen and some other elements such as carbon, nitrogen, and oxygen. Absorptions by molecules so far identified also involve transitions from their lowest levels; each band of their usually complex spectra is represented by only one or two narrow lines. In addition, there are a number of diffuse lines, which might be mistaken for stellar lines except for a difference in radial velocity. Their identifications are not as yet established.

Interstellar molecules
OH
CN
CO
CH
CH^+
H_2
NH_3
HCN
H_2O
H_2CO
H_2CO_2
HC_3N
C_2H_2O
CS
NH_2CHO

The frequent division of the interstellar lines into two or more components was shown by earlier studies of W. S. Adams at Mount Wilson. The division was later given clear and important interpretation by G. Münch, whose results were derived from spectra of distant stars photographed with the 200-inch telescope. They show that the material producing the interstellar lines is situated in or near the spiral arms of our galaxy. The interstellar neutral hydrogen also separates into two or more components. Observations in the direction of stars studied by Münch have shown sharp neutral hydrogen lines having the same velocities as

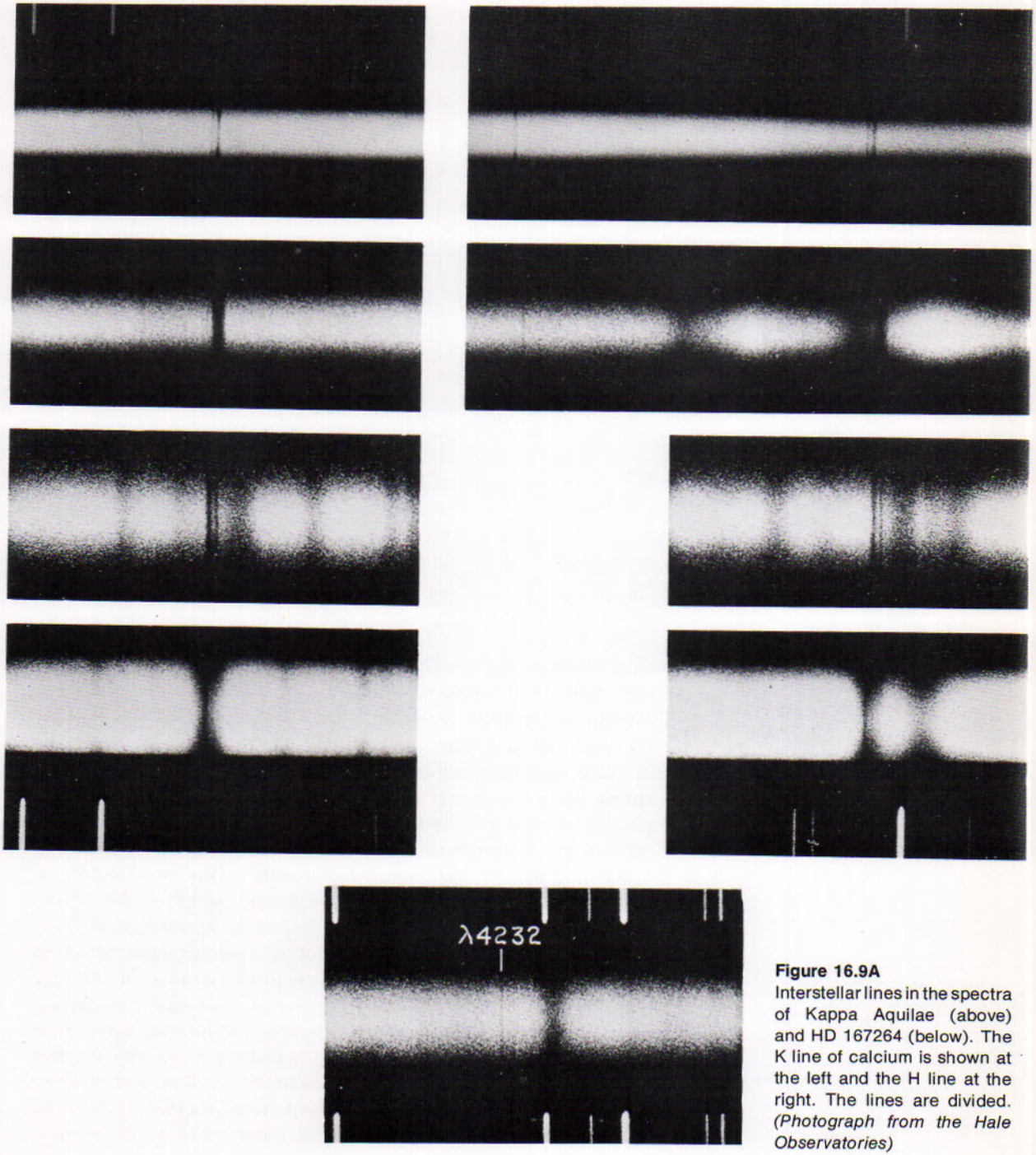


Figure 16.9A
Interstellar lines in the spectra of Kappa Aquilae (above) and HD 167264 (below). The K line of calcium is shown at the left and the H line at the right. The lines are divided. *(Photograph from the Hale Observatories)*

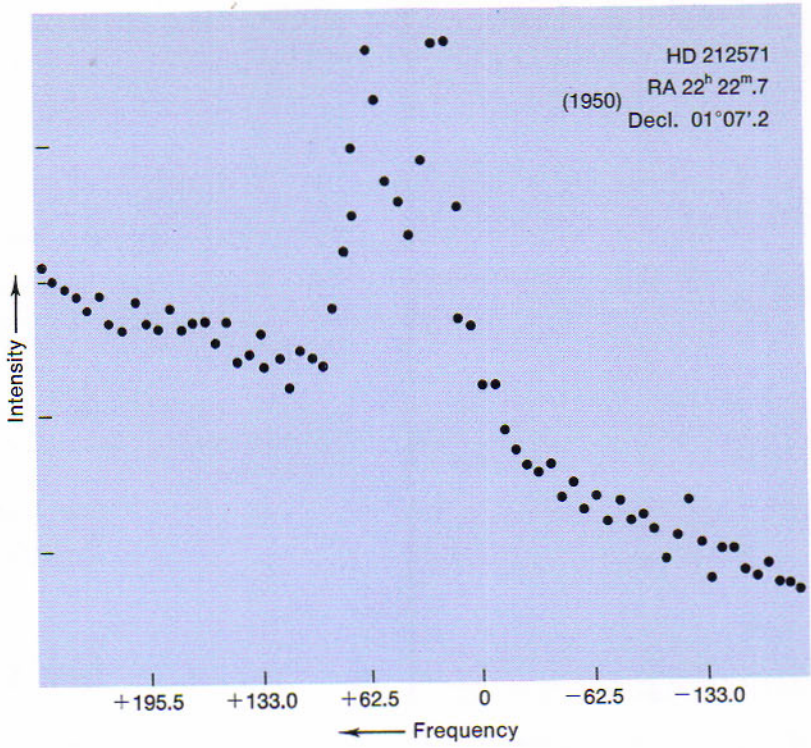


Figure 16.9B
Two sharp interstellar hydrogen lines in the direction of the Star HD212571. The abscissa is in kilohertz with the zero point at 1420.406 megahertz. The ordinate is in arbitrary intensity units. These observations were obtained at the National Radio Astronomy Observatory. *(Leander McCormick Observatory diagram)*

the interstellar sodium and calcium lines. The obvious conclusion is that these elements coexist in the same cloud.

Since the temperature of the interstellar gas is only about 40°K (although some astronomers use a temperature of 120°K) the gas is generally in the ground state and interstellar atomic lines can arise only in electron transitions from the lowest levels. We would expect, for example, that the strongest interstellar hydrogen line would be the Lyman Alpha (Lα) line at λ 1215 Å and this is the case. The strongest helium line will be a triple line near λ 537 Å and fortunately a line at λ 10,830 Å. In general, however, the principal lines occur in the far ultraviolet where they are unobservable from the surface of the earth, hence the astronomers' keen interest in obtaining high resolution spectrographs attached to space telescopes. The high resolution is required because of the low kinetic temperature of interstellar space. We recall that line broadening is a function of the square root of the temperature. While we know that temperature is not the only source of line broadening we can for the moment neglect the other sources and conclude that interstellar lines should be 10 times sharper than lines arising from a gas at 10,000°K. Indeed, even very weak interstellar lines are often quite easily noticed because of their extreme sharpness.

It is necessary to emphasize that hydrogen is so abundant in space that lines other than those of the Lyman series are observed. We have already mentioned the λ 21 cm line. Hydrogen recombination lines have been observed, for example, from the 110 level to the 109 level.

Where two arms intervene between the star and observer, the lines may be divided into two main components by Doppler effects of the rotation of the Galaxy, effects which we examine in the following chapter (17.16). A line produced by the gas of each arm may be further divided into several components by turbulence of the material or by several descrete clouds in the line of sight. As the starlight passes through the arm, it may be absorbed by clouds moving at different speeds in the line of sight and thus causing different Doppler displacements of the lines.

Some gas clouds are situated as much as 900 parsecs from the principal plane of the Galaxy; these have velocities of the order of 50 km/sec toward or away from the plane. At high latitudes and probably at great distances are clouds of hydrogen with velocities exceeding 100 km/sec toward the Galaxy. These high velocity clouds are considered to be intergalactic hydrogen falling into the Galaxy.

16.10 Dust Located by Counts of Stars

Interstellar dust dims the light of the stars beyond. Whether or not the dust is conspicuous in the photographs, its distance and effect can be determined by counts of stars in that area of the sky.

If the stars were equally luminous and uniformly distributed in the space around us, and if the space itself is perfectly transparent, the total number of stars brighter than a limiting apparent magnitude would increase 4 times for each fainter magnitude to which the limit is extended. Consider the total number of stars brighter than the 12th magnitude as compared with the total number brighter than the 11th magnitude.

Because the apparent brightness of equally luminous stars varies inversely as the squares of their distances from us, a 12th-magnitude star, which is apparently 1/2.5 as bright as an 11th-magnitude star, would be the square root of 2.5, or 1.6, times as far away. Thus the stars brighter than the 12th magnitude would occupy a volume of space around us that is 1.6^3, or nearly 4, times as large as the space occupied by the stars of the 11th magnitude; and with the assumption of equal distribution they would be 4 times as numerous. It is true that the stars are not equally luminous; but the luminosity function (11.31) can replace that assumption.

As an example, suppose that the total numbers of stars brighter than successive magnitudes are counted in a certain area and that the ratio of the numbers remains 4 up to the 11th magnitude but is reduced to 2 at the 12th magnitude. We conclude that the stars are uniformly distributed and that space is transparent to the distance represented by the 11th-magnitude stars. From there the stars either thin out or are dimmed by dust, or both. Other evidence is needed for the decision.

16.11 Reddening of Stars; Color Excess

Interstellar dust not only dims the stars beyond, but it makes them appear redder than their normal colors by scattering their blue light more than their red light. Similarly, the scattering of sunlight in our atmosphere reddens the setting sun. If in a particular part of the sky the stars more distant than 300 parsecs are reddened, there is dust at that distance.

The *color excess* of a star is the difference in magnitude by which the observed color index, blue minus visual, exceeds the accepted value for a star of its spectral type (Table 12.II), it is a measure of the reddening of the star by the dust. When the color excess is multiplied by an appropriate factor, we have the *photographic absorption*, that is, how much in magnitudes the star is dimmed by the dust as photographed with a blue-sensitive plate. A useful survey of color excesses is that of Stebbins, Huffer, and Whitford, who employed a photoelectric cell and color filters to determine the colors of 1332 O and B stars.

The distance of a reddened star, r in parsecs, is calculated by the formula: $5 \log r = m - M + 5 - K$, where m is the apparent photographic magnitude, M is the corresponding absolute magnitude for a star of this particular spectral class, and K is the photographic absorption. Average values for the absorption are usually adopted, an often used value being 0.8 magnitude per kiloparsec. Special studies by photoelectric means have indicated values as high as 10 magnitudes in heavily obscured regions such as Cygnus and values as low as 0.1 magnitudes toward the south galactic pole.

16.12 Dust Grains

The reddening of stars by intervening cosmic dust is attributed to particles smaller than 10^{-5} cm in diameter. Dust grains of this size would scatter the starlight inversely proportional to the wavelength, which is not far from the observed relation in the light of the reddened stars. The origin of the grains is not clearly understood; whether they form from gas in the interstellar medium or are particles blown into the medium from upper atmospheres of stars, or both, is conjectural. The dust contributes only one or two per cent to the total mass of the interstellar material.

Once the size of the grains is established it is possible to estimate the total number of grains required to give the average observed dimming. Unless other information is available we assume a uniform distribution, adopt a correction factor for shadowing effects, assume a mass for the average grain and arrive at the conclusion that the dust contributes about 1.3×10^{-26} grams per cubic centimeter to the mass of the Galaxy. Thus the dust contributes only about 1% to the total mass of the interstellar material and less than ½% to the mass of the Galaxy. We will see later that this contribution is strongly limited to the plane of the Milky Way.

Starlight is frequently polarized in its passage through the dust clouds, as J. S. Hall and W. A. Hiltner discovered independently in their photo-

electric studies. L. Davis and J. L. Greenstein have explained that this effect could be produced in the light by elongated grains rotating around their short diameters; the axes of the rotations would set themselves along magnetic lines of force.

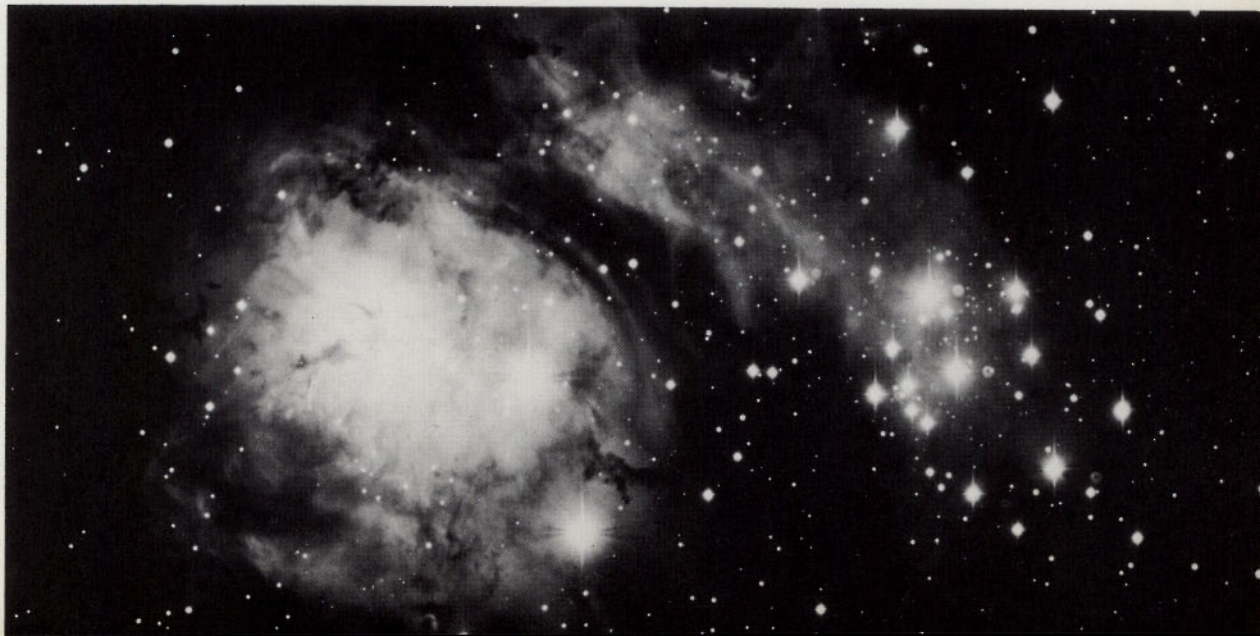
The galactic magnetic field is generally parallel to the plane of the galaxy and is to some extent knotty. All sources of study lead to a rather small average value of the magnetic field on the order of 10^{-3} gauss. Studies of nearby polarized sources, such as the pulsars, in two frequencies lead to an average value of about 7×10^{-4} gauss.

This weak magnetic field operating over great distances and eons probably plays a fundamental role in many cosmic phenomena. Squeezing of the field may trigger the collapse of interstellar clouds and thus begin star formation. The field may act as an accelerator for cosmic ray particles or perhaps as a cosmic storage ring for these particles. We are only beginning to appreciate what role this magnetic field can play from a theoretical point of view and the observational proofs are still to be obtained.

16.13 Globules of Dark Nebulae

Photographs in many parts of the Milky Way show small dark nebulae against backgrounds of star-rich regions and bright nebulosity. Some are like wind-blown wisps; others are oval or nearly circular, as if they had been inked in the photographs with a fine-pointed pen. B. J. Bok and others have drawn attention to the great numbers of such "globules." We see them, for example, projected upon the diffuse nebula M 8 (Fig. 16.13), and an average of one per square degree throughout that vicinity

Figure 16.13
Dark Globules in M 8. Stars are believed to be forming in the dark globules. Note the beautiful open cluster. Donut-like images are ghosts caused by the field flattening lenses. *(Courtesy of B. Bok)*



wherever the background is sufficiently bright. None at all, however, is reported in the region of the Orion nebula.

The globules have apparent diameters of from 5″ to 10′ or more, and linear diameters of the order of 10^4 to 10^5 astronomical units. Their photographic absorptions range from at least 5 magnitudes for the smaller objects to 1 magnitude for the larger ones. Cosmologically the globules have great interest if it can be shown that they represent a stage of evolution preceding the birth of stars.

THE LIVES OF THE STARS

Two features of stellar evolution have persisted in the successive theories of the past two centuries. The first is that stars condense from nebulae; the second that energy derived from gravitational contraction is the guiding process throughout the evolution. A third feature, that the stars are "cosmic crucibles" in which lighter chemical elements are built up into heavier ones, was prominent around the beginning of our century. It appears again in the latest theories, which also stress the importance of the interchange of material between stars and nebulae.

The literature of this recently revived and rapidly developing subject is already extensive. The brief account in the following sections may be supplemented, for example, by G. R. and E. M. Burbidge's review article on Stellar Evolution in Handbuch der Physik, *Volume 51, 1958, and by M. Schwarzschild's* Structure and Evolution of the Stars.

16.14 Birth of Stars in Interstellar Clouds

It is supposed that some external phenomena exert a pressure on a considerably dense cloud of gas and dust to the point where the cloud will contract under gravity. The initial cloud may have a diameter of several parsecs, breaking up into fragments of an average diameter of one parsec. When this material has condensed to a diameter of the order of 10,000 astronomical units, the "protostar" may perhaps be observed as a dark globule. More highly heated by further contraction, the star finally becomes hot enough to shine; and it may then blow away enough excess dust around it so that we can better see what is going on.

Where the process begins with interstellar material of normal density,

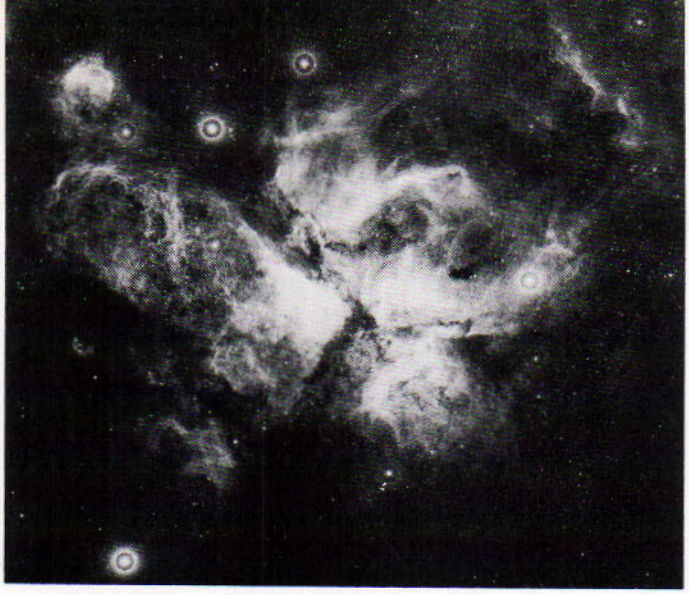


Figure 16.14
Eta Carinae nebula, photographed in red (Hα) light. Curtis-Schmidt photograph, Cerro Tololo. *(Courtesy of B. Bok)*

the cloud requires a mass of at least 1000 suns in order to condense under its own gravitation and is likely to fracture later into many parts. This is the beginning of a cluster or an association of stars.

16.15 From Nebulae to the Main Sequence

The initial energy released by the contraction of the interstellar cloud is mostly absorbed by the cold interstellar grains (40°K). The collapse is essentially in the form of gravitational free fall. For a time, the interior heats up unnoticed by the falling outer layers and suddenly the star becomes visible and establishes a quasi-hydrostatic equilibrium. It must be kept in mind that it has only taken the star 30 to 100 years to evolve from the start of collapse until it becomes visible, being larger and brighter than it will be when it reaches the main sequence. For example, a star destined to be similar to the sun when it reaches the main sequence, will be about 300 times brighter and slightly redder according to Hayashi.

Energy generated inside the distended star is carried up to the surface in large convective cells. The star then continues to collapse, with very little change in its surface brightness thus evolving almost straight down the H-R diagram, as shown in Figure 16.15. After about one million years the star is sufficiently small and dense for the large convective cells to be dissipated, and radiative equilibrium is established. At this point the star stops evolving downward and moves horizontally to its position on the main sequence. The various time scales and traces in the diagram vary depending upon the mass of the star, and are referred to as Hayashi evolutionary tracks.

During the horizontal phase of the evolution toward the main sequence, the stars may vary irregularly in brightness. A. H. Joy called attention to these *T Tauri stars* in 1945. He observed that the light variations of 40 stars in a heavily clouded region of Taurus are accompanied by bright lines and continuous emission of varying intensity in their spectra. More recently, G. Haro, G. H. Herbig, and others observed similar emission

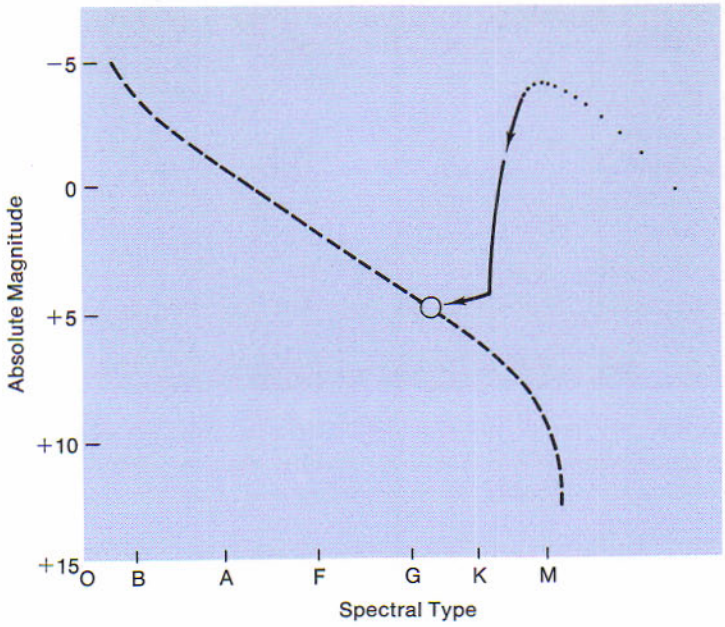


Figure 16.15
The Hayashi evolutionary track for a solar type star. Radiative equilibrium sets in at the sharp break in the track.

in the spectra of yellow and red dwarf stars in clouded regions of Taurus, Monoceros, and Orion. These effects are believed to be caused by instability of the young stars rather more than by their interactions with the dust clouds.

When a group of youthful stars approaches the main sequence of the color-magnitude diagram, the most massive stars settle in the bluest parts of the sequence; they arrive in advance of the others because they contract the more rapidly. The less massive stars array themselves later in order of decreasing mass along the redder parts of the sequence. The cores of all these stars are now hot enough to promote the synthesis of hydrogen into helium with the release of sufficient energy to keep the stars shining. Contraction is halted for a time at this stage; the stars change little in size, temperature, and brightness for a long period. Here is the reason why the majority of stars in the sun's vicinity are members of the main sequence.

Some stars do not have sufficient mass to start a self-sustaining conversion of hydrogen into helium and evolve downward completely outside the main sequence. They are highly degenerate and shine only on account of light resulting from gravitational contraction. S. S. Kumar places the lower limit on stable internal reactions at stars of 0.1 solar masses. He points out that all stars below this mass must evolve directly into "black dwarfs" and cites the binary L726-8 as an example. This system, discovered by W. Luyten in 1949, is a binary whose total mass is 0.08 solar

masses and the brighter component apparently has a much less massive invisible companion. Kumar, Huang and others feel that there must be myriads of such objects and even that Jupiter is such an object.

16.16 Evolution of Cluster Stars

A star remains in the original main sequence as long as the release of energy in its interior is just enough to supply a constant rate of radiation. When the hydrogen in the core is nearly exhausted, the core resumes its contraction, growing hotter and promoting further nuclear reactions outside it. The outer layers expand. The star then becomes brighter and begins to move upward and to the right in the diagram.

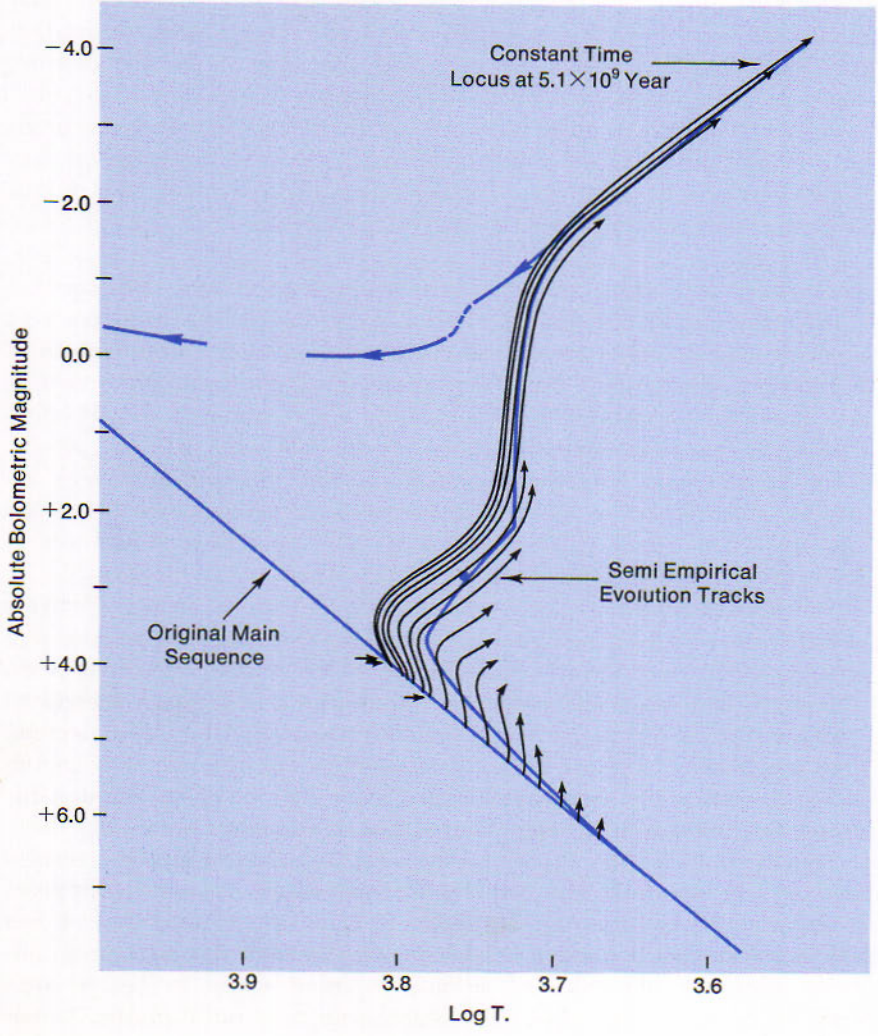


Figure 16.16
Semiempirical evolution tracks of stars in the globular cluster M 3. Values of log T_e correspond to the following spectral types: 3.9 to A5, 3.8 to F5, 3.7 to K0 for the main sequence and G0 for giants, 3.6 to K5 for the main sequence and K0 for giants. *(Diagram by A. Sandage)*

Because the evolution is faster as the stars are bluer, the slope of the new main sequence becomes steeper, as Chandrasekhar and Schönberg have shown, up to the point where the stars have left the sequence entirely. Although this effect is less conspicuous where the stars are continually being born, it shows clearly in the galactic clusters, as we have seen (15.3), and also in the globular clusters.

A. Sandage has determined empirically the evolutionary tracks (Fig. 16.16) in the globular cluster M 3 of stars in the interval of 5.1 billion years since they left the original main sequence with absolute magnitudes between +3.98 and +4.5. By comparing the present luminosity function of the cluster with the initial function (Fig. 11.31) for the main-sequence stars, he found the final magnitude for each track, and located each endpoint horizontally in the diagram by use of the color-magnitude diagram of this cluster. The heavier curve through the points so determined lies along the more nearly vertical branch of the type II giant sequence.

A supergiant star of magnitude −4.0 at the top of this curve was originally of magnitude +3.98. The original main-sequence stars brighter than +3.98 may now be moving toward the left on the horizontal giant branch or may already have collapsed to become white dwarfs.

16.17 Abundances of the Chemical Elements

The present abundances of the elements offer one of the most powerful clues to the history of the stars. Hydrogen accounts for about 93 per cent of all the atoms and 76 per cent of all the mass of matter in the universe. Helium is second with about 7 per cent of the atoms and 23 per cent of the mass. All the other elements together contribute only a little more than 1 per cent to the mass.

Fig. 16.17 shows how the logarithms of the relative abundances are arrayed with respect to the atomic numbers of the elements in the sun and stars. The zigzag line represents the corresponding abundances in the earth and meteorites.

The abundance curve drops rather abruptly for the lighter elements and then levels off at number 60. Most of the points define a curve well enough to promote inquiry about a few more conspicuous departures. The atoms of lithium and beryllium (numbers 3 and 4, below the curve) undergo nuclear disintegration at temperatures around a million degrees; they are likely to unite with protons and then to separate into helium atoms. The atoms of iron (number 26, above the curve) are quite stable except at extremely high temperatures, above 2 billion degrees.

Two theories of the origin of the chemical elements have received considerable attention. The first, by G. Gamow and associates, supposes that the elements were built up from neutrons all in the course of half an hour following a possible explosion that initiated the expanding universe. Although the observed abundance curve agrees in general with their theoretical curve, based on successive neutron captures, it has

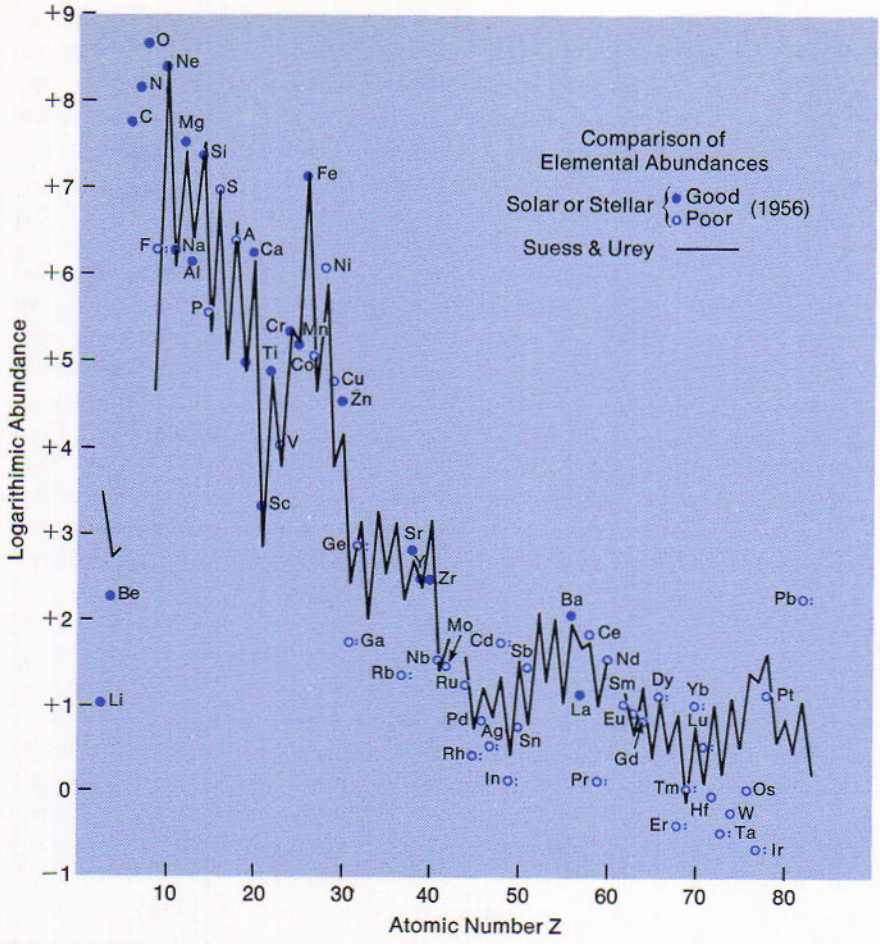


Figure 16.17
Relative abundances of elements in the sun and stars. The zigzag line represents the corresponding abundances in the earth and meteorites. *(Diagram by J. L. Greenstein)*

seemed doubtful to some other investigators that this process could have gone successfully beyond the bottlenecks of the unstable helium isotope, weight 5, and beryllium, weight 8.

The second and more recent theory, as explained particularly by W. A. Fowler and associates, supposes that the elements have been and are still being synthesized in the interiors of evolving stars.

16.18 Synthesis of Helium in the Sun

The more recent theory of the origin of the chemical elements begins with the proton-proton reaction in the cores of stars while they are approaching the main sequence. At 5 million degrees the protons are fusing into deuterons. These later collide with other protons to form

He^3, which finally combine to produce ordinary helium, He^4. This reaction goes on effectively at central temperatures, about 13 million degrees, of ordinary stars such as the sun. When some carbon is present and when the temperature becomes as high as 20 million degrees, the carbon cycle is an additional means of burning hydrogen. The fusion of hydrogen into helium is the main process in all stars and is the only one that is expected in a main-sequence star not considerably more massive than the sun.

The sun itself, estimated to be 6 billion years old, is scheduled to remain near the main sequence for an equal period in the future. By the end of that period the core of the sun, originally containing 12 per cent of the total hydrogen supply, will have become pure helium, according to the theory. Having run out of available fuel, the core will contract rapidly and grow much hotter. The expanding mantle around it will then burn its hydrogen at a furious rate.

Relatively soon thereafter at a central temperature of 100 million degrees the sun will be a red giant 30 times its present diameter and 100 times its present brightness. Its hydrogen will be nearly exhausted and its temperature will not be high enough for any considerable burning of helium. This evolution track is obtained by Sandage by transformation from the color-magnitude diagram of the galactic cluster M 67, where the stars have masses about equal to that of the sun.

With little fuel remaining, the sun will presumably contract as quickly as it had expanded. It may move to the left across the middle level of the color-magnitude diagram, from red to yellow to blue, and will then fade down to become a white dwarf, a faintly glowing cinder of its former splendor. From this point on, it will gradually cool and fade until it becomes a black dwarf. The time scale from the white to black dwarf stage is quite long, probably on the order of 500 billion years.

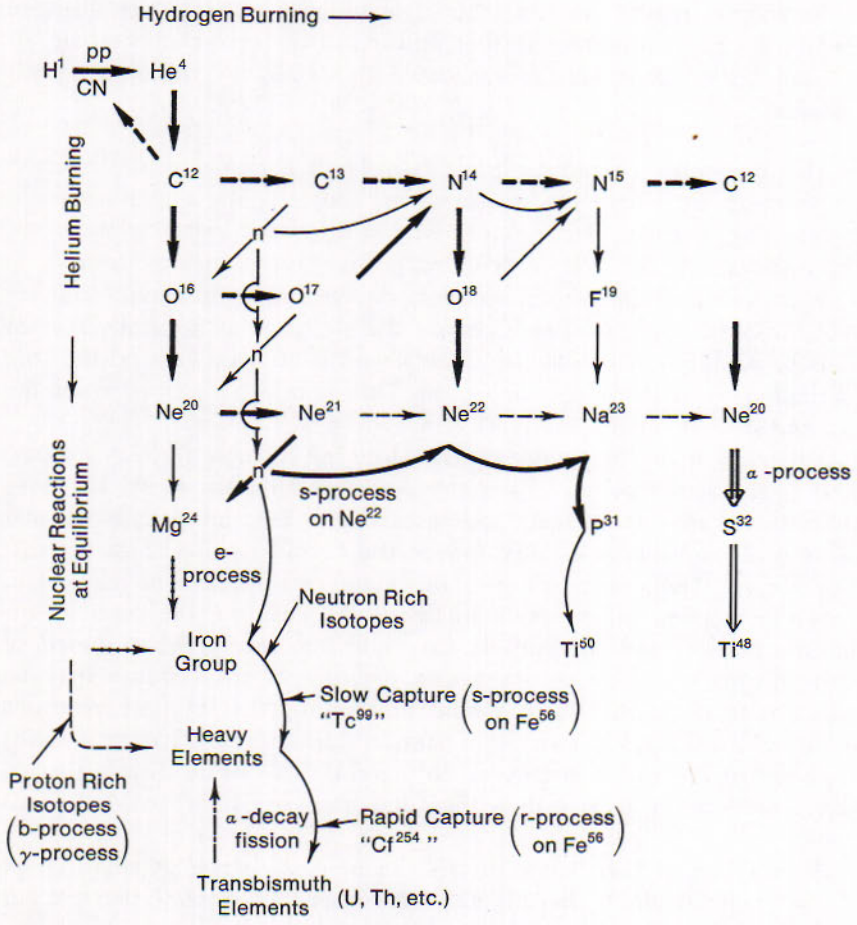
The evolutionary picture given for the sun holds for stars between 0.1 and about 1.5 solar masses. Below this mass interval, the protostar never initiates a self sustaining energy production but evolves directly into the black dwarf stage, as we have already seen. Above this interval, a violent adjustment must take place, since as Chandrasekhar has shown, degenerate objects cannot have masses greater than about 1.4 solar masses.

16.19 Synthesis of Elements in More Massive Stars

Stars more massive than the sun, and particularly the bluest stars of the main sequence, have shorter and more spectacular careers. They attain higher central temperatures, which may run into billion degrees when they become supergiants. Such stars are considered capable of building up practically all known chemical elements. The synthesis of elements may proceed mainly in three successive stages (Fig. 16.19).

1. *Hydrogen burning.* The fusion of hydrogen into helium in the cores of these more massive stars may be principally by means of the carbon cycle. The synthesis of helium spreads into the mantles and is practically complete at the central temperature of 100 million degrees.
2. *Helium burning.* The fusion of helium into heavier elements can occur at 150 million degrees. Helium nuclei may then combine to form

Figure 16.19
Synthesis of elements in stars. Hydrogen burning is the main process. In the more massive stars helium burning may build up heavier elements, and neutron capture may extend the synthesis to the heaviest elements. This chart appeared in an article by G. R. Burbidge, E. M. Burbidge, W. A. Fowler, and F. Hoyle in *Reviews of Modern Physics*, volume 29. *(Courtesy of W. A. Fowler)*



carbon, the reverse of a process already accomplished in the laboratory. Carbon may then fuse with other helium to form oxygen, neon, and magnesium. At a temperature of 5 billion degrees the build-up may go as far as iron, the final synthesis that can release energy for the stars' radiations.

3. *Neutron capture.* The building of elements from iron on may occur by successive fusions with neutrons, as in the Gamow theory. Neutrons are now abundant, having been released in syntheses of the second stage.

As the core builds up to pure iron, the only external evidence is the extreme red giant condition. The atmosphere is essentially unaffected by the previous history of the core and, except for some small changes due to convective mixing, has its original abundance ratio. When the core reaches the pure iron stage further burning in the core cannot occur. At this point the atmosphere is supported essentially by the electron gas pressure.

16.20 Final Evolution of Massive Stars

As the atmosphere contracts gravitationally, it compresses the electron gas to the point where the electrons can penetrate the iron nucleus. In the ensuing reaction, electrons combine with protons forming neutrons, and transforming the iron nucleus into a manganese nucleus unstable to β decay. The new nucleus immediately decays and the process continues until the overlying atmosphere presses the electrons so forcefully that an effective equilibrium is established between the electron gas and the core continuing the isotope of manganese. This removes electrons from the gas, and collapse of the overlying atmosphere takes place.

Depending upon the energy generated by the collapse a nova or supernova phenomenon occurs. If the star stabilizes after the event, the nova phenomenon may occur again and again at very long intervals as the star sheds mass and works its way toward the Chandrasekhar mass limit. If the collapse is sufficiently energetic many different phenomena may occur.

At the moment of explosion following the collapse, the core, if sufficiently massive and compressed, may split into several white dwarfs or white dwarfs and neutron stars. The pressure of the collapse may be sufficient to crush the electrons and protons together to form neutrons and if sufficient mass is involved a neutron star is formed, having a radius of about 10 km and a density of 10^{14} g/cm^3. The white dwarfs formed may have masses greater than the Chandrasekhar limit and be quasi-stable.

The stability of such white dwarfs can be easily upset. A small influx of matter, debris from the original event or matter from the interstellar medium, falling onto the star upsets the delicate balance and collapse

immediately sets in. Under proper conditions the collapse ceases and dampens out at the neutron star configuration. However, if the collapse energy causes an *implosion,* the neutron star stability range is destroyed and the star goes into total collapse and becomes a singularity in the space-time continuum. The singularity is called a *black hole.*

We believe that the above, very abbreviated discussion, is close to the actual sequence of events, being at least the best that theory can tell us for the present. Are there any observations that will support this theory? White dwarfs at any reasonable distance are very difficult to detect and of course so are singularities, but this is not the case with neutron stars.

Neutron stars should be sources of X-ray radiation and probably pulse-like phenomena; perhaps the pulsars are neutron stars. In the Crab nebulae, a magnificent cosmic laboratory, we seem to be able to observe all of the ingredients predicted by the theory. There are several pulsars associated with this nebula as predicted by the fragmentation theory and they are sources of X-ray radiation.

The time scales for these dramatic events are not really known. Flickering is suspected to go on for weeks or months before the nova outburst occurs, but this event must occur within a very short time. The post-nova decay, on the other hand, takes months or years. We have already seen that once a white dwarf falls below the Chandrasekhar limit it is stable and cools over a very long period of time. The time period for neutron stars being "visible" is not known, but from the Crab nebula we know that they have lasted more than 400 years—if the pulsars are neutron stars. Estimates place their lifetime at 1000 years.

Finally, for stars that collapse beyond the neutron star configuration, an interesting paradox occurs. In the time frame of the collapsing star the entire event is over in a flash, but in the time frame of an external observer the event lasts infinitely long as the collapsing star gets redder and redder and fainter and fainter.

16.21 Interchange of Material Between Stars and Nebulae

The stars are believed to be formed by condensation of interstellar material, as we have seen. The stars also keep returning gas and solid particles as well to the interstellar medium. Material issues explosively from supernovae and novae. It streams away continuously from Wolf-Rayet stars, P Cygni stars, red supergiant stars, main-sequence stars like the sun, and from rapidly rotating single stars and close double stars.

A supernova can blow into space material equal to 1 or 2 solar masses. A normal nova returns less than a thousandths of this amount in a single explosion; but such novae are far the more numerous, and explosions may occur repeatedly in the same star. A. J. Deutsch estimates that a red supergiant is losing material at the rate of 1 solar mass per 10 million

years. He concludes that of the "dead stars" in the sun's vicinity, which have left the main sequence and have already completed their evolutions, half of their original mass has been returned to the cosmic clouds and the other half is now in the form of white dwarf stars.

16.22 Metals in Successive Generations of Stars

If, as is now supposed, the heavier chemical elements are formed in evolving stars, then the interstellar medium is being enriched more and more in these elements. If the stars of the first generation condensed from cosmic gas that was pure hydrogen, if they built up metallic atoms in their interiors and eventually returned much of the enriched gas to the cosmic clouds, the second-generation stars formed in these clouds would contain a percentage of metals from the start. Third-generation stars, such as the sun is said to be, would begin their evolutions with a higher percentage of metals than did the second generation.

This conclusion seems to be supported by studies of stellar spectra. Strömgren has reported that the percentage content of metals is from 0.1 to 1 in old stars, 2 in stars of middle age, and 3 in very young stars.

16.23 Planetary Systems

It is considered highly probable that other planetary systems exist throughout the Galaxy and the universe. A few nearby stars reveal the presence of a planet or planets having masses on the order of that of Jupiter. It is almost inconceivable that the solar system is unique and the trend in evolutionary theory seems to indicate that under normal protostar conditions planetary systems may develop. The discovery of extra-solar system planets is achieved by observing changes in a star's proper motion, changes in its radial velocity or from the observing of periodic shallow eclipses.

Another method of discovering such planetary systems would be the reception of radio or optical signals from intelligent life on these planets. However the conditions for life are stringent. Su-Shu Huang and others point out that stars suitable for life on planets attending it must be hot enough to warm a deep habitable zone around them. Such stars must shine for a time long enough to allow the process to proceed. Here on earth, he points out, rational animals evolved from the earliest forms of life in something like three billion years. Too bright stars would emit types of radiation disruptive to this long term process. Most favorable of all stars would be those of types F, G, and early K which constitute about 10% of all the stars.

A possible indication that the sun is not unique in possessing a family of planets has been mentioned (14.26). Blue main-sequence stars are likely to rotate swiftly, whereas corresponding yellow and red stars have more moderate speeds. The break from fast to slow rotation comes abruptly

Figure 16.23
A small computer of the type used extensively in astronomy for specialized computing and controlling telescopes, both radio and optical. *(Digital Equipment Corporation, Maynard, Mass.)*

at type F5. It would seem that the redder stars have planetary systems to which they have imparted much of their original angular momentum. Thus the sun rotates in a period of about a month, and carries only 2 per cent of the entire momentum of the solar system.

To detect intelligent life by observations in the electromagnetic spectrum requires great skill and patience and a very good guess at what frequencies whoever would emit signals would use. F. D. Drake proposed the λ 21 cm line because all intelligent known beings able to build radio-transmitting equipment must know about it; others have proposed the H and K lines of calcium, etc. Drake even went as far as to observe two solar type stars τ Ceti and ε Eridani for 150 hours (without success).

Early success would be most surprising indeed, for several factors should be borne in mind. First, there is the problem of simultaneity; another civilization only 90 parsecs away would have to have reached our level about 350 years ago to communicate with us now. If they did so 2000 years ago they have had about 1700 years to improve their communications techniques and were probably using techniques so advanced 300 years ago as to be beyond our present capability to understand.

Second, our present instruments are really only sensitive enough to observe to a limited range, about 15 parsecs with a 100 m radio telescope and about the same distance with a 250 cm (100 inch) optical telescope and a high dispersion spectrograph looking for modulated signals in the calcium K line. Finally, statistically, we do not expect a system with intelligent life to be any nearer than 300 to 500 parsecs assuming that such systems are scattered randomly throughout the Galaxy.

REVIEW QUESTIONS

1. What are the differences between emission nebulae, reflection nebulae and dark nebulae? How do we discover dark nebulae?
2. Why are oxygen and nitrogen emission lines in nebulae often stronger than hydrogen lines even though hydrogen is the more abundant?
3. How do we determine the density of an element in the interstellar medium from its absorption line or lines?
4. Describe and explain the Strömgren sphere. How is it different for an O and A type central star? How could it be observed by radio techniques?
5. We believe that interstellar grains are elongated objects or perhaps even needle like. Why?
6. What do you deduce from the folowing table of idealized star counts in a small area in the direction of Cygnus?

Magnitude	10	11	12	13	14	15	16
Count	33	132	528	528	528	2112	8448

7. A seeming paradox exists in the statement that a star chronologically young may in fact be very old in its evolution. Explain.
8. Describe the Hayashi evolutionary track for stars of 0.5, 1 and 10 solar masses.
9. What is the Chandrasekhar limit and its importance to stellar evolution?
10. What are the possible end products of stellar evolution?
11. List the evidence that stars are returning material to the interstellar medium.
12. What techniques can be used to discover planets outside the solar system?

REFERENCES

Dufay, Jean, *Galactic Nebulae and Interstellar Matter,* trans. A. J. Pomerans, New York, Philosophical Library, 1957.

Johnson, Martin, *Astronomy of Stellar Energy and Decay,* New York, Dover Publications, 1959.

Page, Thornton and L. W. Page, eds., *The Evolution of Stars,* New York: The Macmillan Co., 1968.

Rush, J. H. *The Dawn of Life,* Garden City, N.Y., Hanover House, 1957.

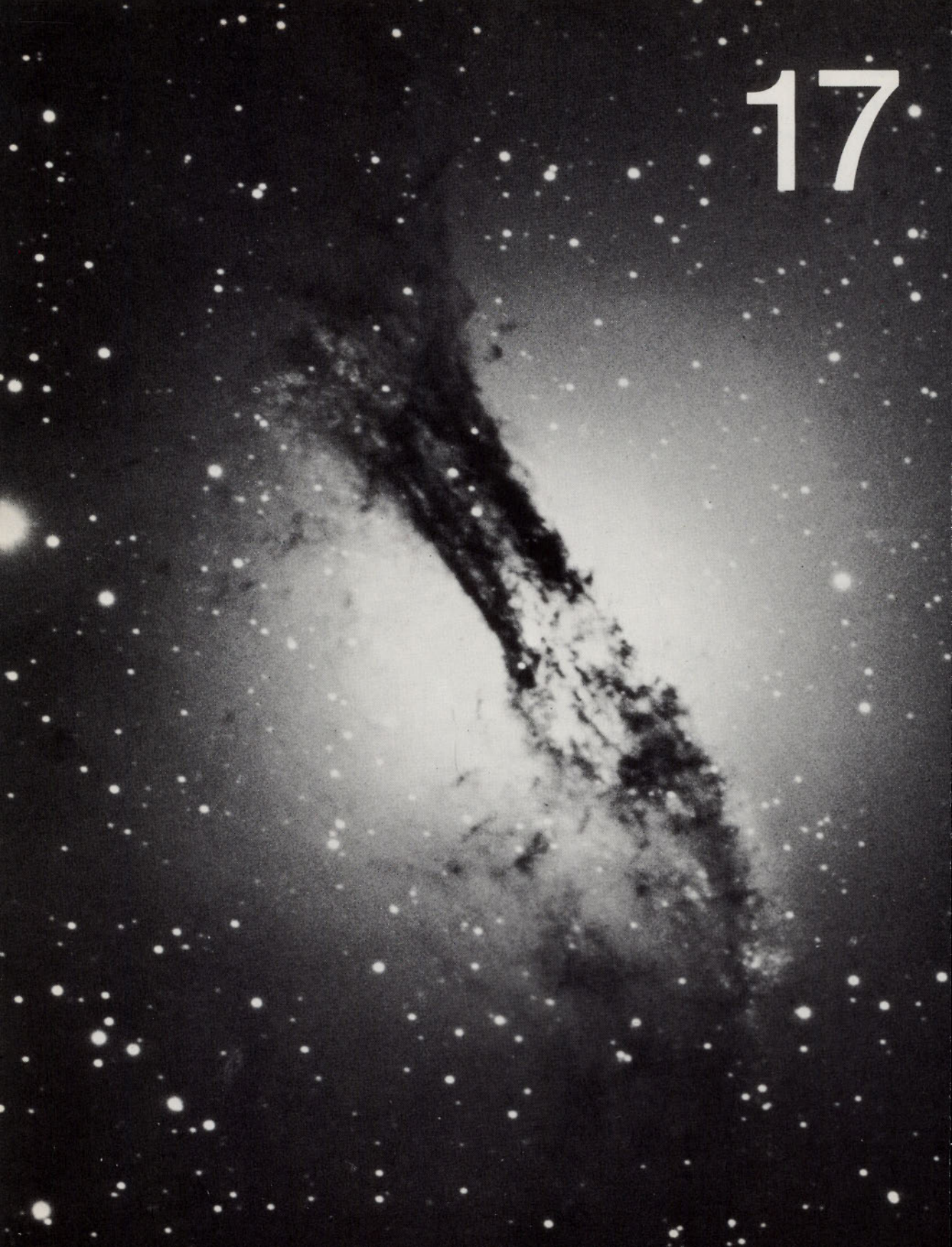
FOR FURTHER STUDY

Aller, Lawrence H., *Gaseous Nebulae,* New York: John Wiley and Sons, 1956.

Middlehurst, Barbara and Lawrence H. Aller, eds., *Nebulae and Interstellar Matter,* (Vol. 7 of *Stars and Stellars Systems*) Chicago: University of Chicago Press, 1968.

Kraus, John D., *Radio Astronomy,* New York: McGraw-Hill Book Co., 1966.
Schwarzschild, Martin, *Structure and Evolution of the Stars,* Princeton: Princeton University Press, 1958.
Spitzer, Lyman, *Diffuse Matter in Space,* New York: Interscience Publishers, 1968.
Stein, R. F. and A. G. W. Cameron, eds., *Stellar Evolution,* New York: Plenum Press, 1966.
Struve, Otto, *Stellar Evolution,* Princeton: Princeton University Press, 1950.
Wickramasinghe, Nalin C., *Interstellar Grains,* London: Chapman and Hill, 1967.

The Arthur J. Dyer Observatory of Vanderbilt University. *(Vanderbilt University photograph)*


17

THE GALAXY

THE MILKY WAY — STRUCTURAL FEATURES OF THE GALAXY — ROTATION OF THE GALAXY — RADIO VIEW OF THE GALAXY

The *galactic system,* or system of the Milky Way, is so named because a prominent feature of our view from inside it is the band of the Milky Way around the heavens. The system is a spiral galaxy of stars. Its spiral arms, one of which includes the sun, contain much gas and dust as well as stars. It is commonly called *the Galaxy* in distinction from the multitudes of other galaxies.

This chapter begins with a description of the Milky Way as it appears to the naked eye and in the photographs. The appearance suggests that the main body of the Galaxy is much flattened and that the sun is far from the center. The galactic structure and rotation are next considered, and finally the progress being made in exploring the Galaxy with radio telescopes.

NGC 5128. *(Photograph from the Hale Observatories)*

THE MILKY WAY

The faint hazy band of light easily visible on a dark night in the summer has the descriptive name of the Milky Way. It is visible on winter nights as well but is less conspicuous in northern latitudes. Photographs of the Milky Way reveal literally millions of stars, glowing gas and dark clouds. The galactic coordinate system uses the Milky Way to define its "equator".

17.1 The Milky Way of Summer

The *Milky Way* is the glowing belt of the sky formed by the combined light of vast numbers of stars. Its central line is nearly a great circle of the celestial sphere, which is highly inclined to the celestial equator. Because of its inclination, the course of the Milky Way across the sky is quite different at different hours of the night and at the same hour through the year.

Figure 17.1
The Milky Way from Scutum to Scorpius. The great Sagittarius star clouds are near the center of the picture. *(Photograph from the Hale Observatories)*

At nightfall in the late summer in middle northern latitudes, the Milky Way arches overhead from the northeast to the southwest horizon. It extends through Perseus, Cassiopeia, and Cepheus as a single band of varying width. Beginning in the fine region of the Northern Cross overhead, it is apparently divided into two parallel streams by the Great Rift, which is conspicuous as far as Sagittarius and Scorpius. The western branch of the Milky Way is the broader and brighter one through Cygnus. Farther south, in Ophiuchus, this branch fades and nearly vanishes behind the dense dust clouds, coming out again in Scorpius. The eastern branch grows brighter as it goes southward and gathers into the great star clouds of Scutum and Sagittarius. Here, in Barnard's words, "the stars pile up in great cumulus masses like summer clouds."

17.2 The Milky Way of Winter

In the evening skies of the late winter in middle northern latitudes the Milky Way again passes nearly overhead, now from northwest to southeast. The stream is thinner here and undivided. From Cassiopeia to Gemini it is narrowed by a series of nearby dust clouds, which cause a pronounced obscuration north of Cassiopeia and angle down through Auriga to the southern side of the band in Taurus. The Milky Way becomes broader, weaker, and less noticeably obscured as it passes east of Orion and Canis Major down toward Carina.

The part of the Milky Way nearest the south celestial pole is either quite out of sight or else too near the horizon for a favorable view anywhere in the United States. This part is conspicuous for observers farther south as it passes through Centaurus, Crux, and Carina; and the Great Rift continues the division as far as Crux (Fig. 17.2A). There is a fine star cloud in Norma and another in Carina, and there is the black Coalsack near the Southern Cross.

Figure 17.2A
Region of the Southern Cross. α and β Centauri are on the left. The absorbing cloud known as the coalsack is in the center. *(Harvard College Observatory photograph)*

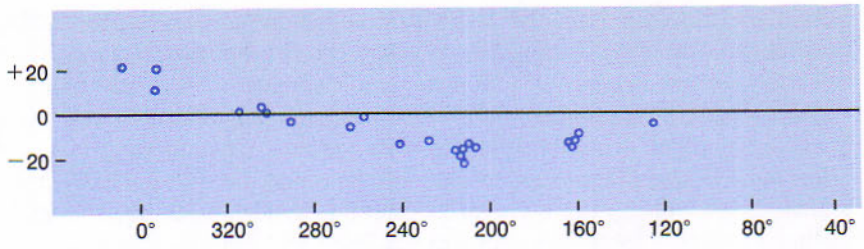


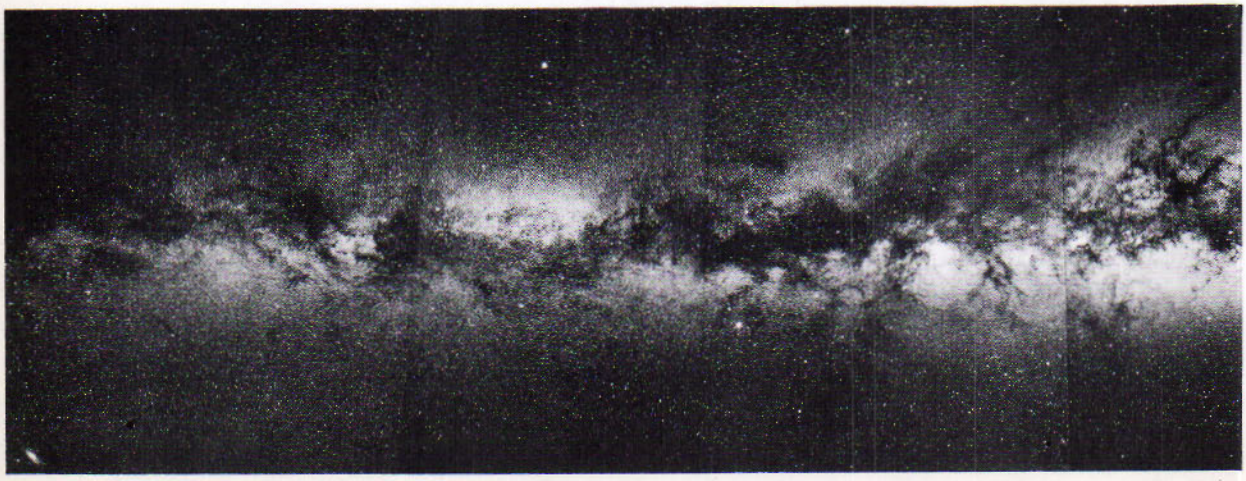
Figure 17.2B
Gould's belt. Plot of *O* and *B* stars nearer than 400 parsecs. Note evidence of tilt. *(After S. Sharpless)*

A plot of O and B stars within 400 parsecs of the sun in galactic coordinates shows a concentration in the winter sky and a clear tilt to the galactic plane (Fig. 17.2B). This is not surprising as this was noted by Gould in 1879 and is referred to as Gould's Belt. A study of the winter sky reveals a string of blue stars running through the Pleiades, the Hyades and down through Orion. This is the nearby Orion arm. The tilt must have been caused by a gravitational disturbance arising from an interaction with a nearby galaxy (or galaxies), perhaps the Magellanic Clouds. Such interactions are quite common among galaxies as we shall see and even the great galaxy in Andromeda (M31) has evidence of such an effect.

17.3 Photographs of the Milky Way

The general features of the Milky Way are best displayed to the naked eye or with very short-focus cameras. The details are well shown in the photographs with wide-angle telescopes. E. E. Barnard was a pioneer in this field. Fifty of his finest photographs are contained in his *Photographic Atlas of Selected Regions of the Milky Way.* These were made with the 10-inch Bruce telescope at Mount Wilson and Williams Bay. A more recent col-

Figure 17.3
Mosaic of photographs of the Milky Way from Sagittarius to Cassiopeia. Note M 31 on the lower left. *(Photograph from the Hale Observatories)*



lection is available in the *Atlas of the Northern Milky Way*, prepared by F. E. Ross and Mary R. Calvert; these photographs were taken with a 5-inch Ross camera at Mount Wilson and Flagstaff.

The latest and most penetrating representation of the Milky Way, north of declination −27°, is contained in the negative prints of the National Geographic Society-Palomar Observatory Sky Survey (4.16) made in blue and red light with the 48-inch Schmidt telescope.

Photographs reproduced in this chapter and elsewhere in the book illustrate the variety in different parts of the Milky Way. Fig. 17.1 shows the most spectacular part of the Milky Way, from Scutum to Scorpius; the Scutum and Sagittarius star clouds and the Ophiuchus dark cloud are prominent features of this region. Fig. 17.2A shows the region containing the Southern Cross and the Coalsack.

17.4 Galactic Longitude and Latitude

In studies relating to the Galaxy it is often convenient to denote the position of a celestial body with reference to the circle of symmetry of the Milky Way. For this purpose we define a system of circles of the celestial sphere in addition to the three systems described in Chapter 1. This system is based on the plane of the galactic equator, which passes through the sun.

The north and south *galactic poles* are the two opposite points that are farthest from the central circle of the Milky Way. By international agreement they are respectively in right ascension 12^h49^m, declination +27°.4 referred to the equinox of 1950, in Coma Berenices, and 0^h49^m, −27°.4, south of Beta Ceti.

The *galactic equator* is the great circle halfway between the galactic poles; it is inclined 63° to the celestial equator, crossing from south to north in Aquila and from north to south at the opposite point east of Orion. The galactic equator passes nearest the north celestial pole in Cassiopeia and nearest the south celestial pole in the vicinity of the Southern Cross.

Galactic longitude was formerly measured in degrees from the intersection of the galactic and celestial equators in Aquila, near R. A. 18^h40^m. By decision of the International Astronomical Union in 1958, the zero of galactic longitude is changed to the direction of the galactic center (17.10) on a slightly revised galactic equator, in R. A. $17^h42^m.4$, Decl. −28° 55′ (1950) in Sagittarius. As before, the longitude is measured around through 360° in the counterclockwise direction as viewed from the north galactic pole; its new values equal the former ones plus about 32°.

Galactic latitude is measured from 0° at the galactic equator to 90° at its poles and is positive toward the north galactic pole. In the sections that follow it will be specified whether the galactic coordinates are given in the former or the new system.

STRUCTURAL FEATURES OF THE GALAXY

Early studies of the Milky Way were aimed at explaining the structure of the universe. These studies revealed, among other things, that the Milky Way was limited and only one galaxy among many. More refined studies yielded the size and general characteristics of the Galaxy.

17.5 Early Hypothesis

The early problem of the extent and structure of the stellar system began when the stars came to be regarded as remote suns at various distances from us. The problem took the following form: Do the stars extend indefinitely into space, or is the system of stars around us bounded? If it is limited in extent, what are the size and the form of the system? The philosopher Kant, in 1755, was one of the first to imagine that the system has finite boundaries and that the nebulae might be other "universes."

William Herschel, in England, was the pioneer in the observational approach. His first attempt to determine the "construction of the heavens" was described by him in 1784. It was based on counts of all stars visible in the field of his telescope when it was directed to different parts of the sky. He supposed that the extension of the system in any particular direction was proportional to the cube root of the number of stars counted in that direction. Herschel's results contributed little more than the obvious conclusion that the stellar system is much extended toward the Milky Way. Analysis of star counts, however, long remained the favored method of the exploration and is still quite useful in studies of obscured regions in order to determine the distance to them. If we count all stars down to a given magnitude (m) and if the stars are uniformly distributed and then we count one magnitude fainter (m + 1) we should now have 3.98 more stars. Any sharp deviation from this ratio indicates the presence of an obscuring body. A uniform decrease in this ratio would indicate a falling off in the density. Kapteyn observed just such a falling off in his count ratios and this led him to propose a "universe" with the sun located very near its center.

The difficulties with the analysis of the early star counts lay with the

assumptions and interstellar absorption. The latter causes the stars to seem to be farther away and hence introduces an artificial thinning out. A hidden assumption was that on the average all of the stars were of the same brightness. Current applications of star counting take into account the interstellar absorption and carefully select the stars by spectral type so that the assumption of equal brightness is very nearly true.

17.6 The Recent Advances

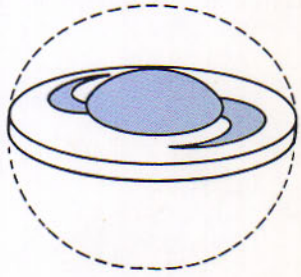
The modern era in the studies of the galactic structure began, in 1917, with H. Shapley's researches on the globular clusters. Shapley showed that the system of the Milky Way has finite dimensions and that the sun is far from its center.

V. M. Slipher's earlier discovery that the mysterious spirals and associated "nebulae" were receding from the sun at fantastic velocities led E. Hubble to study the biggest such object (M31) in detail. Hubble's discovery, in 1924, of cepheid variable stars in M31 conclusively showed that it was an independent assemblage of stars—a galaxy—and, consequently placed our own galaxy in a more proper perspective.

Another forward step was made by R. J. Trumpler, in 1930. His investigations of galactic clusters revealed for the first time that dimming of the view in many directions by cosmic dust must be allowed for.

Meanwhile, the studies of many astronomers with optical and radio telescopes are providing an increasingly clear picture of the Galaxy. The principal features are as follows:

The Galaxy is an assemblage of 100,000 million stars together with much gas and dust. Its spheroidal central region is surrounded by a flat disk of stars 24,000 parsecs in diameter, in which spiral arms of stars, gas, and dust are embedded. The galactic center is about 10,000 parsecs from the sun in the direction of Sagittarius. This flat main body of the Galaxy is rotating around an axis joining the galactic poles. Around the main body is a more slowly rotating and more nearly spherical halo containing "high velocity" stars and the globular clusters.



17.7 Stellar Populations in the Galaxy

Baade's recognition of his two types of stellar population in the universe was promoted by his studies of the spiral galaxy M31 in Andromeda, which structurally resembles our own Galaxy. His conclusion was that stars in the spiral arms of that galaxy, where there is an abundance of interstellar gas and dust, belong to the young type I population. Stars in the central region, where little gas and dust remain, belong mainly to the old type II population. A similar situation seemed to exist in our own galaxy.

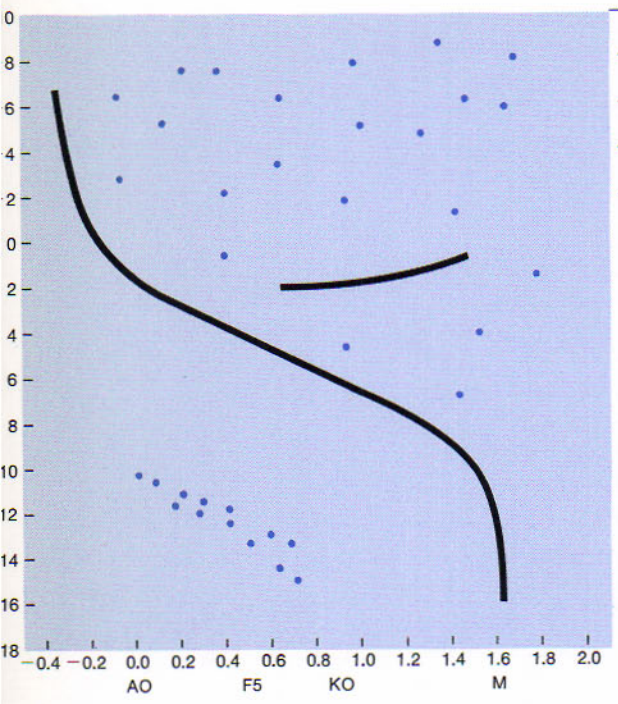


Figure 17.7A
H-R diagram showing population I distribution.

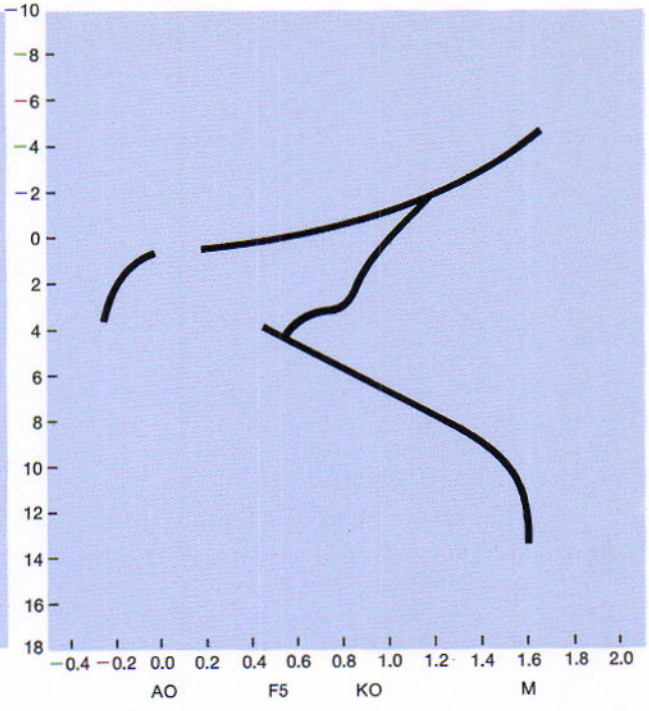


Figure 17.7B
H-R diagram showing population II distribution.

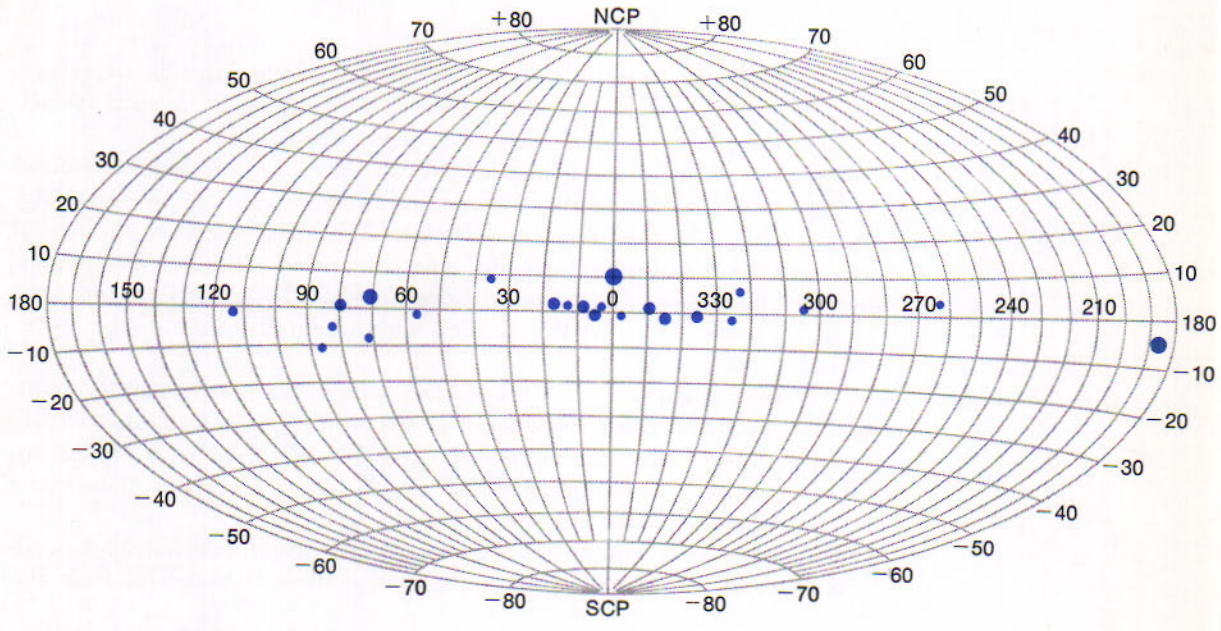
Meanwhile, the accumulation of observational data and the development of an attractive theory of stellar evolution have suggested a gradation of types, the precise number being a matter of convenience. The following sequence of five types of population in the Galaxy in order of increasing age of the stars was proposed by a conference of astronomers at Rome in 1957.

1. *Extreme population I.* This very young population is contained in the spiral arms, where much gas and dust is still uncondensed into stars, and to a limited extent around the center of the Galaxy as well. Its brightest members are blue supergiant stars, such as Rigel, only a few million years old.
2. *Intermediate population I* comprises somewhat older stars, such as Sirius, situated near the principal plane of the Galaxy, but not confined to the arms.

3. *Disk population.* The majority of the stars between the arms and many in the central region of the Galaxy belong to this type; they range from 3 to 5 billion years in age. The sun is believed to be a member.
4. *Intermediate population II* comprises many older stars in the halo and central region of the Galaxy.
5. *Extreme population II,* the oldest population, is represented by the older globular clusters and the separate stars of the halo. An age of at least 7 or 8 billion years is assigned to this group.

Attempts to fit classes of objects into population types can be misleading. However, often there is nothing else that can be done. For example, when we plot all of the observed X-ray sources (1970) on galactic coordinates as in Figure 17.7C we find that all but one of the sources fall within 10° of the galactic equator. From this we can conclude at least tentatively that the X-ray sources are either disk population, intermediate population I or extreme population I.

Figure 17.7C
X-ray sources in the Galaxy, plotted on galactic coordinates.



17.8 The Central Region of the Galaxy

This is a spheroidal concentration of stars 3,700 parsecs or more in diameter. Here the stars are crowded rather uniformly two or three times as closely as they are in the sun's vicinity. This region near the borders of Sagittarius, Scorpius, and Ophiuchus would be remarkably bright if it were not mostly obscured from the direct view and ordinary photography by dense dust clouds of the Great Rift. The bright star cloud of Sagittarius is an exposed portion.

Radiation from the central region has been recorded through the dust by use of a photocell and an infrared filter, together especially sensitive to radiation having a wavelength of 10,300 Å. Employing this apparatus on the 60-inch Mount Wilson telescope, J. Stebbins and A. E. Whitford made a series of sweeps across the part of the Rift immediately west of the Sagittarius cloud. They recorded an area of maximum radiation extending 8° in galactic longitude, 4° or 5° in latitude, and centered in longitude 326°.5.

J. Dufay and associates at the Haute Provence Observatory in France have made infrared photographs of the central region with a 12-inch Schmidt telescope, which are in general agreement with the photoelectric sweeps. These show the region divided, where the infrared radiation from stars behind did not penetrate the dust. Many radio wavelengths do penetrate the dust quite easily so that radiation originating in the very center of the Galaxy is observed. We discuss these observations in the sections following 17.17.

17.9 The Flat Disk of the Galaxy

The appearance of the Milky Way itself tells of two features of our galaxy. First, the disk of stars surrounding the central region is much flattened; second, the sun is far from the center of the disk.

The stars crowd toward the Milky Way. From the dull regions around the galactic poles the numbers of stars in equal areas of the sky generally increase with decreasing galactic latitude. The stars visible to the naked eye are 3 or 4 times as numerous near the galactic equator as they are near its poles; and the increase exceeds 40-fold for stars visible with large telescopes, despite the greater obscuration by dust in the lower latitudes.

The concentration of stars toward the Milky Way shows that the main body of the Galaxy is flattened in the direction of its poles and is widely extended toward its equator. When we look toward the equator, we are looking the long way out through the Galaxy and therefore at many more stars. Thus the band of the Milky Way.

The similarity of the numbers of stars in corresponding latitudes north and south of the galactic equator shows that the sun is not far from the plane of this equator.

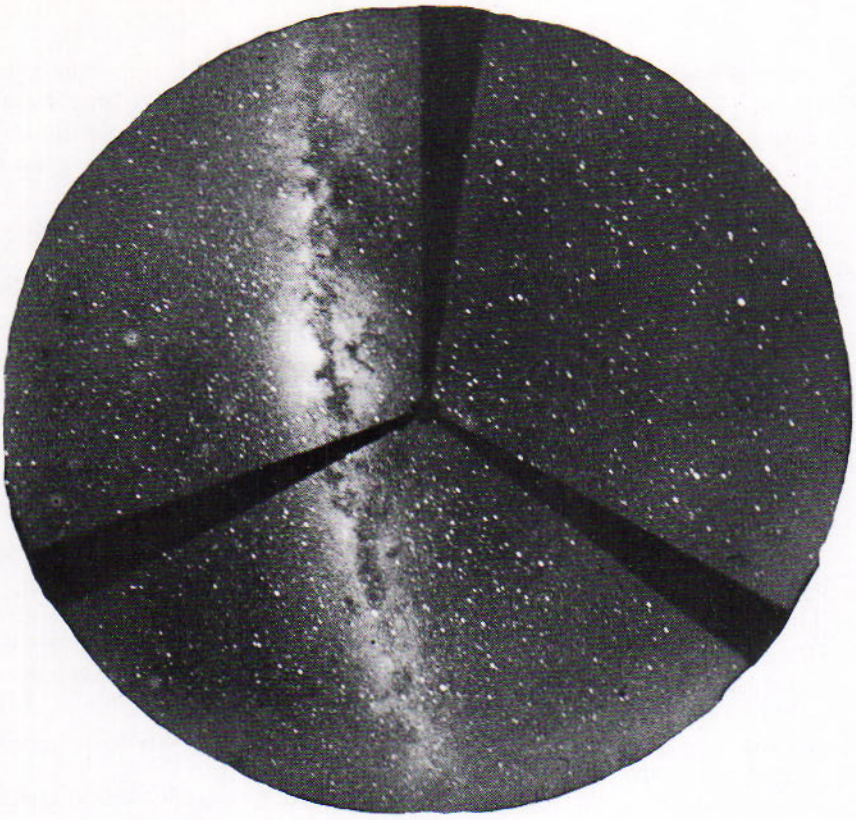


Figure 17.9
The southern Milky Way in infra-red light, photographed by T. Houck and A. Code with a Greenstein-Henyey wide-angle camera at Bloemfontein, South Africa. *(Washburn Observatory photograph)*

17.10 Eccentric Position of the Sun

Although the sun is near the principal plane of the Galaxy, it is about three fourths of the distance from the center toward the edge of the disk. A frequently cited value of the distance is 8,200 parsecs, as determined by Baade from his observations of RR Lyrae stars near the center and by the Leiden radio astronomers from their analysis of the galactic rotation. It seems possible, however, that this value is somewhat too small. A provisional distance of as much as 10,700 parsecs may be indicated by recent observations of B stars and cepheid variables by A. E. Whitford, A. D. Code, and J. D. Bahng. Astronomers have adopted a value of 10 kiloparsecs (32,000 light years) for computational convenience.

The place of the center was originally located by Shapley in galactic longitude 325° and latitude 0°, in Sagittarius; he supposed that this center has the same direction from us as the center of the system of

globular clusters, which he had determined. Shapley also remarked that from our place in the suburbs of the Galaxy the greatest brightness and complexity of the Milky Way is observed in this direction, and the dullest aspect in the opposite direction, in Auriga and Taurus, where we look the shortest way out through the Milky Way.

The more recent determinations have placed the galactic center remarkably close to the position originally assigned to it. The center, marked by the radio source Sagittarius A, is in galactic longitude 327°.8, latitude −1°.4 in the former system, or in right ascension $17^h42^m.4$, declination −28° 55′ (1950). Note in Map 3 the location of the center between the characteristic star-figures of Scorpius and Sagittarius.

17.11 The Spiral Structure

Two arms emerge from opposite sides of the central region of the Galaxy and coil around it in the same sense and in about the same plane. This spiral is considered to be of intermediate type in respect to the closeness of its coiling, resembling the great Andromeda spiral. The arms contain a considerable amount of gas and dust as well as stars.

These features of the galactic structure became definitely known only a few years ago. After Hubble, in 1924, had demonstrated the existence of other galaxies, it began to seem possible that our galaxy might have a spiral structure similar to that of some of the galaxies outside it. This possibility strengthened to a probability in the minds of many astronomers and finally, in 1951, became established fact.

The tracing of the spiral arms of the Galaxy in the heavens is in progress by at least three means: (1) by direct photography of emission nebulae and the hot stars associated with them (17.12), which are also prominent in the arms of exterior spirals; (2) by studies of interstellar lines in stellar spectra (17.16); (3) by radio reception from neutral hydrogen in the otherwise dark gas clouds of the arms (17.21) and also from ionized hydrogen in the bright clouds.

17.12 Spiral Arms Traced by Photography

The first tracing of the spiral arms of the Galaxy was announced in 1951 by W. W. Morgan, S. Sharpless, and D. E. Osterbrock at Yerkes Observatory. Employing a wide-angle Greenstein-Henyey camera and a filter for transmitting the light of the red line of hydrogen, they had photographed the part of the Milky Way that could be observed from that latitude. The positions and distances of emission nebulae and associated blue stars shown in the photographs permitted the tracing in space of two lengths of arms and the suggestion of a third. The three parts of the arms and the names originally assigned them are as follows.

The *Orion arm* was at first supposed to extend from Cygnus past Cepheus, Cassiopeia, Perseus, and Orion to Monoceros. It included the North America nebula, the Great Nebula in Orion, the Great Rift as

part of its dark inner lining, and it passed near the sun. Some later observers have considered this arm as passing through the sun's position and extending from here in a somewhat different direction. They have renamed it the Carina-Cygnus arm and have regarded as one of its spurs the part in the Orion region.

The *Perseus arm* was so named by Morgan because it contains the double cluster in Perseus. Outside the first arm, it passes about 2,100 parsecs from the sun. Emission nebulae are less conspicuous in this arm and probably would be difficult to trace in any arm outside it.

The *Sagittarius arm* is nearer the center than the sun's distance. Not well placed for observation in northern latitudes, its tracing was extended optically by B. J. Bok and associates at Bloemfontein. Hydrogen emission is very strong from Sagittarius through Scorpius and into Norma. From there to the Southern Cross it is weak, suggesting a break in the spiral structure. The arm goes on with some interruptions through Carina and Canis Major to Monoceros.

Optical tracing of the arms of the Galaxy is made difficult by the obscuring dust of the Milky Way. The tracing of the spiral pattern has since been extended by radio reception (17.21), which is not hampered by the intervening dust.

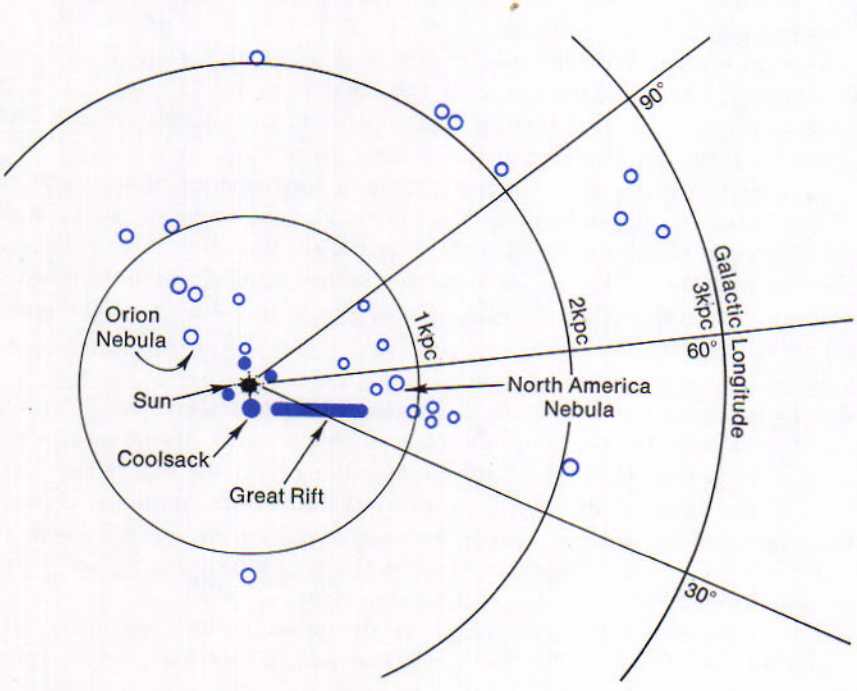


Figure 17.12
Parts of spiral arms of the galaxy. Traced by directions and distances of emission nebulae. *(Adapted from a model by W. W. Morgan, S. Sharpless, and D. E. Osterbrock, Yerkes Observatory)*

ROTATION OF THE GALAXY

The flattened form of the main body of the Galaxy indicates its rotation. Other effects of the rotation are found in the two star streams, in the trend of the motions of stars of high velocity and of the globular clusters, and in the systematic changes in the radial velocities of distant stars with changing galactic longitude. We note these effects in the order of their discoveries.

17.13 Two-Star Streams

Up to the beginning of the present century, no evidence of the systematic motions of the stars had been presented, aside from the apparent drifting of stars away from the standard apex of the solar motion and the common motions of stars in binary systems and clusters.

In 1904, the Dutch astronomer J. C. Kapteyn announced that the peculiar motions of stars (from which the effects of the solar motion are eliminated) around us are not random. There are two streams of stars moving in opposite directions in the plane of the Milky Way with a relative speeed of 40 km/sec. With something like the convergence of the Hyades cluster, the stars of the two streams are closing in toward two opposite points in the heavens.

The convergent point of one stream is in right ascension $6^h\,15^m$ and declination $+12°$, in Orion, and of the other is in $18^h\,15^m$ and $-12°$, in Scutum. The line joining them is in the plane of the galactic equator and is not far from the direction of the galactic center.

This preferential motion of the stars is a consequence of the rotation of the Galaxy, as B. Lindblad, at Stockholm, was the first to explain. Most stars move in slightly eccentric orbits around the galactic center. Lindblad showed that these stars have a greater spread in their motions toward and away from the center than at right angles to this direction. The effect for us is the star streaming.

17.14 The Motions of High-Velocity Objects

The majority of the stars in the sun's vicinity have space velocities of the order of 20 km/sec, which are directed in general away from the standard apex in Hercules. These stars are moving along with the sun in the rotation of the Galaxy, and all have moderate individual motions as well. Exceptional stars, having speeds exceeding 60 km/sec, are known as *high-velocity stars;* their motions were first studied extensively by G. Strömberg at Mount Wilson and J. H. Oort at Leiden.

The motions of high-velocity stars in the galactic plane, when they are corrected for effects of the sun's motion toward Hercules, are directed away from the half of the Milky Way having Cygnus at its middle. Dif-

ferent classes of these objects have different average speeds. For example, the RR Lyrae variables are moving at the rate of 100 km/sec, and the globular clusters, the swiftest of all, are going twice as fast relative to the sun.

Previous to these studies the motions of the globular clusters seemed surprisingly rapid. Now it is understood that we ourselves are the ones who are moving so swiftly. The sun is speeding toward Cygnus in the whirl of the highly flattened disk of the Galaxy at the rate of 216 km/sec, the value commonly adopted. The less flattened array of RR Lyrae stars is rotating with the lower speed of 116 km/sec and is therefore falling behind us at the rate of 100 km/sec. The more nearly globular assemblage of the globular clusters (Fig. 15.13B) is turning even more slowly. The idea of subsystems of the Galaxy, rotating on a common axis at different rates and thus having different degrees of flattening, was proposed by Lindblad.

If we assume that the sun is on a circular orbit about the center of the Galaxy, then high-velocity objects cannot be. This tells us that the orbits of the high-velocity stars must be highly elliptical. Computation shows that some of these stars must pass through the dense central region of the Galaxy. These stars are population II stars and are metal poor when compared to the sun.

17.15 Differential Effects of the Rotation

Two extremes in the distribution of material through the Galaxy would produce the following effects in the rotation:

1. If the material is uniformly distributed, the Galaxy would rotate like a solid wheel. All parts would rotate in the same period, keeping the same relative positions. Evidence of such rotation would be difficult to observe except by reference to the external galaxies.
2. If most of the material is concentrated around the center, the rotations of the outer parts would resemble the revolutions of the planets around the sun. The periods would increase and the speeds would diminish with greater distance from the center. Thus the stars nearer the center than the sun's position would go around faster than the sun's speed, so as to overtake and pass on ahead of us. The stars farther from the center than the sun's position would move more slowly and would therefore fall steadily behind us in the rotation.

J. H. Oort, in 1927, was the first to show more nearly the second effect of the galactic rotation in the radial velocities of stars in different parts of the Milky Way. He observed that stars having galactic longitudes 45° and 225° greater than the direction of the center (Fig. 17.15) are receding from us with the greatest speeds. Stars having longitudes 135° and 315° greater than the direction of the center are approaching us with the greatest speeds.

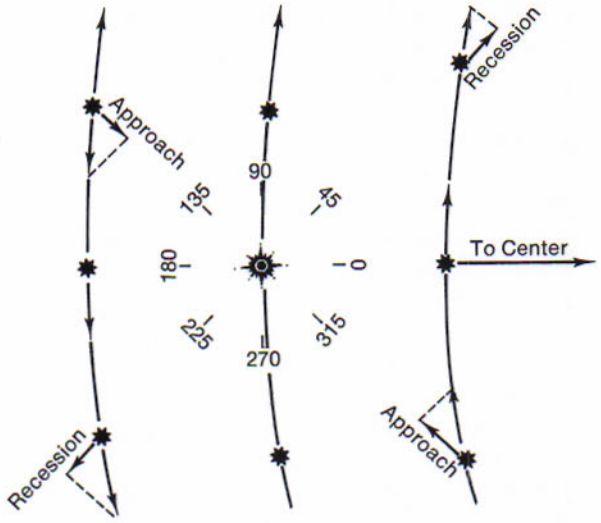


Figure 17.15
Effect of the rotation of the Galaxy on the radial velocities of stars. Stars nearer the center than the sun's distance are going around faster and are passing by the sun. Stars farther from the center are moving more slowly and are falling behind the sun. Thus stars having directions 45° and 225° greater than that of the center are receding from the sun, and stars around 135° and 315° are approaching the sun.

The differential effect, V, in the radial velocities of stars is expressed by the formula:

$$V = rA \sin 2l,$$

where r is the difference between the star's and the sun's distance from the galactic center. A is Oort's constant, 15.0 km/sec per 1000 parsecs, which is the rate of velocity change with increasing distance difference. The quantity l is the star's galactic longitude measured from the direction of the galactic center. The result for a particular distance difference, where the radial velocities are plotted against the longitudes, is a curve with a double wave, having two maxima and minima around the circle of the Milky Way.

The differential effect in the observed radial velocities, whether of approach or recession, increases with the difference of distance between the star and the sun from the center of the Galaxy, and its amount can inform us of this difference. This is the basis for tracing spiral arms of the Galaxy by radio reception (17.21).

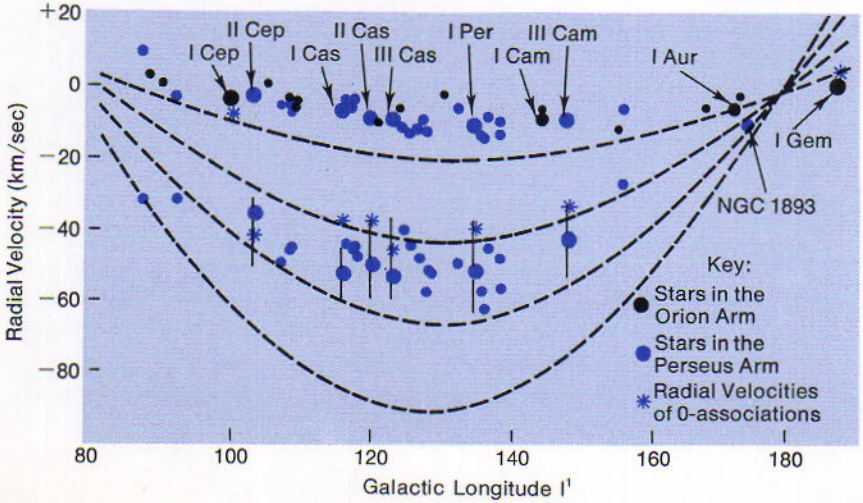
17.16 Interstellar Gas in the Spiral Arms

A spectroscopic survey of a large section of the Milky Way by G. Münch with the 200-inch Palomar telescope has shown that the interstellar lines in stellar spectra (16.9) are caused by gas in the spiral arms of the Galaxy. These studies have located parts of two arms at different distances beyond the sun's distance from the center of the Galaxy. These are the Carina-Cygnus arm and the Perseus arm.

Because the gas in the more distant of these two arms is moving the slower relative to the sun's motion in the rotation of the Galaxy, the dark lines it absorbs in the spectra of remote stars have the greater Doppler displacements. Thus the interstellar lines are divided into two components. The corresponding radial velocities for the two sets of lines are plotted in Fig. 17.16 with respect to the galactic longitudes of the stars in the former reckoning. The two arrays of points are well represented by sine curves showing the differential effects of the galactic rotation at distances of 400 and 3000 parsecs, which are evidently the distances of the two arms from the sun in these directions. The distance of the nearer arm is in practical agreement with its distance derived in other ways.

The maximum values of approach to the sun occur in about galactic longitude 100° in the old coordinates (about 135° from the direction of the galactic center), as Fig. 17.15 shows that they should. The departures of the points from the sine curves of Fig. 17.16 and the frequent further splitting of the interstellar lines into several parts are ascribed to turbulent velocities of the gas in the two arms; these are more conspicuous in the more distant arm.

Figure 17.16
The dependence on galactic longitude of the radial velocities of the strong components of interstellar in the Northern Milky Way. Larger circles represent 0-associations. The broken-line curves represent the galactic rotation effect on the radial velocities of points at distances of 1, 2, 3 and 4 kpc from the galactic center. *(Diagram by G. Munch, in* Galactic Structure, *University of Chicago Press, 1965).*



17.17 Rotation of the Galaxy

Viewed from the north galactic pole, the direction of the galactic rotation is clockwise. As we in the northern hemisphere look from our place in the Galaxy toward its center, the direction of the rotation between us and the center of the Galaxy is toward the left. The period of the rotation is probably about the same throughout the central region. In the disk it becomes greater with increasing distance from the center. At the sun's distance the period is of the order of 250 million years. If the Galaxy is 10 billion years old, it has rotated 40 times in our vicinity since its beginning. The difference in the rotation at different distances from the center is enough to efface the spiral arms in the course of a single rotation, according to Oort. The question remains as to how the dispersed spiral pattern is restored (18.3).

The sun's velocity in the rotation is generally taken to be 216 km/sec, or 134 miles a second, toward galactic longitude 55° in the former reckoning, in Cygnus. This is the value given by van de Hulst, Muller, and Oort in 1954 from their radio records. It is based on the sun's distance of 8200 parsecs from the galactic center; and some other investigators have reported values of this order. From the radial velocities of galaxies of the local system, however, N. U. Mayall, in 1946, and M. L. Humason, in 1955, derived the sun's velocity as about 296 km/sec. H. F. Weaver points out that the wide discrepancy in these results could be lessened by adopting a greater distance of the sun from the center.

RADIO VIEW OF THE GALAXY

Radio astronomy has provided in recent years an unexpected and completely new approach in the studies of the Galaxy. Improved apparatus and methods are giving increasingly detailed views of the radio Galaxy. The use of the 21-cm emission line in the spectra of the otherwise dark hydrogen clouds permits the tracing of the spiral arms.

17.18 An Early Radio Survey

After Jansky, in 1932, had reported radio reception from the Milky Way at 14.7 meters, G. Reber completed, in 1944, the first extensive radio survey of the Milky Way. With a 30-foot paraboloidal antenna and apparatus for recording at 1.85 meters, he scanned the Milky Way as it passed his meridian. Reber's pioneer equipment could not distinguish separately

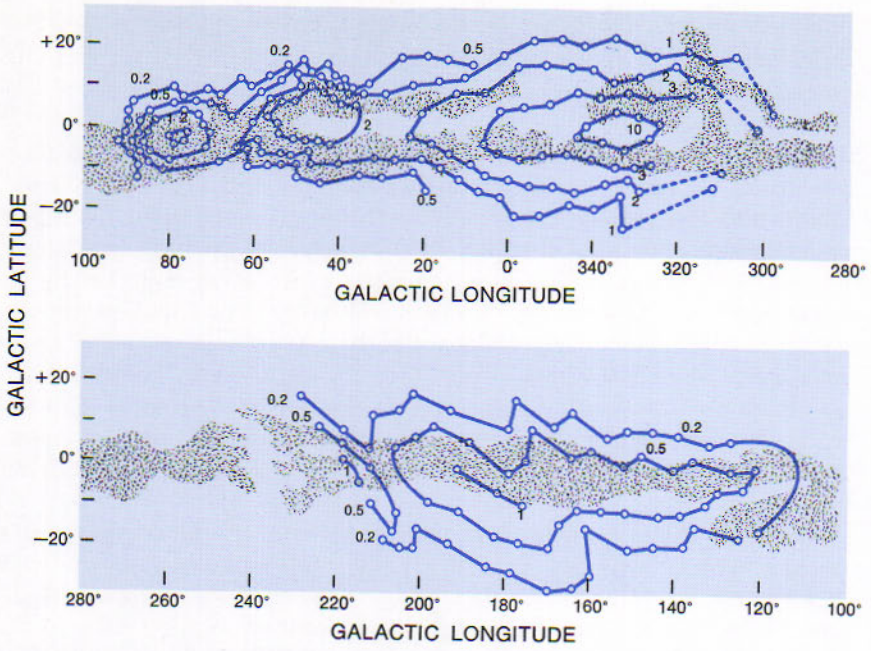


Figure 17.18
Contour lines of equally intense radiation from the Milky Way at 1.85 meters. The galactic coordinates are in the former system. *(Diagram from Reber's radio data prepared by R. E. Williamson and R. J. Northcott, David Dunlap Observatory)*

sources of radiation closer together than 12°. Although the view was accordingly blurred, it revealed some of the brightest features of the radio Milky Way. His antenna is now remounted at the National Radio Astronomy Observatory at Green Bank, West Virginia.

Reber's original results are shown in the contour map of Fig. 17.18, where the intensities of the radiation are indicated with respect to the galactic longitudes of their origins. We note that regions of equal intensity are roughly symmetrical relative to the galactic equator. The greatest intensity, 10 on Reber's scale, was recorded from the direction of the galactic center, about longitude 330° by the former reckoning. Secondary maxima of intensity 2 were found in Cygnus, longitude 50°, and in Cassiopeia, 80°; these are now known to be effects of two strong radio sources in the former constellation and one in the latter.

Later surveys with improved apparatus have been made at a variety of wavelengths. The Milky Way is the dominant feature of all the radio maps, but radio sources appear instead of individual stars of the optical view. The radio records differ, depending on the wavelengths with which they are obtained. At the shorter wavelengths, where the thermal radiation (4.27) is most intense, the Milky Way is narrower and the sources

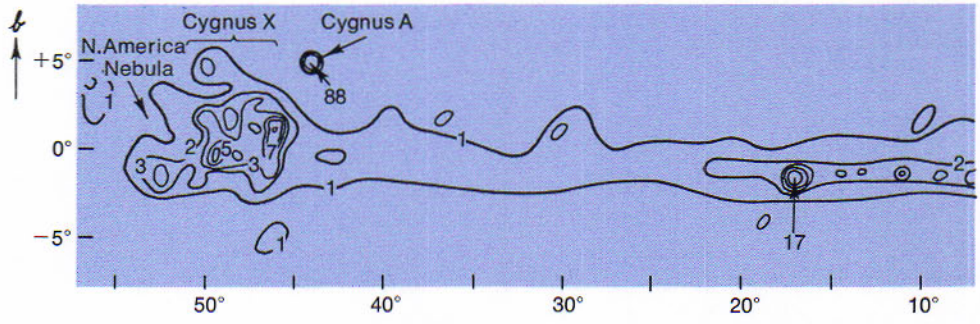
are mainly emission nebulae. At the longer wavelengths, where the non-thermal radiation is most intense, the Milky Way is broader and the sources are of less usual types.

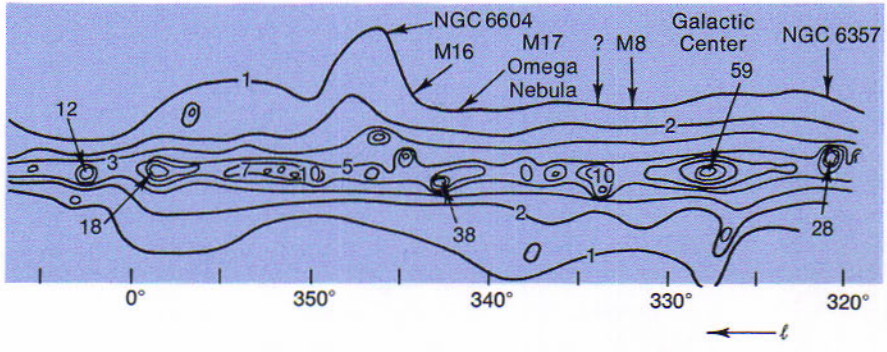
17.19 Radio Maps of the Galaxy

Numerous radio surveys of the Galaxy have been made since Reber's early work. Careful selection of the wavelength used enhances desired features and suppresses other features. The preponderance of hydrogen makes it natural to make a map of the Galaxy at a wavelength of 21 cm. Figure 17.19A shows the result of a survey of λ 21 cm radiation by G. Westerhout from Cygnus past the galactic center. The numbers on the contour lines are relative intensities above the natural noise of the receiver. The plane of the Galaxy is clearly defined and certain very high intensity areas stand out; for example, Cygnus A. In addition to the sources marked, the galactic center appears to be a concentrated source as well. The region between 45° and 55° in the former reckoning is where we are looking along a spiral arm.

In figure 17.19B we have reproduced a portion of the same region from a survey made at λ 11 cm. This survey, by W. Altenhoff, was done with

Figure 17.19A
λ 21-cm survey of the Galaxy between cygnus and the galactic center. *(G. Westerhout)*

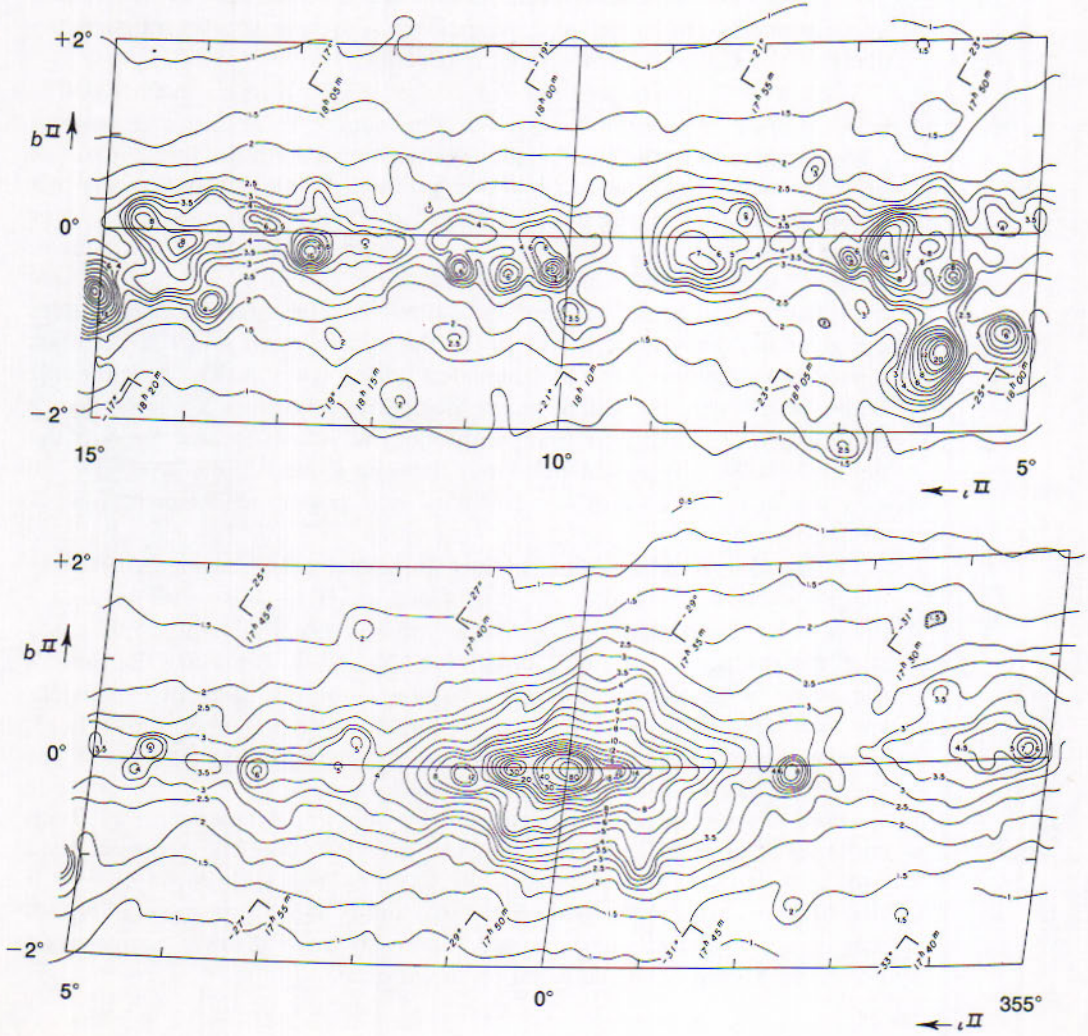




a higher resolution than that in figure 17.19A so allowance should be made for this fact. The sources that we see in figure 17.19A are discernible here also and new sources appear. The center of the Galaxy begins to show greater detail. It is now apparent from extensive work by radio astronomers that the center of the Galaxy contains numerous highly concentrated sources. One of these sources is probably similar to the intense point source that we see in the center of the Andromeda Galaxy (Figure 18.4A).

Figure 17.19B

λ 11-cm survey of the Galaxy. Only a small section of the extensive survey conducted on this wavelength by W. Altenhoff is reproduced here. Note the higher resolution and the new galactic coordinates, as opposed to the λ 21-cm survey. 0° in this survey corresponds to 327.7° in Fig. 17.19A. *(Courtesy of W. Altenhoff and NRAO)*



Radio maps of the Galaxy bear little resemblance to what we see with the eye. It is true that we see the Milky Way and the concentration in Cygnus and the galactic center, but none of the bright stars are discernible. The explanation, as we have seen, is that the radio waves pass through the gas and dust and that the stars are very low radio emitters. A complete λ 21-cm map shows that neutral hydrogen covers the entire sky (17.24).

17.20 Discrete Radio Sources

Radio maps of our galaxy, similar to those just discussed, reveal certain discrete sources of radio emission which range in size from more than 1° to less than 1″ in diameter. Most of the sources in or near the Milky Way plane are associated with the Galaxy and the majority of these appear to be remnants of novae and supernovae. Discrete sources outside the plane of the Galaxy are mostly extragalactic.

$S \propto \lambda^{\alpha}$

α = *spectral index*

The discrete galactic sources are primarily nonthermal, their radiation being due to the *synchrotron process.* The source is observed at several wavelengths and a plot of intensity against wavelength (or frequency) has been made, which is analogous to a magnitude-wavelength plot. If the plot slopes down with increasing wavelength the source is thermal; if it does not it is nonthermal or a thermal source with a nonthermal component.

As we might expect, the Crab nebula turns out to be a bright discrete radio source. It has the designation Taurus A. The most familiar sources are generally known by the name of the constellation and the brightest source in the constellation is designated by the letter A. The Crab nebula is the brightest radio source in the northern hemisphere's winter sky. It is the remnant of the type II supernova of the year 1054 and contains the pulsar NP0532. We might point out that the Crab nebula, accessible on long winter nights in northern latitudes, is a most excellent astrophysical laboratory.

The brightest radio source at low frequencies is Cassiopeia A. It is unique and, as R. Minkowski points out, must be the result of a rare event. It has been identified optically with a wispy, fragmented, nebulous shell expanding at a rate of 7500km/sec Minkowski computes the time of the event to be around the year 1700 by assuming uniform expansion. He associates Cassiopeia A with a supernova of the very rarest kind. Cassiopeia A shows a secular decrease in its energy output of 1 to 2% per year. This had been predicted by Shklovsky earlier.

Other discrete galactic sources are Tycho's and Kepler's novae (both supernovae of type I), the supernova of the year 1006 (the brightest nova ever recorded and also of type I), the Cygnus Loop (supernova of type II that occurred around the year 65,000 BC) and IC443, a supernova of type II that occurred around the year 400. There are several other sources suspected of having been remnants of supernovae.

Figure 17.20
NGC 5128, an unusual galaxy in Centaurus and the bright radio source Centaurus A. *(Photograph from the Hale Observatories)*

Sources similar to those just discussed occur in the direction of the Magellanic Clouds and they are clearly the result of novae and supernovae in those objects. Other bright sources such as Centaurus A, Cygnus A and Virgo A are associated with extra-galactic objects. The bright source in the center of the galaxy is designated Sagittarius A.

Scans of the Galaxy at certain frequencies (wavelengths) such as the microwave transitions in molecules reveal other discrete sources. Scans at λ 17.5 cm, an OH transition, reveal small sources of radiation in the center of the Galaxy. Many of the sources have very small angular sizes when measured with an interferometer and are located in the direction of small clouds of neutral hydrogen catalogued by G. Westerhout. We are tempted to associate such coincidences with star formation.

17.21 Tracing of Spiral Arms by Radio

Surveys of the spiral structure of the Galaxy with radio telescopes began soon after the detection of the emission line at λ 21 cm from the otherwise dark clouds of neutral hydrogen that are abundant in the arms. The original survey by H. C. van de Hulst, C. A. Muller, and J. H. Oort with the 25-foot radio telescope at Kootwijk, Holland, reported in 1954, shows features of the spiral pattern between longitudes 12° to 167° and 190° to 250° beyond the sun's distance from the center of the Galaxy.

As an example of how the tracing is done, consider a particular pointing of the radio telescope toward the central line of the Milky Way in Cassiopeia, in longitude 80° in the former system, or 112° from the galactic center. The hydrogen clouds in this direction are relatively approaching the sun. As the distances of the clouds from us are greater, their speed of approach, and accordingly the Doppler shift of the spectrum line to shorter wavelengths, increases.

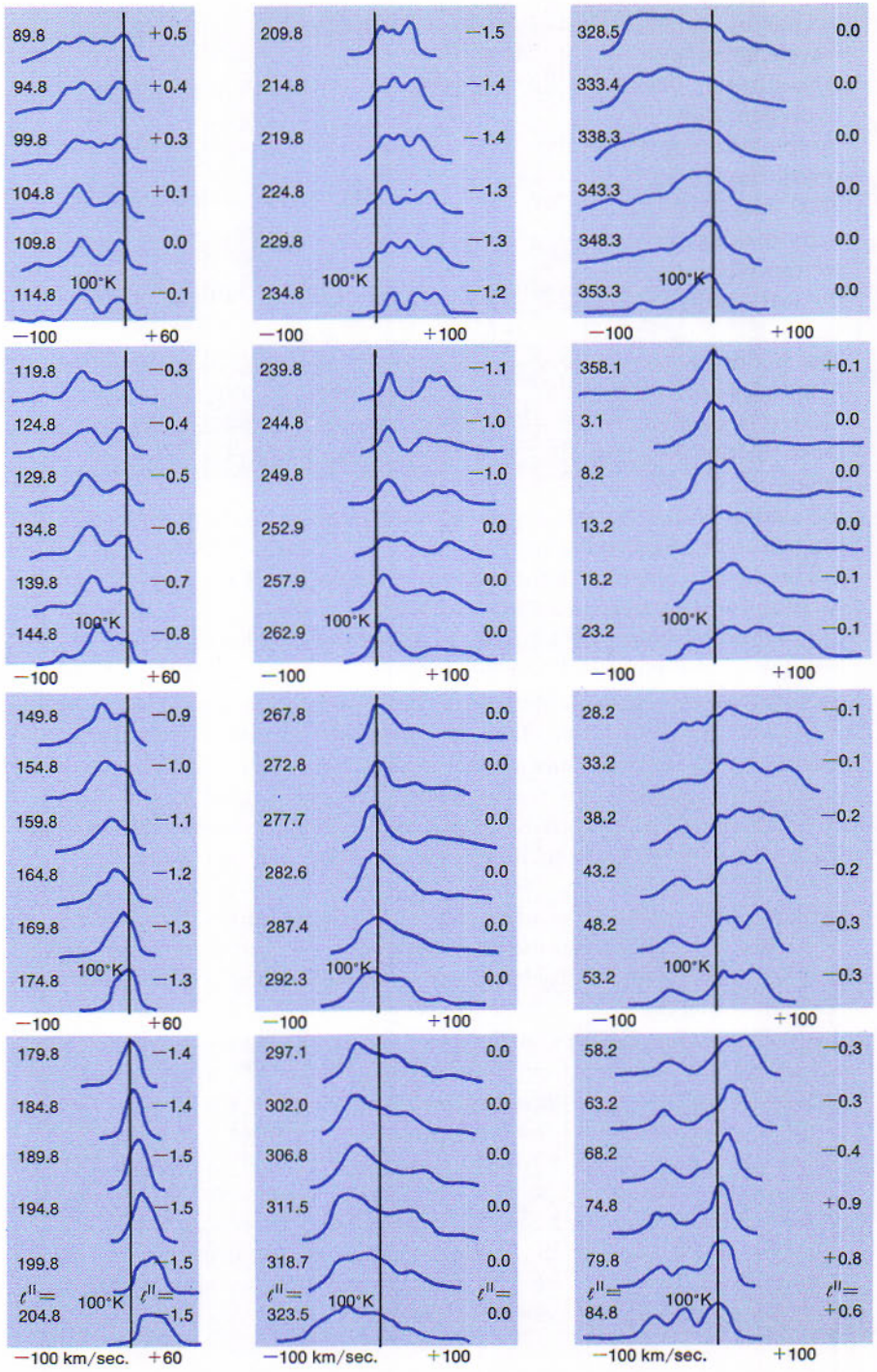


Figure 17.21
λ 21-cm line profiles around the galactic plane. Note smooth shift in concentration in velocity as function of direction. (G. Westerhout, *Galactic Structure*).

Thus by tuning the radio telescope first to λ 21 cm and then to shorter wavelengths successively, the survey reaches to greater and greater distances in this direction. Where the signal becomes stronger, there are the hydrogen lanes of a spiral arm. The line profile at $l^{II} = 85°$ recorded by the telescope shows three maxima. The corresponding radial velocities reveal the distances from the sun, which in this case are given as 500, 3200, and 7500 parsecs.

By making observations all around the plane of the Galaxy (Fig. 17.21) we can watch the intensity maxima change in velocity and hence plot out the arms of hydrogen.

17.22 Spiral Structure of the Galaxy

The hydrogen lanes traced by Dutch radio observers in recent years in longitudes available to them are shown in Fig. 17.22A. The center of the Galaxy is marked by a cross and the sun's position by a dot above the cross. In the direction 80° from the sun we note three hydrogen lanes, which trace three spiral arms. The first is the Carina-Cygnus arm and the second is the Perseus arm. The lane tangent to the 18° direction from the sun is the Sagittarius arm.

The view in this figure is from the north galactic pole, from which the direction of the galactic rotation is clockwise. The distances of the Carina-Cygnus and Sagittarius arms from the center increase with increasing longitude, showing that the arms are trailing in the rotation. The narrow gap in the lanes just above the sun's position represents directions near the anticenter where the radial velocities are so nearly zero that their differences are not significant.

Figure 17.22A
Spiral structure of the Galaxy as deduced from NFRA observations in Leiden and CSIRO observations in Sydney. The position of the sun is marked by arrow. *(Courtesy of G. Westerhout)*

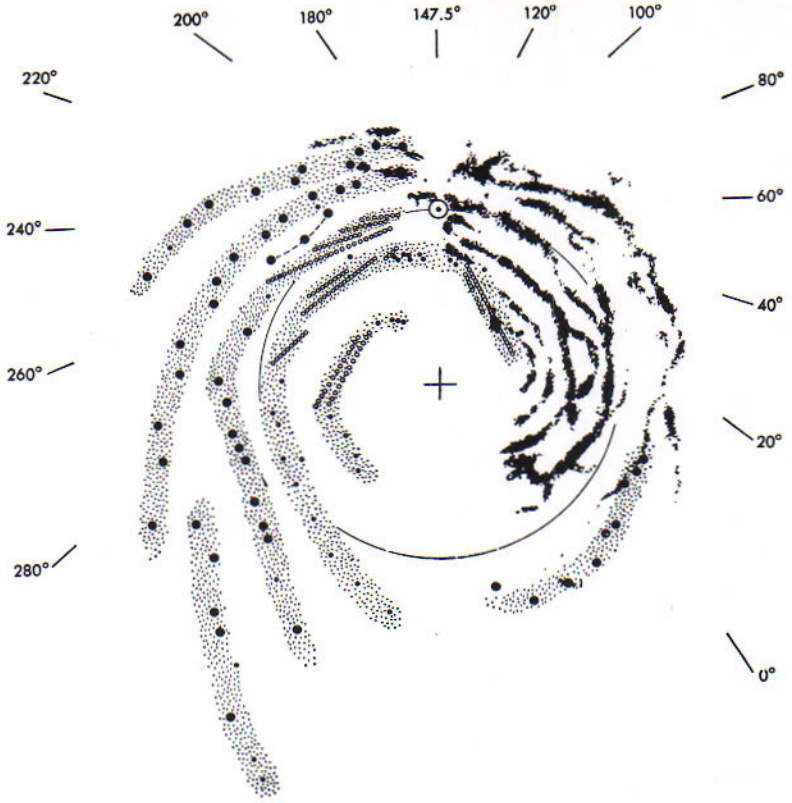


Figure 17.22B
Spiral Structure of the Galaxy. Composite diagram of the part of the radio pattern traced at Leiden (right) and at Sydney (left). The numbers are galactic longitudes in the former system. *(Diagram by J. L. Pawsey, Radiophysics Laboratory, Sydney, Australia)*

The composite diagram of Fig. 17.22B shows the spiral arms traced by the radio observers in Holland (slightly retouched) and at the Radiophysics Laboratory in Australia. In the Sydney pattern the larger dots represent the more reliable observed positions.

Radio studies of the galactic structure other than with the λ 21-cm spectrum line can employ the continuous nonthermal radiation (4.27) from the spiral arms. Having separated this type of radiation from the rest, B. Y. Mills at the Radiophysics Laboratory in Sydney has constructed a model of the spiral arms that resembles in a general way the optical and other radio models.

17.23 Gas Around the Galactic Center

A significant addition to the radio picture of the Galaxy is recorded by J. H. Oort and associates with the 82-foot paraboloid at Dwingeloo. They find that the radio source Sagittarius A marking the galactic center is a region of neutral hydrogen 2° in diameter enclosing a few clouds of ionized hydrogen. Gas streams are moving outward from here with speeds of from 50 to 200 km/sec and are also turning in the general rotation of the Galaxy. A bright ring of gas around the center with a radius of 3,700 parsecs glows by the synchrotron process. How this gas originates and is being replenished as it streams outward is unknown, unless it keeps falling in from the rare gaseous medium of the galactic halo. The opinion is being expressed that we may have here a clue to the evolutionary process by which a galaxy develops spiral arms.

17.24 Gas Distribution Out of the Galactic Plane

Careful observations of neutral hydrogen out of the plane of the galaxy reveal the thickness of the gas to be roughly 500 parsecs. The density of the gas is highest in the plane and then falls off as one goes above and below the plane. Fully half of the gas lies within a disk only 220 parsecs thick. This extremely flat disk extends outward from the center of the Galaxy to a distance of about 7 kiloparsecs. It then bends upward on one side and downward on the opposite side in what has become known as the "hat brim" effect. A similar effect can be seen in other galaxies; for example, in M31.

There have been several efforts to explain the turned edge of the gas disk, and we simply cite the two most quoted ones. One explanation would be a tidal interaction with the Magellanic Clouds. A second explanation due to F. Kahn and L. Woltjer attributes the distortion to pressure on the gas disk caused by the motion of the Galaxy through the "intergalactic gas". A third would be interaction with a nearby unseen galaxy.

Out of the plane of the galaxy knots of in-falling hydrogen gas have been observed. These are at high latitudes and some of the velocities are quite high, on the order of 150 km/sec. These high and intermediate velocity clouds were first observed by the Dutch astronomers and are interpreted by J. Oort as knots of intergalactic hydrogen falling into the Galaxy. It has been suggested also that these clouds are relatively near to the sun and are the result of hydrogen being ejected from the disk by a violent event.

The high velocity clouds are superposed upon a smooth gradual inflow of hydrogen. The inflow is at a velocity of 6 km/sec and is occurring on both sides of the disk.

17.25 The Origin of the Galaxy

A hypothesis by C. F. von Weizsäcker of the evolution of the Galaxy offers a convenient means of reviewing the present information about the system of the Milky Way. Von Weizsäcker's tentative account begins several billion years ago with a large turbulent cosmic cloud, mainly of hydrogen gas, in rotation around its center. The original cloud condensed until it fractured into smaller clouds. Most of the clouds eventually fell in toward the equatorial plane of the rotating mass, forming the flat disk of gas in which spiral arms would develop and also condensing into the stars of the Milky Way.

Some of the smaller clouds remained in the nearly spherical halo of the Galaxy to evolve into globular clusters of stars. Statistical studies by von Weizsäcker and others based on Mayall's radial velocity measures showed that the globular clusters are revolving in highly eccentric orbits around the galactic center, somewhat remindful of the orbits of comets around the sun. By Kepler's law of equal areas, the clusters spend most of their lives near their present positions far from the center, where they are practically immune to disturbing influences. After long intervals each revolving

cluster dips for relatively short times into the crowded central region of the Galaxy. By collisions here the interstellar gas of the cluster is heated and thereby dissipated, but the formation of the cluster stars themselves is not likely to be seriously affected.

Table 17.I, which lists the contribution of various types of objects in solar masses per thousand cubic parsecs, is very revealing. Those objects that are most obvious to the naked eye contribute very little to the mass of the Galaxy at least in the solar neighborhood. Most surprising is the large amount of unidentified matter. Some astronomers feel that most of this matter is unobservable gas and dust. Others suggest that the mass comes from degenerate and "burned out" stars and perhaps numerous small planet-like objects ejected from multiple systems.

TABLE 17.I
Composition of the Solar Neighborhood in Solar Masses* (per 10^3 pc^3)

Interstellar Hydrogen	25
O-B5 Stars	11×10^{-2}
Cepheids	3×10^{-4}
Open Clusters (O-B6)	3×10^{-2}
Open Clusters (B7-F)	5×10^{-2}
B8 - A5 Stars	1.7
F Stars	2.5
Dwarf G Stars	3.5
Dwarf K Stars	9.0
Dwarf M Stars	29
Giant G - M Stars	0.7
White Dwarfs	8.0
Sub-dwarfs	0.5
Planetary Nebulae	—
Long Period Variable Stars	1×10^{-3}
RR Lyrae Stars	3×10^{-5}
RV Tauri Stars	1×10^{-5}
Globular Clusters	1×10^{-3}
Unidentified Matter	64

* From notes taken at a seminar given by A. Blaauw

Finally, the hypothesis must explain the content of the Galaxy in some detail. A theory of stellar evolution coupled with the hypothesis must yield the relative composition and distribution of matter. Except for the gas distribution, our knowledge of this distribution and composition is most complete in the solar neighborhood out to an approximate distance of 2000 parsecs.

REVIEW QUESTIONS

1. What constellation is in the direction of the center of the Galaxy?
2. What is Gould's Belt?

3. The dark regions in the Milky Way such as the Coalsack were originally thought to be holes in the universe. Why?
4. How did H. Shapley determine the direction and distance to the center of the Galaxy?
5. Describe the five stellar populations as presented here.
6. Describe briefly:
   a) high velocity stars
   b) star counts
   c) star streams
   d) high velocity clouds
7. Explain how we deduce that the outer parts of the Galaxy are rotating slower than the inner parts.
8. How thick is the galactic gas disk? What are the possible explanations of its "hat brim" appearance?
9. How do we trace the spiral arms optically? by radio? Why is the radio technique the more powerful?
10. Describe the galactic center as seen by radio means.

REFERENCES

Bok, Bart J. and P. F. Bok, *The Milky Way*, 4th edition. Cambridge, Mass.: Harvard Univ. Press, 1957.

Davies, Rodney D. and H. P. Palmer, *Radio Studies of the Universe*, Princeton, N.J.: D. Van Nostrand Co., 1959.

Page, Thorton & L. W. Page, eds., *Stars and Clouds of the Milky Way*, New York: The Macmillan Co., 1968.

van de Kamp, Peter, *Basic Astronomy*, New York: The Macmillan Co., 1952.

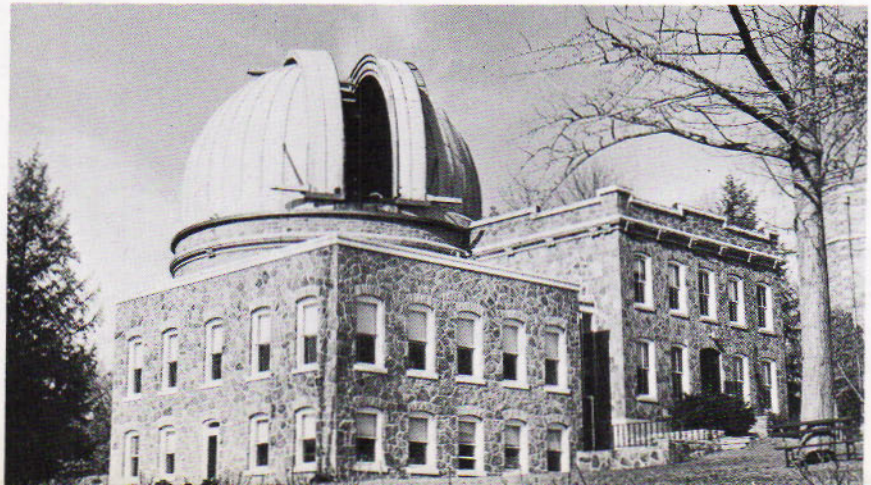
FOR FURTHER STUDY

Blaauw, Adriaan and M. Schmidt, eds., *Galactic Structure*, Chicago: University of Chicago Press, 1965.

Kraus, John D., *Radio Astronomy*, New York: McGraw-Hill Co., 1966.

Mihalas, Dimitry and P. Routly, *Galactic Astronomy*, San Francisco: Freeman Cooper & Co., 1968.

Woltjer, Lodewijk, ed., *Galaxies and the Universe*. New York: Columbia Univ. Press, 1968.



Sproul Observatory, Swarthmore College. *(Sproul Observatory photograph)*


18

THE GALAXIES

STRUCTURES OF GALAXIES — PARTICULAR FEATURES OF GALAXIES — DISTRIBUTION OF GALAXIES — SPECTRA OF GALAXIES

Other galaxies outside our own are scattered through space as far as the largest telescopes can explore. These major building blocks of the physical universe are structurally of three main types: elliptical, spiral, and irregular galaxies. They are generally assembled in large clusters and in smaller groups such as the local group to which our galaxy belongs. Their obscuration by the dust clouds of the Milky Way led to the original designation of galaxies as "extra-galactic nebulae." The redward displacements of their spectral lines, which increase as the galaxies are more distant from us, provide the principal observational basis for the studies of the expanding universe.

) Doradus nebula. *(Courtesy of B. Bok)*

STRUCTURES OF GALAXIES

There are three general types of galaxies as exhibited by their structure. In addition there is a great variety of structure among galaxies that do not fit neatly into the three major classes. Compact galaxies, peculiar galaxies *and* double galaxies *for example, exhibit unique structures.*

18.1 Three General Types

Galaxies are divided into three broad types.

1. *Elliptical galaxies* are ellipsoidal masses of stars, which are densest around the centers and thin out toward the edges. Their outlines range in ellipticity from nearly circular to lenticular.
2. *Spiral galaxies* have lens-shaped central regions surrounded by flat disks, in which spiral arms are embedded; the arms contain gas and dust as well as stars. Spiral galaxies are of two kinds: normal spirals and barred spirals.
3. *Irregular galaxies* have no particular forms, except that some of them appear to be flattened.

The following more detailed accounts of the different kinds of galaxies are based on E. Hubble's structural classification. A modification proposed by W. W. Morgan, which also considers the central concentrations and related stellar populations of the galaxies, is described later (18.20).

18.2 Elliptical Galaxies

These galaxies are so named because the nearer ones appear through the telescope as elliptical disks. They are designated by the letter E followed by a number, which is 10 times the ellipticity (2.3) of the disk. The series runs from the circular class E0 to the most flattened E7, which looks like a convex lens viewed edgewise. If all the ellipticals were of the most flattened class, their disks would show the observed range of ellipticity, depending on how they were presented to us. The frequencies of the different classes, however, are not consistent with this supposition. Some elliptical galaxies are actually nearly spherical, although it is impossible to decide for a particular individual whether it is more flattened than it seems to be.

Extreme types of elliptical galaxies, E7, are shown in Figs 18.2A and B. M 32, one of the two companions of the Andromeda spiral (Fig. 18.4A), is an example of the slightly flattened class E2, and the other companion, NGC 205, is of class E5.

Figure 18.2A
Galaxy NGC 3115. *(Photograph from the Hale Observatories)*

Figure 18.2B
Elliptical galaxy NGC 147, which is resolved into stars. *(Photograph from the Hale Observatories)*

These are systems of stars, generally dust-free and of Baade's type II population. Almost all the gas and dust available for starbuilding would seem to have been exhausted. Small dust clouds remain in some systems and have young stars in their vicinities. The nearer ellipticals, including NGC 147 (Fig. 18.2B), are resolved into stars in the Mount Wilson and Palomar photographs.

The richer systems have their stars highly concentrated toward their centers. The Sculptor system is an example of the less populous objects that have more moderate concentration. This and the Fornax system were the first of the kind to be discovered. Other dwarf ellipticals were found later.

18.3 Normal Spiral Galaxies

These galaxies have lens-shaped central regions, from opposite sides of which two arms emerge and at once begin to coil around the centers in the same sense and the same plane. The central regions are brighter than the arms, and in many cases are directly visible with the telescope. The fainter and bluer arms can be seen to advantage only in the photographs.

The origin or cause of the distinctive spiral arms in the highly flattened galaxies such as our own is an interesting problem. The spirals always have two arms originating on exactly opposite sides of the central region. These arms are composed of gas and dust and contain the bright O and B stars. The assumption that the arms originate with the galaxy presents a problem because they should then be wound up some 30 to 50 times, whereas it appears that this phenomenon takes place only three or four times. C. C. Lin has presented an interesting hypothesis to explain the spiral arms.

Lin proposes that a density wave exists in the disk and that it has a quasi-stationary spiral structure. The gravitational field is associated with the density wave and indeed sustains it over a long period of time. The pattern is not to be considered permanent, however. The O & B stars originate in the high density region but with velocities differing from that of the density wave. Given sufficient time, about half a galactic orbit, or 100 million years, they would move out of the spiral pattern. Since these stars have lifetimes of only 10 million years they do not drift far from the pattern before they evolve and new O & B stars have already formed in the high density region thus continuing to define the spiral pattern. The hypothesis leads to conclusions that are observed in galaxies, especially the fact that only two arms will be generated.

Normal spirals are divided into three classes. *Class Sa* spirals have large central regions and thin, closely coiled arms; an example is NGC 4594 (Fig. 18.5). In *class Sb* the central regions are smaller, and the arms are larger and wider open. The great spiral in Andromeda is typical of

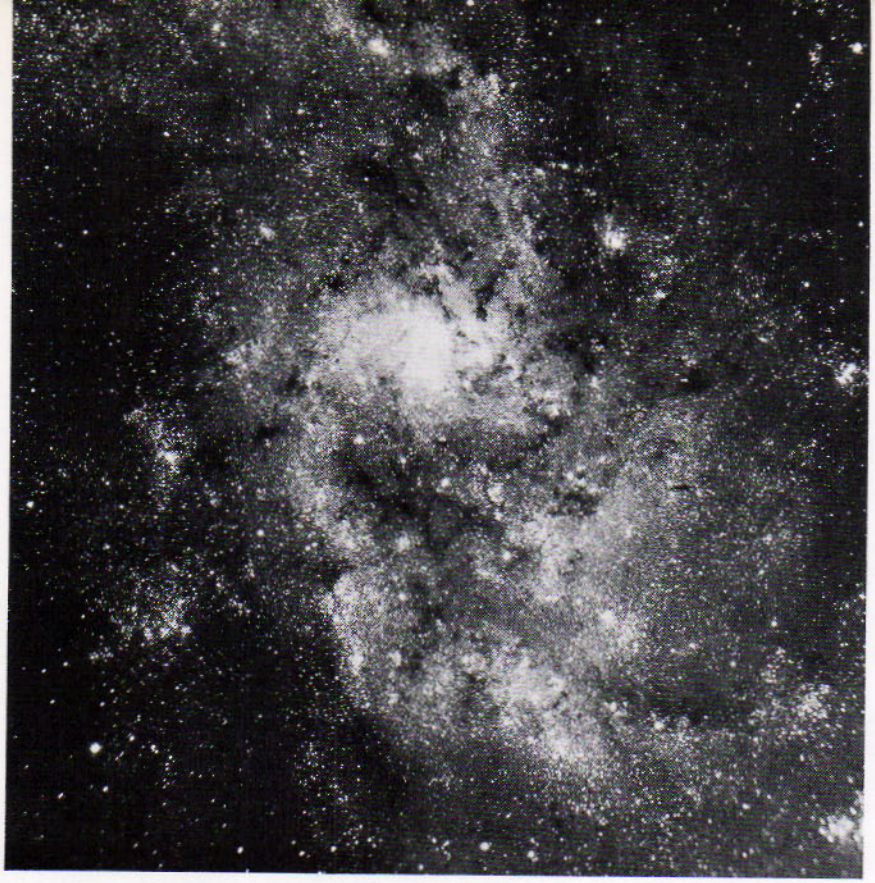


Figure 18.3A
Galaxy NGC 598 in Triangulum. *(Photograph from the Hale Observatories)*

Figure 18.3B
The beautiful Sb galaxy NGC 2841. *(Official U.S. Navy photograph by J. Priser)*

this class to which our own galaxy belongs. In *class Sc* the central regions are reduced to small kernels, and the arms are large and loosely coiled. M 33 in Triangulum (Fig. 18.3A) is representative of this class.

18.4 The Great Spiral in Andromeda

The Andromeda galaxy, or M 31, is the brightest and may also be the nearest of the spiral galaxies. An elongated hazy spot to the naked eye, it is marked in Map 4 with its original name, the "Great Nebula" in Andromeda. It is the central region that appears to the naked eye and, for the most part, to the eye at the telescope. Fainter surrounding parts come out clearly in the photographs, where this galaxy is shown in its true character as a flat Sb spiral inclined 13° from the edgewise presentation.

Because of its inclination, the nearly circular spiral appears oval to us. Its length in the photograph is about 3°, and the gradual fading at the edge suggests greater dimensions. Baade's studies of the faint outskirts increase the diameter to 4°.5, which at the distance of 6.6 million parsecs corresponds to a linear diameter of 55,000 parsecs.

Separate stars in the arms of the Andromeda spiral were first observed by Hubble in his photographs with the 100-inch Mount Wilson telescope. Hubble also discovered the halo of globular clusters surrounding the spiral. Separate stars in the central region and in the disk outside the conspicuous arms were first observed by Baade with the same telescope. The spiral arms of that galaxy contain all the features found in the Milky Way right around us. These include dark dust clouds and emission nebulae made luminous by the radiations of type O and B stars in their vicinities.

Figure 18.4A
The great spiral galaxy in Andromeda, M 31. Two elliptical companion galaxies are M 32 (above M 31) and NGC 205 (below M 31). *(Photograph from the Hale Observatories)*

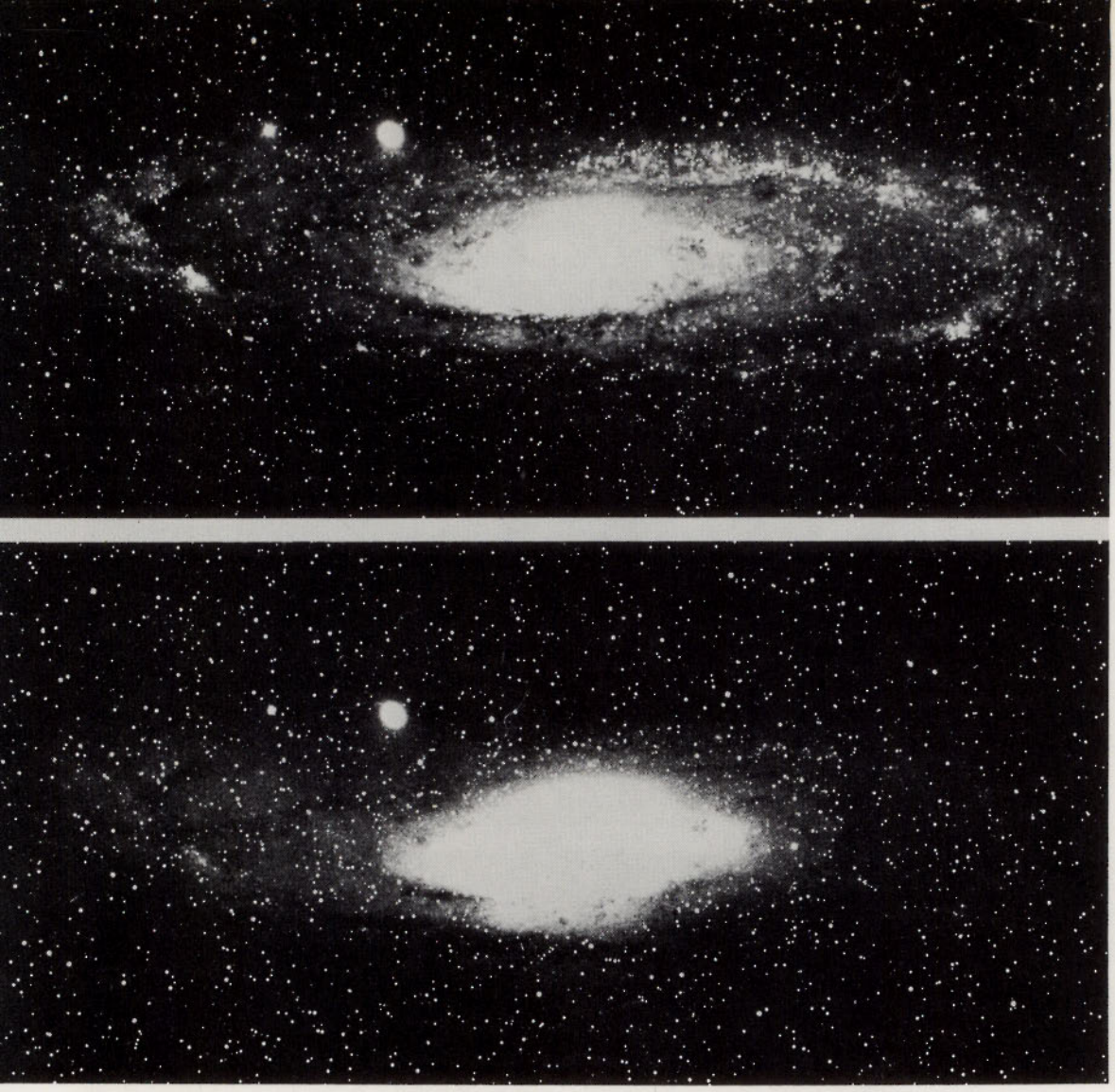


Figure 18.4B
The Andromeda spiral galaxy in ultraviolet (above) and red light (below). Blue stars in the spiral arms clearly show the spiral structure in the ultraviolet photograph. The central region is more conspicuous in the red photograph. *(Courtesy of F. D. Miller, University of Michigan Observatory)*

Except for its larger dimensions, the neighboring Andromeda spiral seems to resemble closely our own spiral galaxy. The relations between the structural features can be more clearly observed in the view from outside. What has been learned about the Andromeda spiral has served to guide the investigations of the system of the Milky Way itself.

18.5 Edgewise Spirals

Many spiral galaxies are presented with their flat disks edgewise to us, or nearly so, as would be expected. They show clearly (Fig. 18.5) the polar flattening of the central region and the fidelity with which the material of the disk keeps to the principal plane. Characteristic of the edgewise spirals is the dark streak that sometimes seems to cut them in

Figure 18.5
Two spiral galaxies seen edgewise; NGC 4565 (top) and NGC 4594 (bottom). The latter is called the Sombrero. *(Photographs from the Hale Observatories)*

two. Just as the dust near the principal plane of our galaxy prevents us from looking out along this plane, so the dust in the arms of galaxies obstructs the view from outside them.

18.6 Barred Spirals

More than two thirds of the recognized spiral galaxies are of the normal type. The remainder are *barred spirals.* Instead of emerging directly from the central region, the arms of this type usually begin abruptly at the extremities of a broad bright bar that extends through the center. Barred spirals are arranged in Hubble's classification in a series paralleling that of the normal type. The classes are designated SBa, SBb, and SBc.

As the series of barred spirals progresses, the arms build up and unwind. In class SBa the arms are joined to form an elliptical ring, so that the galaxy resembles the Greek letter theta (θ). In class SBb the ring is broken and the free ends are spread, so that the galaxy is more nearly like the normal spiral. In class SBc the ends are so far separated that the galaxy has the form of the letter S.

Some spiral galaxies of both types have bright rings inside the spiral patterns. These rings, from which the arms begin tangentially, are said to have so nearly the same linear diameters that they might serve as distance indicators.

Figure 18.6A (left) NGC 1300, barred spiral galaxy in Eridanus. *(Photograph from the Hale Observatories)*

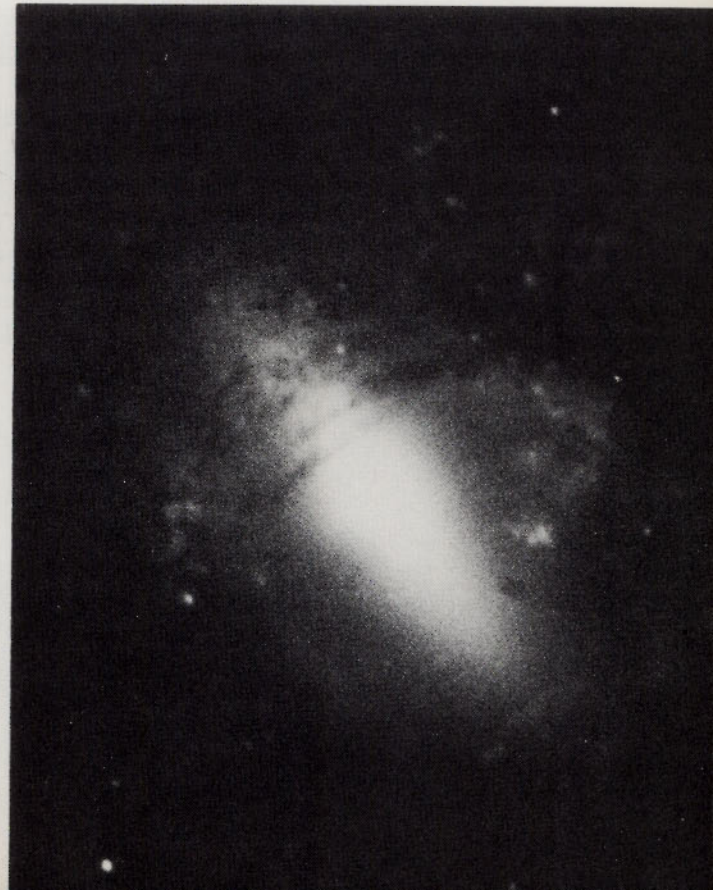


Figure 18.6B (right) NGC 2685, a barred spiral galaxy in Ursa Major, and a very peculiar galaxy. *(Photograph from the Hale Observatories)*

18.7 The Magellanic Clouds

Nearest of the other galaxies and regarded as satellites of our own galaxy, the two Magellanic Clouds are plainly visible to the unaided eye, but are too close to the south celestial pole to be viewed clearly north of the tropical zone. They were mentioned by early voyagers into the southern hemisphere and are named in honor of the navigator Ferdinand Magellan.

The Large Cloud is in the constellation Doradus. As it appears in the photograph (Fig. 18.7) its apparent diameter is 12°, or half the length of the Great Dipper. The Small Cloud is in Tucana, and its apparent diameter is 8°. At distances somewhat greater than 46,000 parsecs the optical linear diameters of the two objects are, respectively, 9,800 and 7,600 parsecs. A quantity of neutral hydrogen equal to a billion solar masses occurs in and around both Clouds, also forming a link between them, as F. J. Kerr and associates at Sydney have found.

The Magellanic Clouds have been classed generally among the irregular galaxies. However, the Australian radio astronomers have observed the rotations of both Clouds, and G. De Vaucouleurs' photographs with a small camera at Mount Stromlo seem to him to show that the Clouds are flat one-armed barred spirals having larger angular dimensions than were previously assigned them. H. M. Johnson sees in his photographs of the Large Cloud some resemblance to an asymmetrical Sc spiral galaxy.

Figure 18.7
The 30 Doradus nebula in the Large Magellanic Cloud. *(Photograph by Hogg at Mt. Stromlo, courtesy of B. Bok)*

One aspect of the Magellanic Clouds has been repeatedly emphasized and should not be overlooked: they are the nearest galaxies to the Milky Way and form a stepping stone to the other galaxies. A telescope of only 20-inch aperture can study the "clouds" with the same detail that the 200-inch can on M31. The Large Cloud shows all of the features that we could ask for in a nearby galaxy: supergiant stars, neutral hydrogen, an abundance of Cepheid variable stars, globular clusters and emission nebulae including the magnificent 30 Doradus Nebula.

18.8 Irregular Galaxies

Irregular galaxies lack the orderly structure that characterizes the elliptical and spiral systems. They constitute a relatively small class among the galaxies where the forms can be observed with present instruments. Aside from the Magellanic Clouds, the nearest examples are NGC 6822

Figure 18.8
The Large Magellanic Cloud. This neighboring galaxy is cited as an example of an irregular galaxy although some astronomers believe that it may be a barred spiral. *(Lick Observatory photograph)*

in Sagittarius and IC 1613 in Sextans. These are examples of a type that contains O and B stars and emission nebulae.

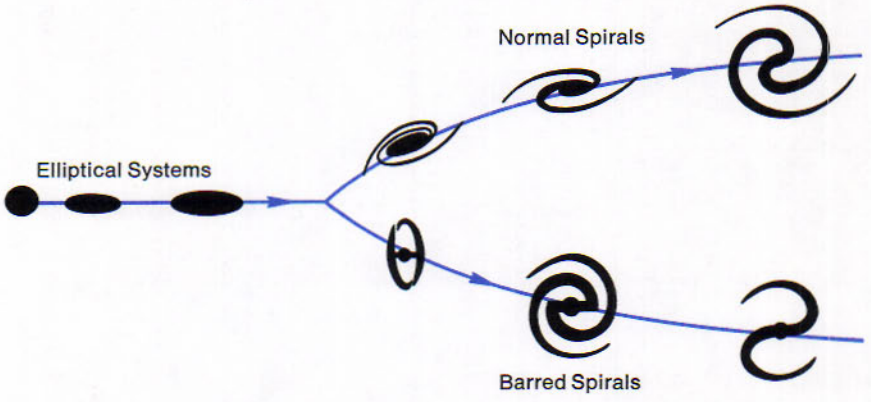
Several hundred irregular and otherwise peculiar galaxies are recognized in the Palomar sky-survey with the 48-inch Schmidt telescope. Some of these are being studied by E. M. and G. R. Burbidge in their photographs with the 82-inch telescope of the McDonald Observatory. In a preliminary report these investigators conclude that the irregular galaxies represent very early stages of evolution, because they cannot last long in their present form and have not yet attained rotational symmetry. The existence of such youthful types implies that galaxies may be forming even today. The question of the evolution of the galaxies will be discussed further in Chapter 19.

18.9 Sequence of Regular Galaxies

The separate series of galaxies having rotational symmetry were joined by Hubble into a continuous sequence. From the compact spherical form at one end to the most open spirals at the other, it might be considered a progression of expansion. At the termination of the elliptical series the sequence divides into parallel branches of normal and barred spirals. Classes S0 and SB0, not represented in Fig. 18.9A, were introduced later after the division before the spiral structures begin.

This sequence provided a basis for thinking about the evolution of galaxies. A progression from left to right in the figure might suggest the streaming of material from opposite sides of the most flattened elliptical systems and the gradual building up of the spiral arms at the expense of the central regions. A progression in the opposite direction, however, is suggested by the trend of the stellar compositions of galaxies (18.20).

Figure 18.9A
Sequence of elliptical and spiral galaxies. Hubble later placed the classes S0 and SB0 near the division of the sequence into two branches. *(Adapted from a diagram by Edwin P. Hubble)*



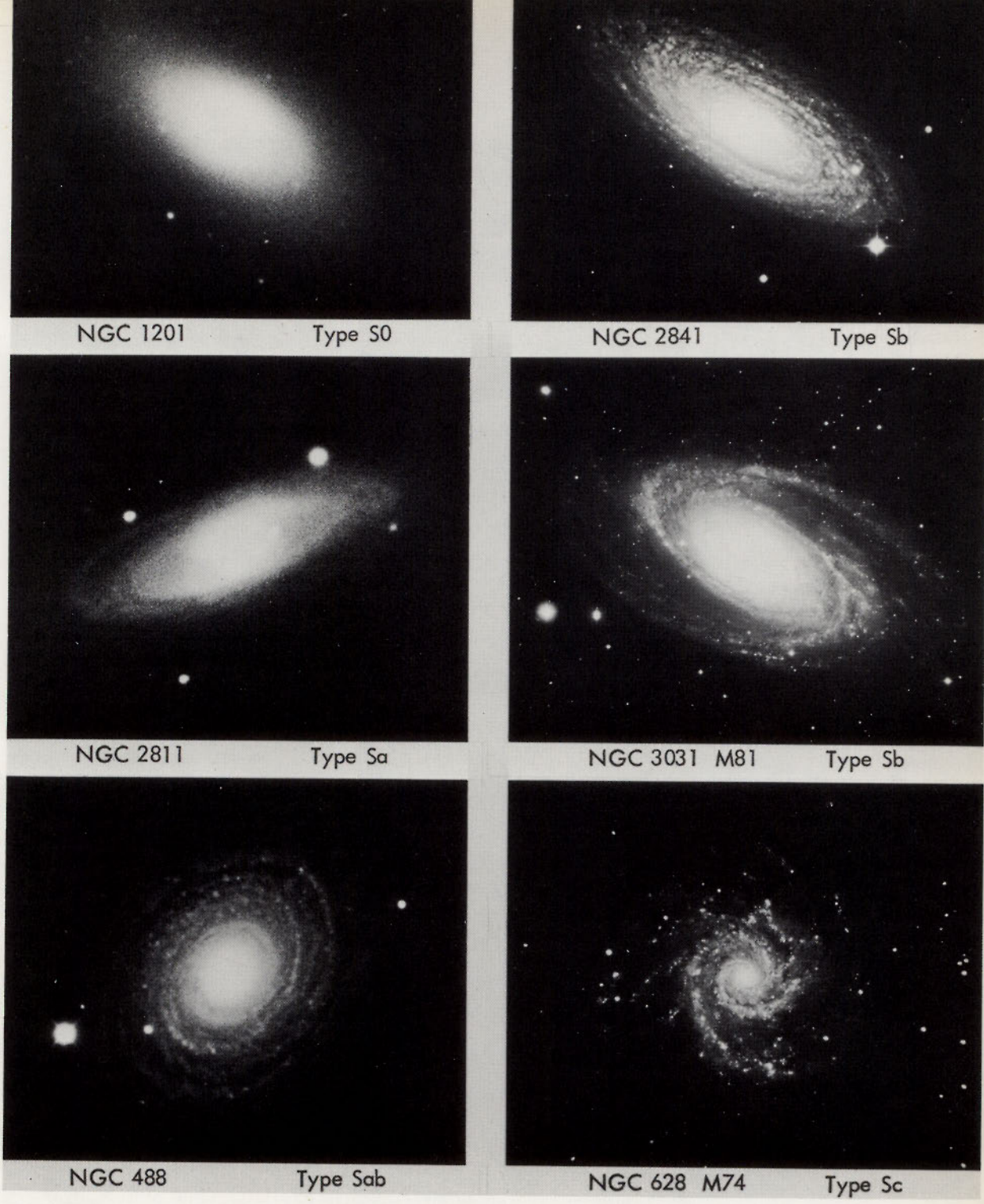


Figure 18.9B
Classification of normal galaxies. *(Photograph from the Hale Observatories)*

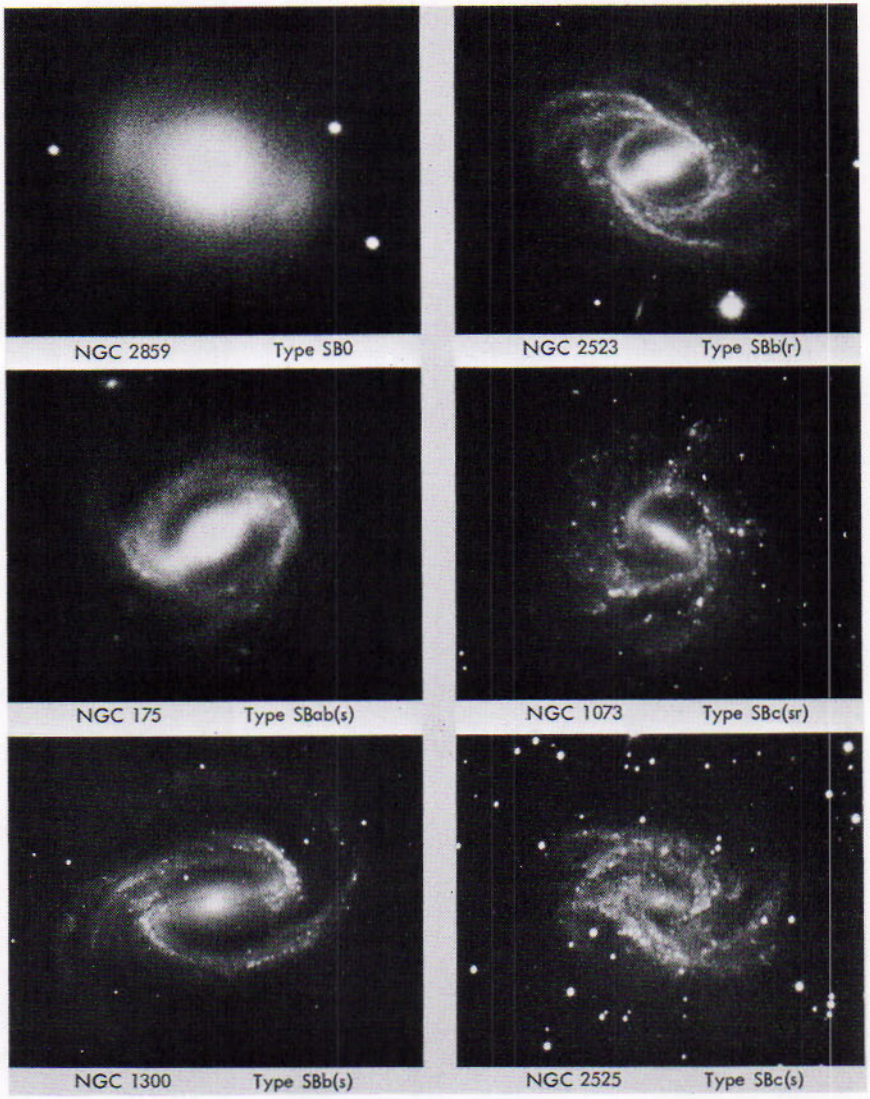


Figure 18.9C
Classification of barred galaxies. *(Photograph from the Hale Observatories)*

18.10 The Status of SO Galaxies

The earliest interpretation of S0 galaxies was suggested by their introduction in Hubble's sequence between the flattest ellipticals and the most compact spirals. It seemed that the S0 galaxies might be an intermediate stage in the evolution between ellipticals and spirals. The greater abundance of these galaxies in the denser clusters of galaxies, however, led to a different interpretation. Following the hypothesis of L. Spitzer and of W. Baade, the S0 galaxies came to be regarded as the central regions of former spiral galaxies.

Although the ratios of the diameters of galaxies to the distances of their nearest neighbors are much greater than are those of the stars, they still average small enough to preclude the possibility of frequent collisions of galaxies in general. In the denser clusters, however, it was estimated that the random motion of a galaxy might bring it into collision with another galaxy several times in the course of a billion years. Collision of a spiral galaxy with a neighbor could be violent enough to heat the gases of the spiral arms intensely, accelerating the atomic motions to exceed the velocity of escape from the galaxy. Thus the gas and dust would be dissipated by the collisions, but the loss of the stellar structure of the arms would require further explanation.

Two objections to the collision hypothesis are: (1) that S0 galaxies are observed in less dense clusters as well, and (2) that collisions even in the denser clusters could become insignificant if the distance scale of the universe needs sufficient enlarging, which may be the case. A current interpretation of S0 galaxies is that they are the oldest of the ellipticals and are especially abundant in the denser clusters because in this environment they age more rapidly.

Another trend away from the collision hypothesis of galaxies is found in the present explanation of radio sources such as Cygnus A.

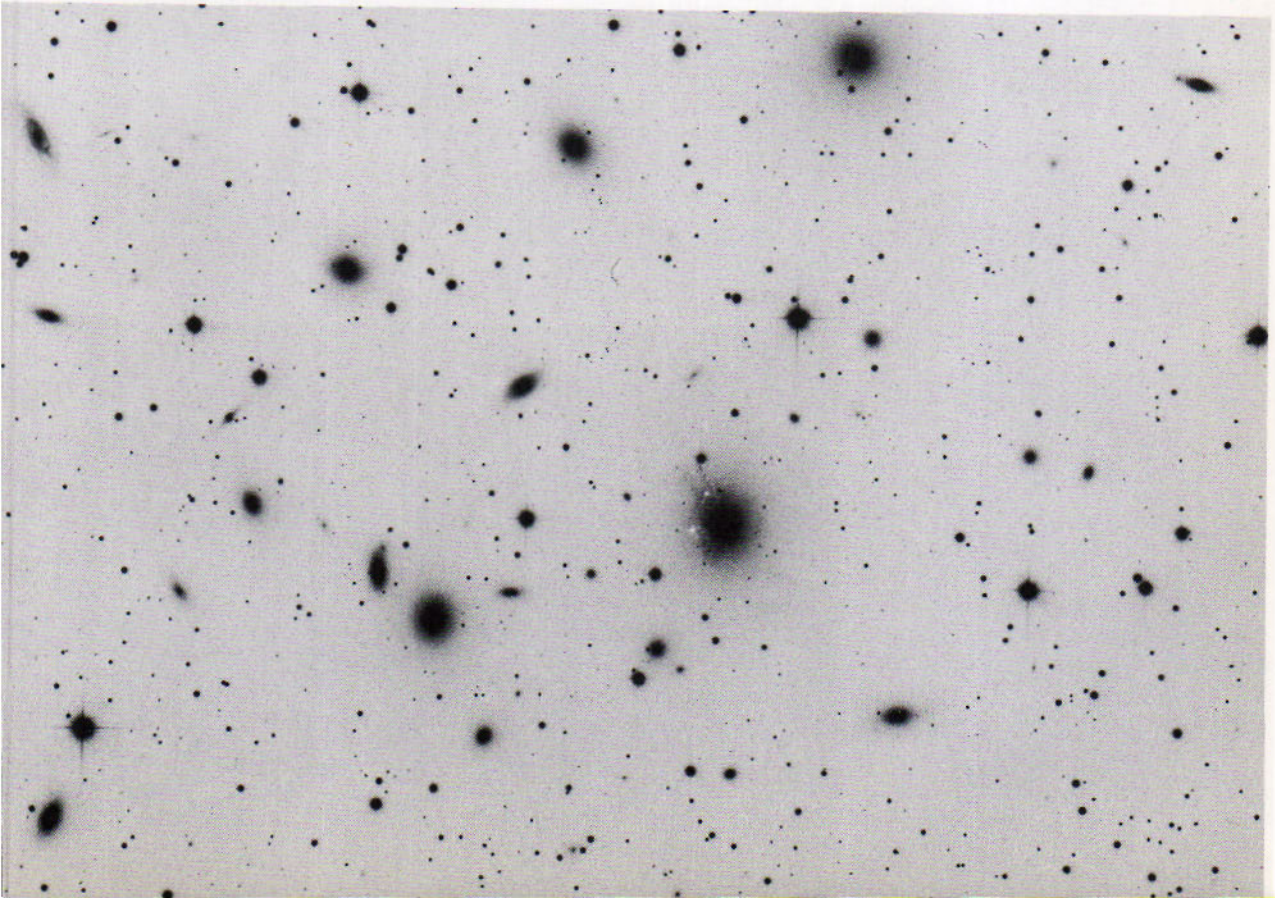


Figure 18.10
NGC 1275, a cluster of galaxies in Perseus. So spirals dominate this remarkable cluster. *(Photograph from the Hale Observatories)*

18.11 Radio Source Cygnus A

A powerful source of radio emission, known as Cygnus A because it was the first of these to be recognized in the constellation Cygnus, has its optical counterpart as photographed in Fig. 18.11. The photograph shows a pair of overlapping condensations centered about 2″ apart and surrounded by a large dim halo. The object is one of the brightest members of a remote rich cluster of galaxies. Baade and Minkowski interpreted the condensations as two colliding galaxies that have interpenetrated until their centers are only a few thousand light years apart. Their gases are intensely heated; almost half of the optical radiation appears in the spectrum as widened bright lines of hydrogen and other elements.

The collision hypothesis for Cygnus A and some other abnormal galaxies, such as NGC 5128 (Fig. 17.20), was accepted by most astronomers until 1960, when apparent discrepancies began to be noticed. It was found that the radio emission comes from two regions centered about 80″ on either side of the optical object in Cygnus, which itself is only 30″ in its longest diameter. The Soviet radio astronomer J. S. Shklovsky concluded that the optical feature corresponding to Cygnus A is most likely a double galaxy with the nucleus forming a close pair and that the components are generically related. From his calculations and those of G. R. Burbidge it seemed improbable that the very strong radio emission of Cygnus A could result from the collision of galaxies. Instead, Shklovsky attributed this radiation to tenuous gas clouds spreading from the sites of earlier supernovae outbursts.

An equally dramatic hypothesis of the instability of systems of galaxies was proposed in 1954 by V. A. Ambartsumyan. The hypothesis conjectures that at least some clusters of galaxies possess enormous and unexplained energy that is tearing them apart. The implications of the proposal might extend from clusters to individual galaxies and their evolution. Radio sources such as Cygnus A are viewed as violently exploding galaxies; their nuclei are dividing to produce new galaxies.

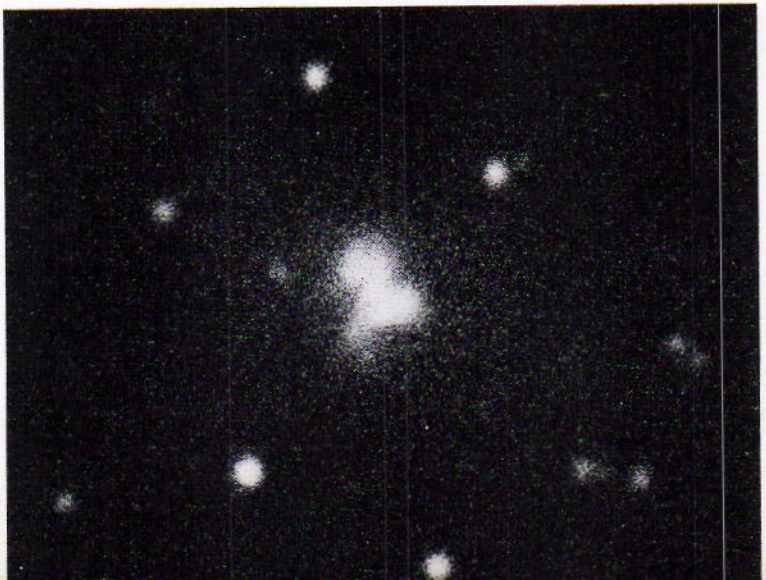


Figure 18.11
Optical identification of the radiosource Cygnus A. *(Photograph from the Hale Observatories)*

18.12 Peculiar and Interacting Galaxies

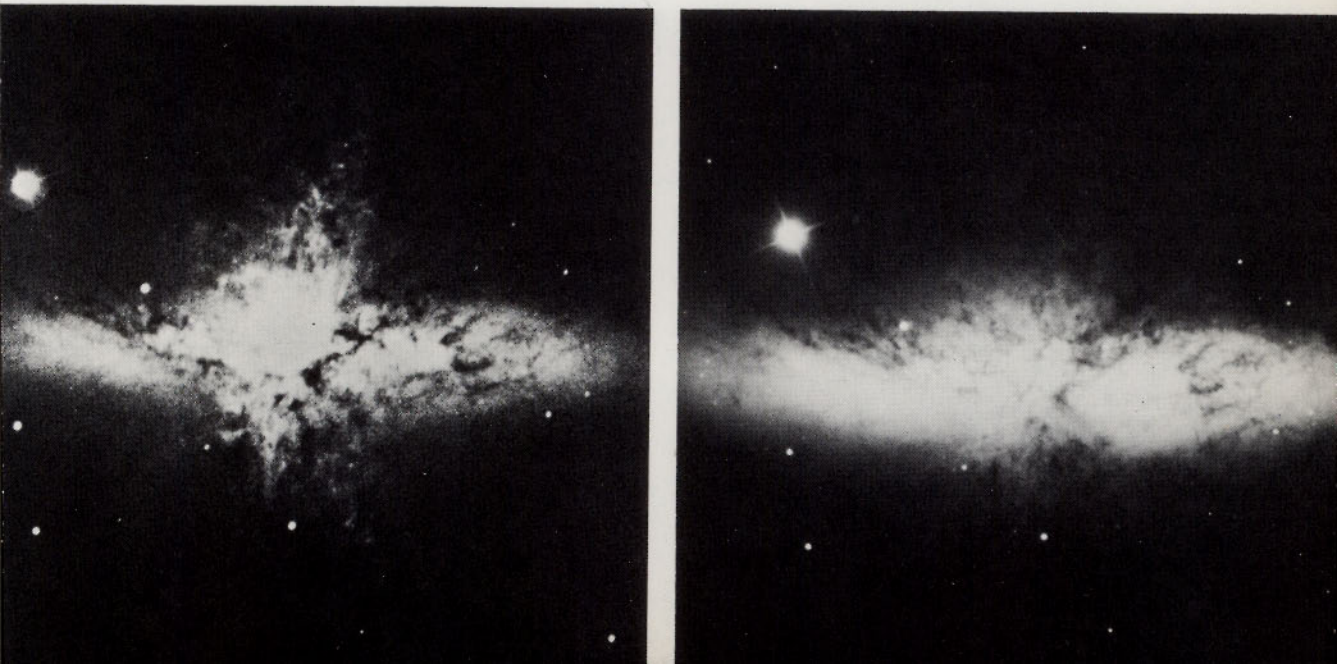
Some objections have been noted to earlier ideas that S0 galaxies might be former spirals which lost their arms in collisions and that radio sources such as Cygnus A might be powered by interpenetrating galaxies. Although collisions between galaxies are perhaps rare enough to be disregarded, effects of interactions between close pairs of galaxies are frequently observed. In some cases the structural features of the galaxies are distorted. While these galaxies do not constitute a separate class morphologically, they nevertheless form an interesting category. Peculiar galaxies are best understood by discussing the galaxy known as M82 (Figure 18.12A).

Visual inspection of a photograph of M82 leads one to surmise that a violent event on a grand scale has taken place right in the center of the galaxy. Indeed, spectroscopic studies indicate that material is leaving the central regions at more than 4000 km/sec. A diffuse light component suggests an origin in synchrotron radiation and J. S. Hall and Aina Elvius have measured polarization along the filaments. One must conclude that this galaxy is peculiar because of some internal event. Peculiar galaxies are often associated with the bi-lobed radio sources, as has been pointed out by H. Arp.

See Plate XIII

F. Zwicky, B. A. Vorontsov-Velyaminov, H. Arp and others have drawn attention to numerous cases of interacting galaxies (Fig. 12.18D). The classic cases are M51 (Fig. 18.12E), the quintet VV172 (Fig. 18.12F), and NGC 3808. A brief study of any reasonably dense cluster of galaxies shows additional cases. Often interacting galaxies are classed as peculiar

Figure 18.12A
The peculiar galaxy M 82 (NGC 3034) seen in blue light (left) and Hα light (right). *(Photographs from the Hale Observatories)*



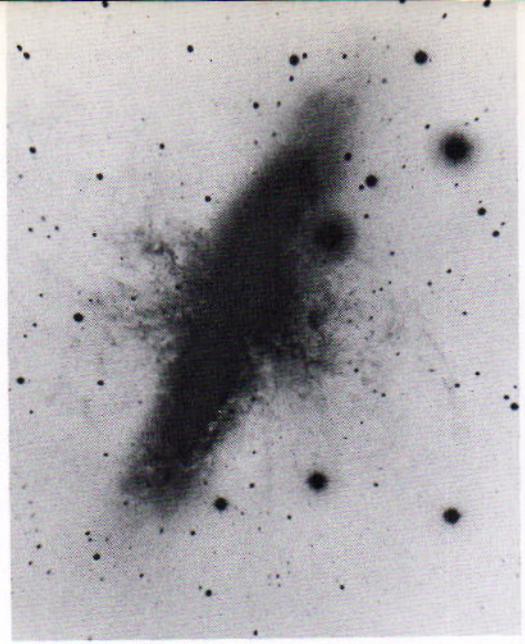
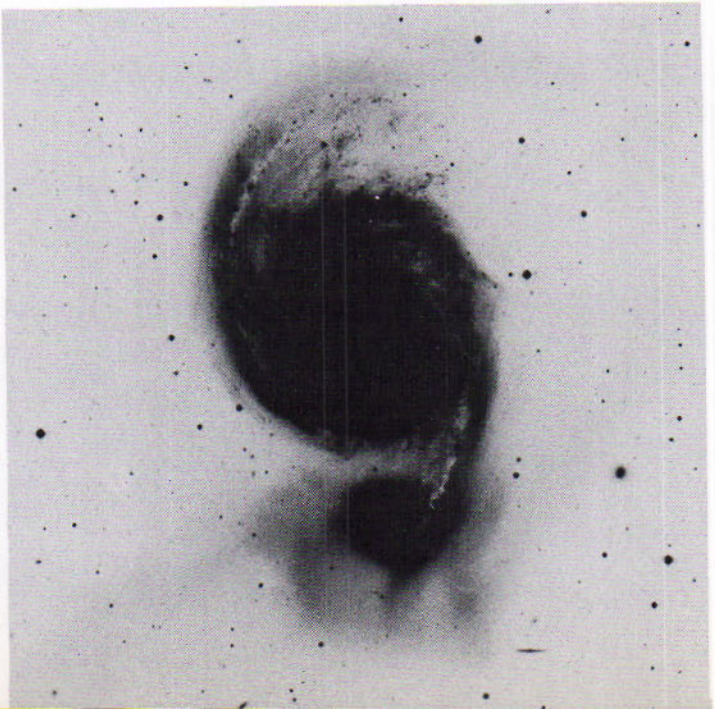


Figure 18.12B (Top left)
M 82, with the long filaments clearly defined in this long exposure. *(Photograph from the Hale Observatories)*

Figure 18.12C (Top right)
Two views of M 82 in polarized light, showing polarization of the blue filaments. The electric vector is indicated by arrows. *(Photograph from the Hale Observatories)*

Figure 18.12D (Bottom left)
Peculiar interacting galaxies called Atlas 87. *(Courtesy of H. Arp, the Hale Observatories)*

Figure 18.12E (Bottom right)
M 51 and its interacting companion. This beautiful photograph by S. van den Bergh clearly shows the halo disturbances due to interaction between the galaxies. See also Figure 18.22C. *(Photograph from the Hale Observatories)*



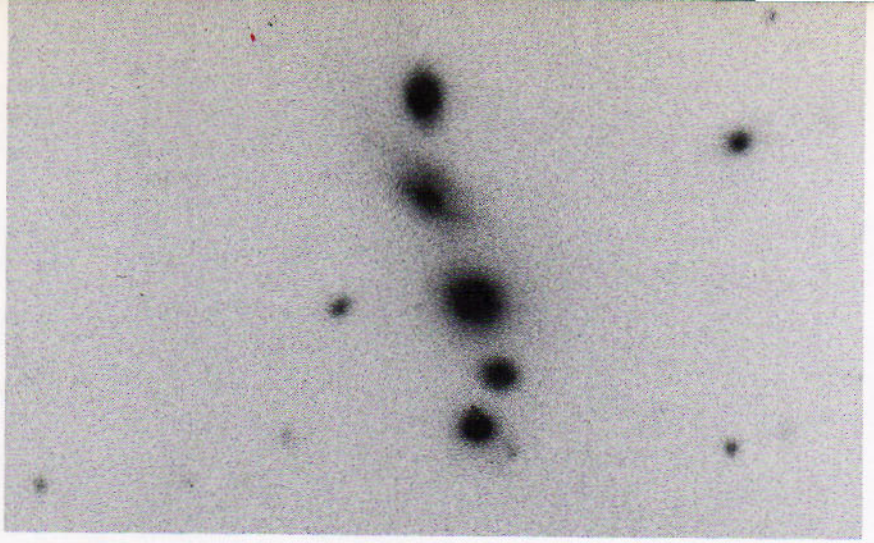


Figure 18.12F
An unusual quintet of galaxies, VV 172. The upper galaxy has a redshift three times larger than the other four. *(Photograph from the Hale Observatories)*

galaxies as well or at least the smaller of two interacting galaxies is. This suggests that the smaller galaxy may have been ejected from the larger. The smaller galaxy exhibits high excitation emission lines and has a high surface brightness. T. L. Page and the Burbidges believe the interacting pairs are in orbital motion; at the scale of galaxies they would have to be called cosmic binaries.

PARTICULAR FEATURES OF GALAXIES

Normal galaxies exhibit a stellar and gaseous content similar to the Milky Way Galaxy, therefore objects in those galaxies should be the same as in the Milky Way. This analogy allows us to determine the distances of galaxies.

18.13 Distance Indicators of Galaxies

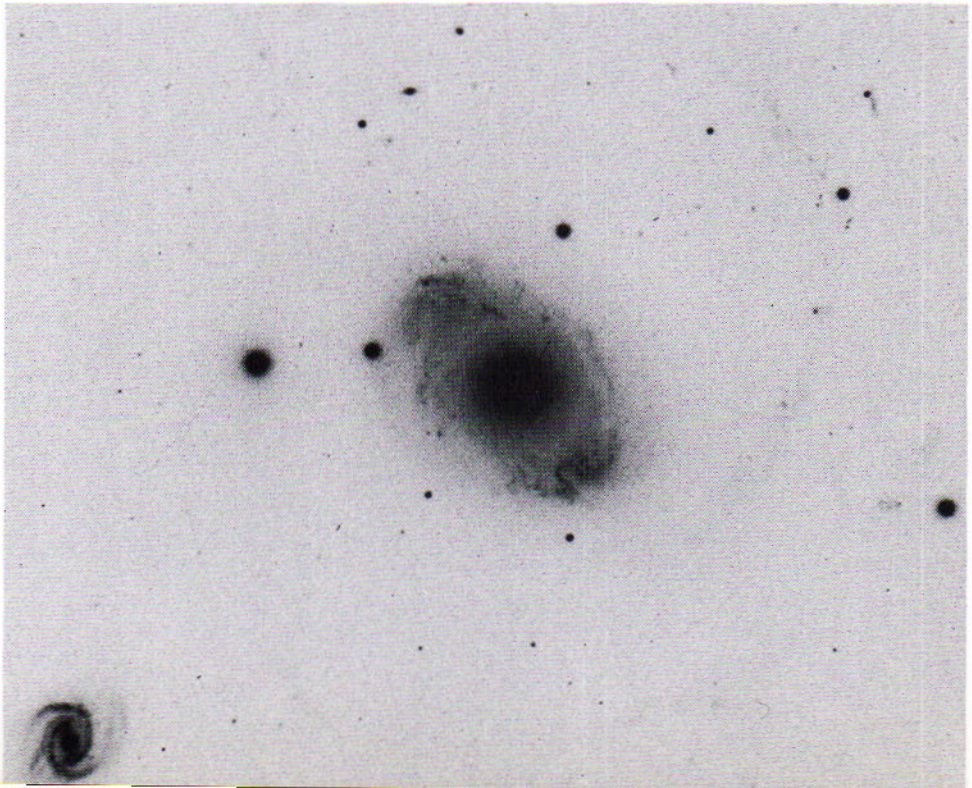
E. Hubble, in 1924, established the spirals and other extragalactic nebulae as stellar systems beyond our Milky Way. His photographs with the 100-inch Mount Wilson telescope showed the arms of the spirals M 31 and M 33 partly resolved into stars. Some of the stars proved to be classical cepheid variables. Hubble observed that the curve representing the logarithms of the periods of the light variations of these cepheids plotted against their median apparent magnitudes seemed to have the same form as Shapley's log(period)-absolute magnitude curve. It remained to determine the modulus, the difference $m - M$ between corresponding ordinates

of the curves in the two cases. The distance, r, in parsecs, could then be found by the useful formula: $\log r = (m - M + 5)/5$. The great distances that resulted showed conclusively for the first time that the two spirals are exterior to our own galaxy.

Hubble later extended his measures of the distances of galaxies in three steps: (1) By use of cepheid variables that he found in five other nearby galaxies; (2) by the apparent magnitudes of what seemed to be the most luminous stars in somewhat more distant galaxies where cepheids were too faint to be seen in his photographs; (3) by the total apparent brightness of entire galaxies on the assumption that galaxies of the same class had the same absolute brightness. More recent investigations with the 200-inch telescope have shown the need for revising the original distance scale of galaxies, as Hubble himself had considered quite possible.

The upward revision of the luminosities of classical cepheids by Baade in 1952 required multiplication by the factor 2 of the previously assigned distances of almost all galaxies; and the factor was soon raised to 3 for other reasons. In addition, it became evident that the period-luminosity curves of classical cepheids occupy a band at least 1 magnitude wide (Fig. 13.8) instead of being on the same line, so that the luminosity of

Figure 18.13
Galaxy NGC 4151 showing bright HII regions. Spiral structure can be traced to the edge of the region pictured because of the HII regions. *(Photograph from the Hale Observatories)*



a cepheid derived from the original single curve may give a distance as much as 25 per cent in error.

Sandage's red photographs with the 200-inch telescope seem to have demonstrated that the supposed brightest stars originally employed as distance indicators in intermediate galaxies were frequently bright nebulae (Fig. 18.13). He estimates that the nebulous spots may appear 1.8 magnitudes brighter than the brightest stars in a particular galaxy.

In his paper of 1962 on the distance scale, Sandage adopts classical cepheids as primary indicators of distances of the nearer galaxies. He employs an improved calibration of the period-luminosity relation for the cepheids and more precise photometric measures of very faint stars in and near these galaxies. The new determination for the Andromeda spiral suggests a 50 per cent increase in the distances of the nearer galaxies (Table 18.I) over the values accepted after the revision of 1952.

18.14 Novae in Galaxies

The normal novae that flare out in the exterior galaxies resemble the galactic novae (13.16) in the order of their luminosity at maximum brightness and in the character of their light variations. Hubble published in 1929 a survey of 82 novae in M 31, and H. C. Arp reported in 1956 on his studies of 30 other novae, which he had discovered in that galaxy with the 60-inch Mount Wilson telescope. Because the distance factor is eliminated, such surveys in a single galaxy give a clearer account of the behavior of novae.

Arp concludes that about 26 normal novae appear annually in the Andromeda spiral; one fourth of these are likely to be concealed by dust clouds of that galaxy or else within its more congested central region. The durations of the observed nova outbursts, while they were brighter than magnitude 20, ranged from 5 to 150 days, and the corresponding absolute magnitudes at the maxima from −8.5 to −6.2. The faster novae faded more smoothly after the maxima than did the slower ones.

18.15 Supernovae

Supernovae sometimes rise to considerable fractions of the total brightness of the galaxies in which they appear. The first of these to be recorded in another galaxy flared out in 1885 near the central region of the Andromeda spiral; at greatest brightness it appeared as a star of the 6th visual magnitude.

Supernovae are stars more massive than the sun, which in the later stages of their evolution attain very high central temperatures and explode. In the explosions they blow off into space excess gaseous material that may amount to as much as one or two solar masses. The outbursts occur in all types of galaxies at the average rate of one per galaxy in 300 to 400 years, according to Zwicky; this frequency is less than has been estimated for our own galaxy (13.21), partly because ours is larger

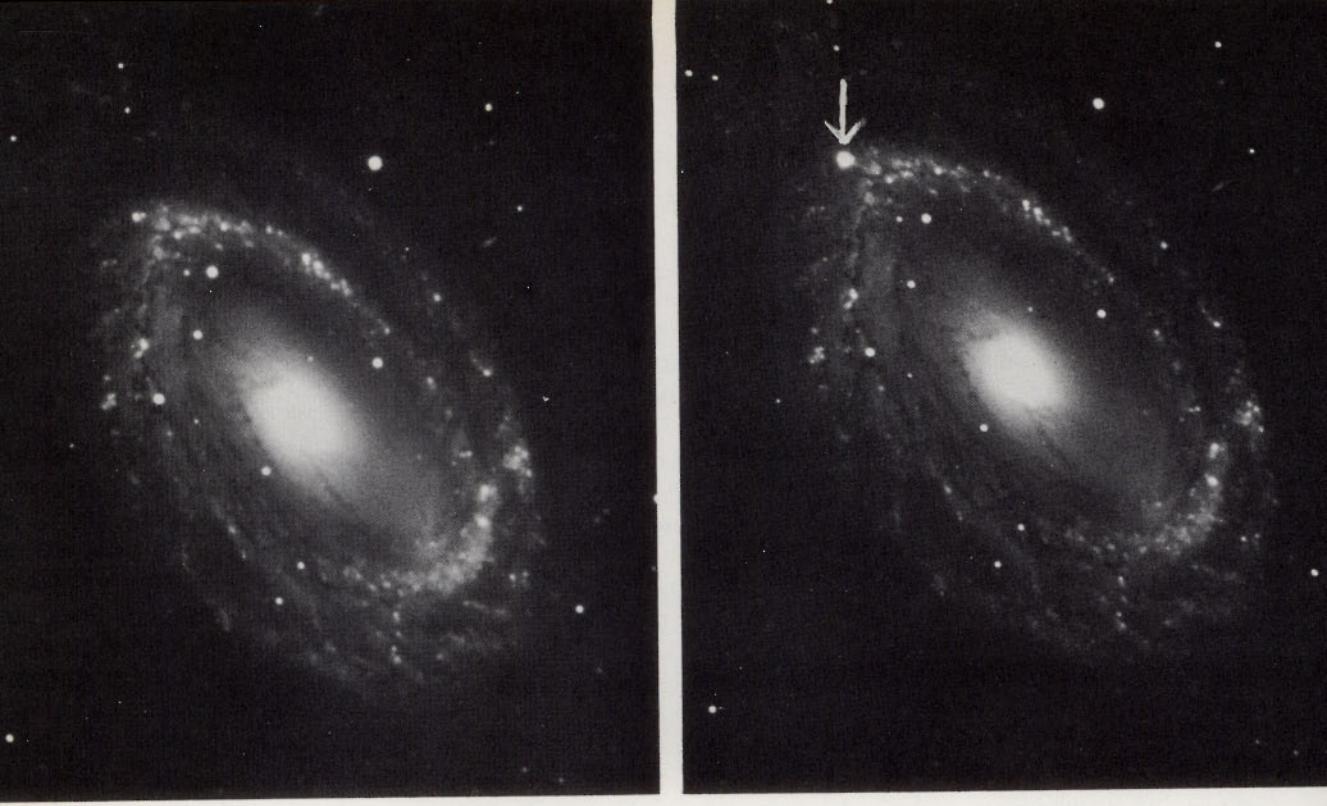
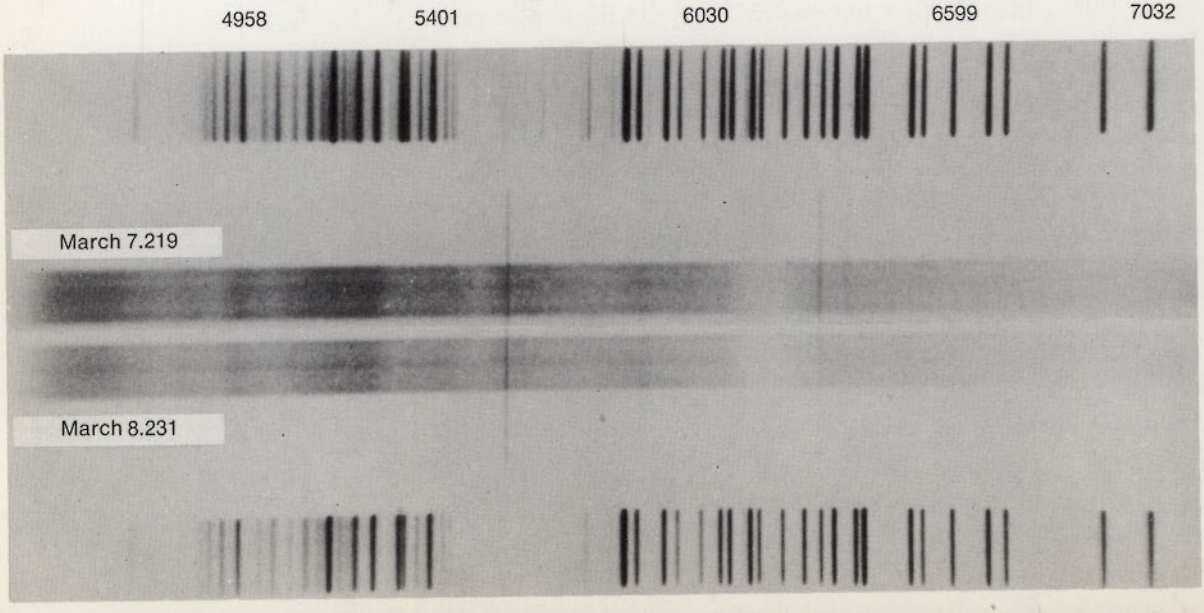


Figure 18.15A
Galaxy NGC 4725 in Coma Berenices, photographed on May 10, 1940 (left) and on January 2, 1941, showing the supernova (arrow, at right). *(Photographs from the Hale Observatories)*

Figure 18.15B
Spectrum of a supernova in the galaxy NGC 3389, located in the Leo cluster of galaxies some 10^7 parsecs distant. The broad emission bands are those of a typical Type I supernova. Spectrograms were taken by Drs. K. Ford and Vera Rubin using the Kitt Peak 84-inch (213.3 cm) telescope and the Carnegie Institution's image tube system. *(By permission of the Publications of the Astronomical Society of the Pacific)*



than average. The outbursts of supernovae are most likely to be detected in repeated photographs of clusters of galaxies. Supernovae are of two types:

Type I. The members of this type rise to about absolute magnitude −15.5, or more than 100 million times as luminous as the sun. After a rapid drop for 100 days following maximum brightness, they decline less rapidly and more smoothly. The extreme width of the lines in the spectra (Fig. 18.15B) testify to the violence of the explosions. Supernovae of type I are particularly deficient in hydrogen.

Type II. The members of this type attain about absolute magnitude −13.5, and may blow away most of the original mass. They decline after maximum brightness more slowly at first than do those of the first type.

18.16 Globular Clusters Around Galaxies

Hubble was the first to report the presence of at least 250 fuzzy disks projected against and near the borders of the Andromeda spiral. He believed that they were globular clusters of stars like those in the halo around our galaxy, as they proved to be; the outer parts of two of them were later resolved into stars with the 200-inch telescope. The clusters of M 31 are as luminous as our own clusters. On the former scale of distances they seemed to be less luminous than ours, one of the several indications that the scale was too small. Globular clusters are observed around other galaxies as well.

These great clusters are presumably as old as the galaxies themselves. Massive enough to remain stable under their own gravitation, they may provide important clues to the early history of the galaxies.

Figure 18.16
Globular Clusters around M 87. Several hundred globular clusters can be detected around this elliptical galaxy. *(Lick Observatory photograph)*

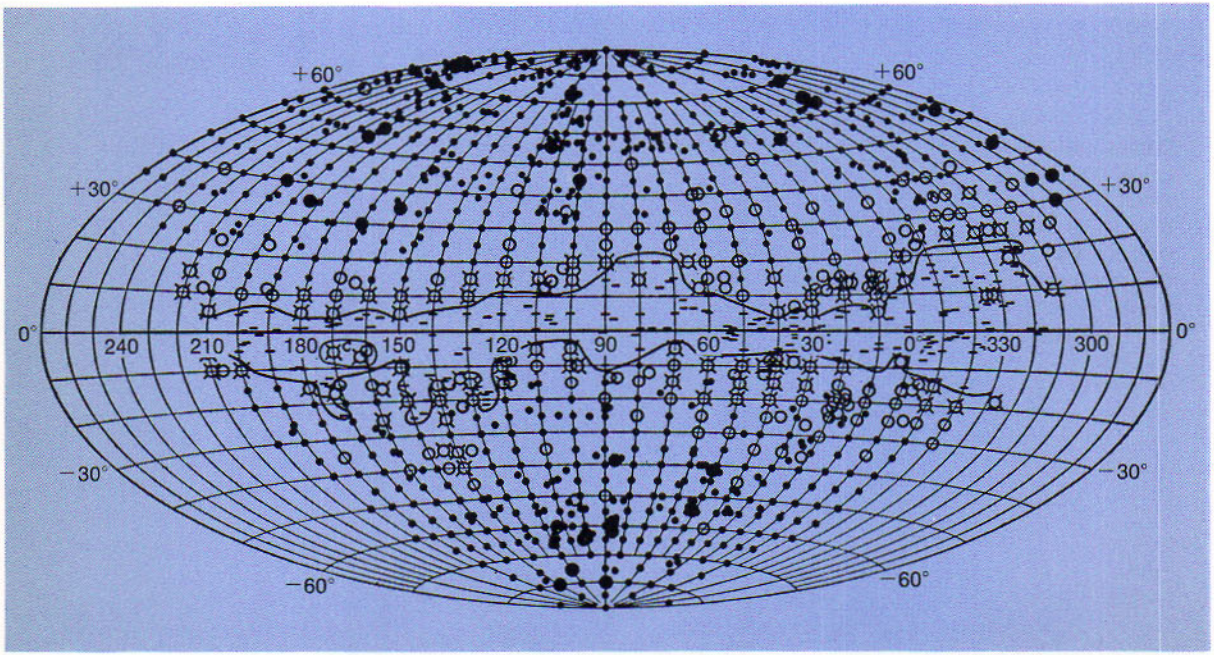
DISTRIBUTION OF GALAXIES

Galaxies are seen in all directions when allowance for obscuration by the Milky Way is made. The seventeen nearest galaxies seem to form a group. Grouping, or clustering, of galaxies is common; the richer clusters being made up of many thousands of galaxies.

18.17 Their Surface Distribution

The diagram of Fig. 18.17 represents the results of a photographic survey by Hubble with the Mount Wilson telescopes. He counted the galaxies visible in each of 1283 sample areas over three fourths of the heavens. The numbers are greatest around the galactic poles; they become fewer with decreasing galactic latitude, until near the equator there are almost no galaxies to be seen.

Figure 18.17
Obscuration of galaxies by dust clouds of the Milky Way. In this representation of the whole celestial sphere the numbers around the edge are galactic latitudes. Filled dots indicate regions where the galaxies are more numerous than average, circles where they are less numerous than average, and dashes where they are not seen at all. The obscuration is nearly complete in an irregular band along the galactic equator, and is partial for some distance north and south of this "region of avoidance." *(Diagram by Edwin P. Hubble, Mount Wilson Observatory)*



The counts near the poles average 80 galaxies per unit area. Obscuration of 1 magnitude in their brightness would reduce the number to 20, 2 magnitudes to five, and 3 magnitudes to not much more than one visible galaxy per area. The conclusion is that the dimming of the exterior systems by the dust in our galaxy is not less than 3 magnitudes near the galactic equator. The dust in front of the galactic center reduces the photographic brightness of the galaxies behind it as much as 8 to 10 magnitudes.

18.18 The Local Group

The galactic system is a member of a group of at least 17 galaxies, which occupy an ellipsoidal volume of space more than 2 million light years in its longest dimension. Our galaxy and M 31 are near the two ends of this diameter; these and M 33 are the normal spirals of the group. The Magellanic Clouds and two smaller systems have usually been classed as irregular galaxies. The remaining 10 are elliptical galaxies. Six dwarf ellipticals are typified by the Sculptor and Fornax systems discovered by Shapley in 1938. Four others, two in Leo and one apiece in Draco and Ursa Minor, were found more recently in photographs with the Palomar 48-inch Schmidt telescope. Henrietta Swope, who with Baade studied his photographs of the Draco galaxy, reports that it contains over 260 variable stars, almost all of the RR Lyrae type. In having so many of these variables and in the general features of its color-magnitude diagram this galaxy resembles the globular clusters of stars, but it is much larger and less dense than the clusters.

TABLE 18.I
The Local Group

DESIGNATION	TYPE	DISTANCE (LIGHT YEARS)	APPARENT DIAMETER	LINEAR DIAMETER (LIGHT YEARS)
Galactic system	Sb	.....	.....	80,000
Large Mag. Cloud	I	* 160,000	12°	32,000
Small Mag. Cloud	I	* 190,000	8°	25,000
Ursa Minor system	E	300,000	55′	4,000
Draco system	E	320,000	48′	4,000
Sculptor system	E	340,000	45′	4,000
Fornax system	E	680,000	50′	10,000
Leo II system	E	1,200,000	10′	3,000
NGC 6822	I	1,400,000	20′	9,000
NGC 185	E	1,600,000	14′.5	7,000
NGC 147	E	1,600,000	14′.1	7,000
Leo I system	E	1,800,000	10′	4,000
IC 1613	I	* 2,200,000	17′	10,000
M 31	Sb	* 2,200,000	4°.5	180,000
M 32	E2	* 2,200,000	12′	7,000
NGC 205	E5	* 2,200,000	15′.8	10,000
M 33	Sc	* 2,400,000	62′	50,000
Maffei 1†	E	—	—	—
Maffei 2†	S	—	—	—

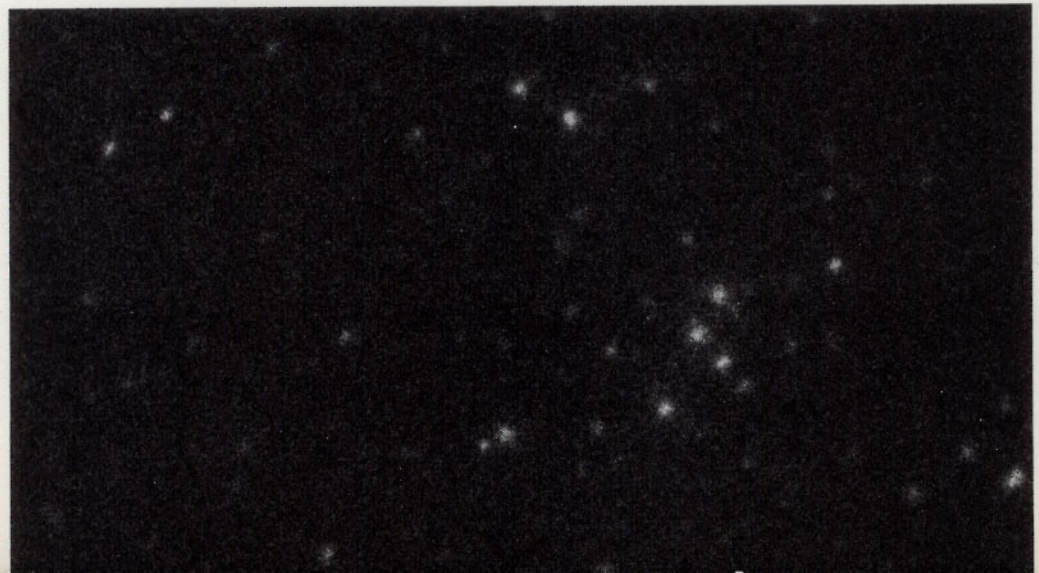
* New values of distances from data by Sandage. † Possible members

The members of the local group are listed in Table 18.I, which contains their types, their distances from us corrected for absorption by intervening dust, and their diameters. The distances marked with asterisks are from new data supplied by A. Sandage. The distance of the Draco system and also of M 31 are as determined by Henrietta Swope. The distances given do not differ significantly from recent values given by S. van den Bergh. van den Bergh, however, calls attention to the fact that the local group tends to form sub-clusters. The Magellanic Clouds and the Galaxy constitute a sub-group with the Magellanic Clouds forming a double sub-system. The Andromeda sub-group is formed of the Andromeda triple (M 31, M 32 and NGC 205) and a dwarf double (NGC 147 and NGC 185). The tendency for galaxies to cluster on a small scale extends to the large scale as well.

18.19 Clusters of Galaxies

Many, and perhaps all, galaxies occur in clusters that fill all space, according to F. Zwicky. The cluster populations range up to several thousand galaxies. His catalog of galaxies and clusters of galaxies lists about 10,000 rich clusters north of declination −30°, which can be recognized on yellow-sensitive photographs with the 48-inch Schmidt telescope of Palomar Observatory. Individual clusters are designated here by the equatorial coordinates of their centers; an example is Cl 1215.6 + 3025, where the first number is the right ascension in hours and minutes, and the second is the declination in degrees and minutes. A few of the most prominent clusters are commonly known by the names of the constellations in which they appear; an example is the Virgo cluster. (See Table 5, Appendix)

Figure 18.19A
Faint cluster of galaxies in Pisces, well over 300 million parsecs from the earth and near the limit of the presently observable universe. If the Great Andromeda galaxy were moved out to where this cluster is, it would not exceed the largest galaxy in this photograph in size or brightness. *(Photograph from the Hale Observatories)*



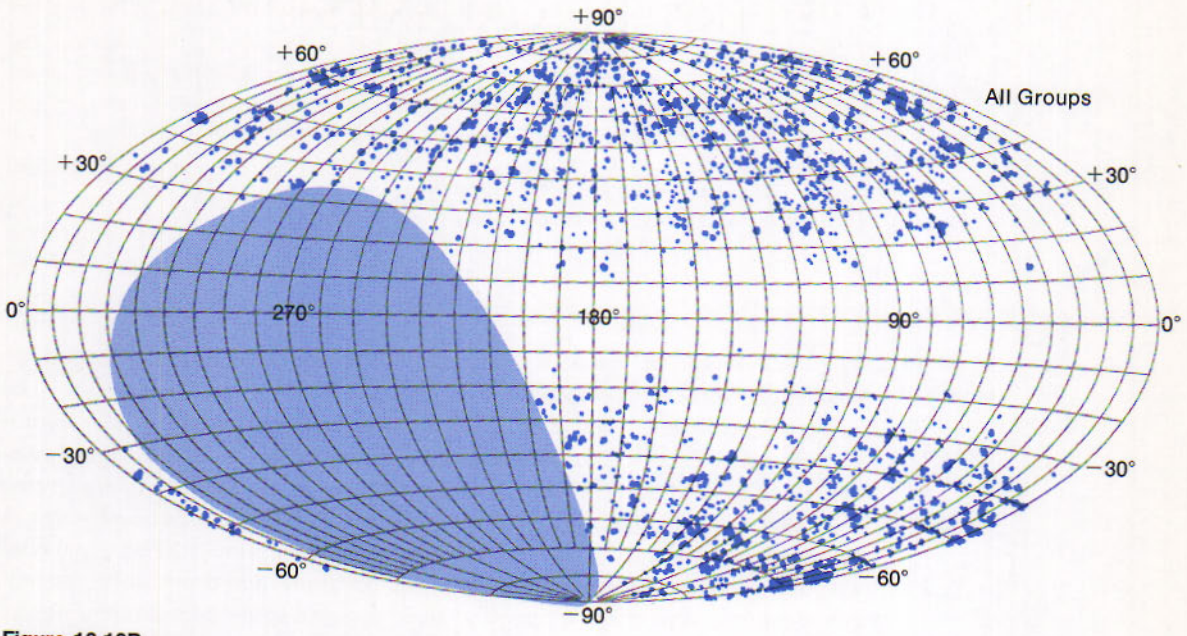


Figure 18.19B
Distribution of Clusters of galaxies. Note a similar distribution in Figure 18.17A. *(Courtesy of G. Abell)*

The Virgo cluster, at a distance previously estimated as 6 million parsecs years, is the nearest and most conspicuous of the larger clusters. Its center is in right ascension $12^h 24^m$, declination $+12°$. Spiral galaxies are numerous in this only moderately compact cluster, and their brightest stars can be observed in photographs with large telescopes. The Coma Berenices cluster is reported to have an apparent diameter of 2° and a membership of 1500 galaxies. This and the Corona Borealis cluster are examples of rich and compact clusters. Their most congested central regions contain a preponderance of S0 galaxies. No evidence of any systematic expansion or rotation of the clusters has been found in extensive spectroscopic studies by Zwicky and Humason.

G. Abell divides the clusters in two classes: *regular* and *irregular*. Regular clusters are generally rich ones in which giant spiral galaxies almost never appear. The Coma cluster of galaxies is a large regular cluster. The famous Virgo cluster of galaxies is an irregular cluster according to Abell's classification. The local group is classified by Abell as a very poor irregular cluster. The same author points out that only four of the seventeen members of the local group would be among the 300 brightest galaxies in the Coma cluster. As we might expect, the distribution of clusters of galaxies (Fig. 18.19B) looks very much like the distribution of galaxies.

SPECTRA OF GALAXIES

The spectra of galaxies are composites, as would be expected for assemblages of stars. The spectrum lines are widened and weakened by the different radial velocities of the individual stars. Doppler effects in the spectra show the rotations of galaxies, and redshifts of the lines increase as the distances of the galaxies from us are greater.

18.20 Modified Classification of Galaxies

W. W. Morgan devised a modified classification of galaxies based on studies of direct photographs and spectrograms of many galaxies. The principal feature of this later classification is the increase in the evolutionary age of the stellar population as the galaxies are more highly concentrated toward their centers. At one extreme are the slightly concentrated irregular and Sc galaxies; their composite spectra contain strong hydrogen lines characteristic of blue stars of Baade's population I. These galaxies are designated as group a. At the other extreme are the highly concentrated spirals and giant ellipticals; their spectra show prominent lines of ionized calcium and molecular bands characteristic of yellow and red giant stars of population II. These galaxies are designated as group k.

Increasing central concentration of the galaxies is represented in the newer system by a succession of groups: a, af, f, fg, g, gk, and k, the lettering being in the same order as in the sequence of types of stellar spectra. The particular population group for a galaxy is followed by a capital letter denoting the form: S for normal spiral, B for barred spiral, E for elliptical, and so on. Finally, a number from 1 to 7 denotes the inclination to the plane of the sky, from flatwise, or else spherical, galaxies to edgewise presentation of flat objects. Thus the Sc spiral M 33, having rather small central concentration and younger stellar population, is classified as fS4. The Sb spiral M 31, having high central concentration and older population, is gkS5.

The population group assigned in each case is determined entirely by inspection of the central concentration of luminosity of the galaxy; but an equivalence with spectral type is expected in the average. The implication is that galaxies having higher central concentration are more advanced in evolution.

Morgan has listed 642 galaxies with their NGC numbers and their classes in the new system. His catalog, prepared mainly from the original negatives of the Palomar Sky Survey, includes practically all galaxies brighter than magnitude 13.1 north of declination −25°.

18.21 Spectral Properties of Galaxies

Spectra of galaxies are usually those of the integrated properties of the galaxies or, in the case of the nearer ones, the integrated properties of a portion of the galaxy being observed. Studies of the helium-to-hydrogen ratio in the three general types of galaxies and in specific areas of the nearer galaxies show an essentially constant ratio. This ratio is about the same as we observe in the Galaxy, namely 1 to 10.

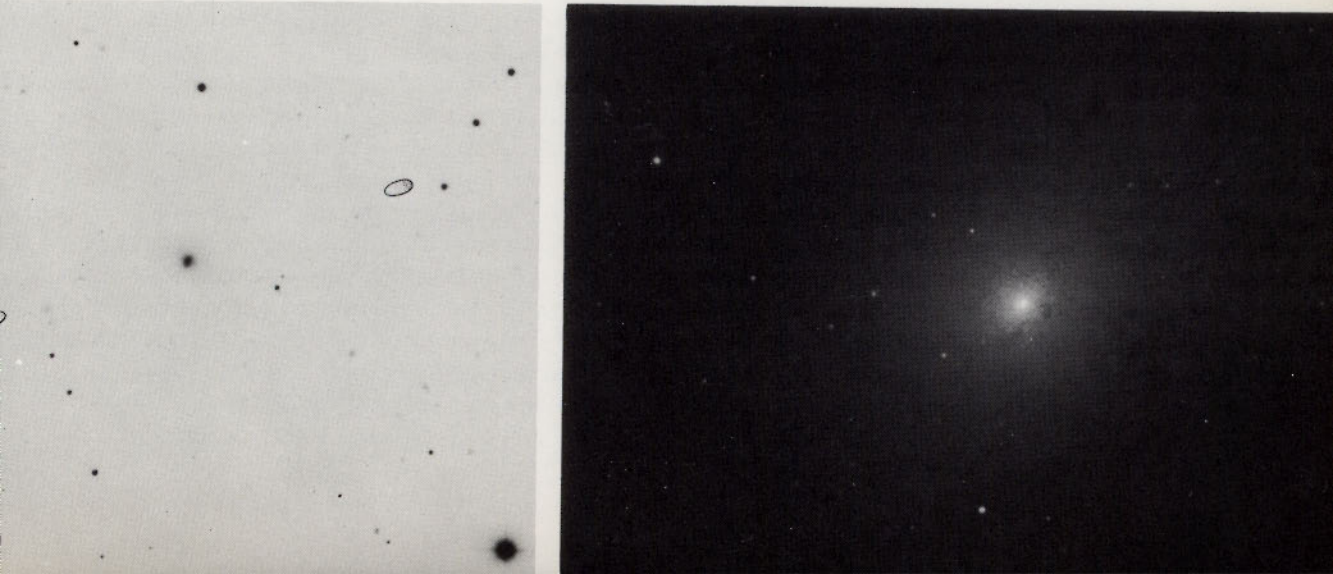
Certain galaxies seem to be metal deficient, but this does not seem to be a property of one of the general types. Elliptical galaxies resemble globular clusters physically but appear to have a normal metal abundance.

Interstellar lines appear in galaxies, the most notable ones being the forbidden λ 3727 Å line of oxygen and the λ 21 cm line of hydrogen. The forbidden oxygen line does not appear in all galaxies, but appears more frequently in the irregular galaxies than in the ellipticals. However, its appearance in the elliptical galaxies clearly shows that the type of galaxy does not strongly correlate with the general types. Some galaxies called Seyfert galaxies show forbidden high excitation lines of [NeIII] and [FeV] and emit strongly in the radio spectrum as well.

18.22 Radio Galaxies

Radio studies at various wavelengths have revealed point-like sources, many of which coincide with galaxies that can be seen on photographs. Visually, these galaxies do not appear to be very unusual, but they are clearly strong radio emitters. Many radio galaxies when studied with long baseline interferometers turn out to be actually double sources of radio emission with the optical source situated in the middle (Figure 18.22A). The radio sources always cover a much larger area than the

Figure 18.22A
Double-lobed radio emission pattern (3C33) with galaxy in the center. *(A. Moffet, California Institute of Technology)*
Figure 18.22B
NGC 224. Note the stellar nucleus. *(Photograph from the Hale Observatories)*



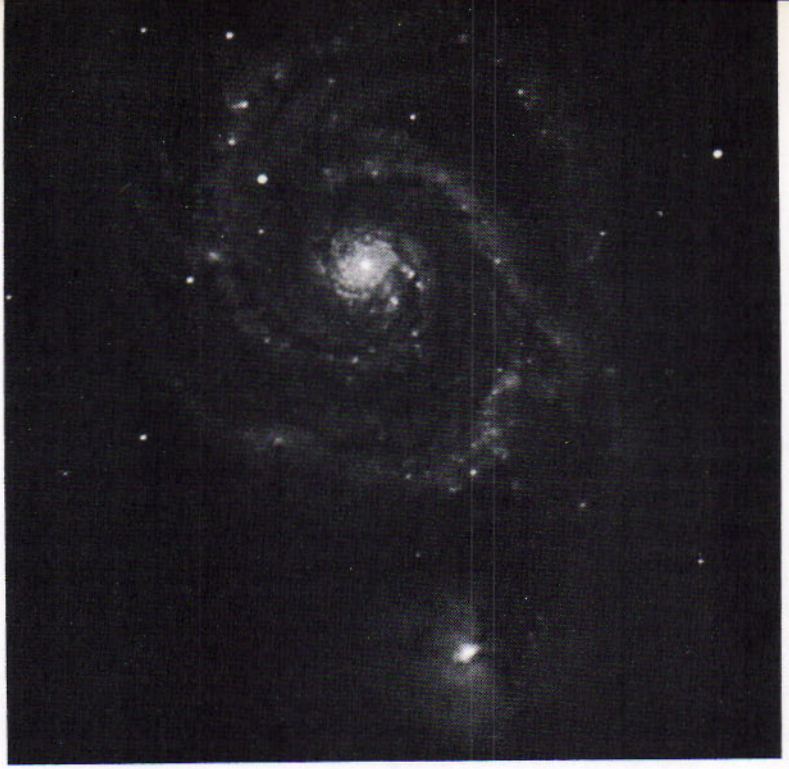


Figure 18.22C
The Spiral Galaxy M 51. Note the point-like central sources in this pair of galaxies. *(Photograph from the Hale Observatories)*

optical source, and in some radio galaxies three or even more components of radio emission are detected. The optical source is not centrally located in such systems, implying, perhaps, a non-symmetrical violent origin for the present configuration.

Some radio galaxies show long jets with a very high surface brightness. The jets are always emitting synchrotron radiation. Radio galaxies are occasionally found to be variable sources.

Generally, radio galaxies are giant elliptical galaxies; no spirals or irregular galaxies are known that get within a factor of 10 of emitting energy in the radio spectrum at the rate of radio galaxies (10^{40}–10^{45} ergs/sec). A minor exception is the Seyfert type galaxy.

C. Seyfert called attention to a class of objects that looked like spiral galaxies but were distinguished by a very sharp bright source in the nucleus of the galaxy. Spectra of these galaxies revealed forbidden emission from oxygen [OII], [OIII], neon [NeIII], and other elements. These *Seyfert galaxies,* as they are now called, were found in the direction of radio sources and almost all Seyfert galaxies are now known to be strong radio emitters. There are a large number of galaxies, called compact galaxies by F. Zwicky, that resemble to a certain extent Seyfert galaxies. They are radio quiet, that is, they have very little or no detectable radio emission but exhibit the same emission lines.

We find, upon close inspection, that the great majority of radio galaxies have very intense point sources in their central region. Galaxies

not known to be radio galaxies often have these very small, extremely bright regions right in the center of their nuclei: M 31, for example, has such a region. Perhaps the radio point sources in the center of the Milky Way are akin to these regions that are common in other galaxies.

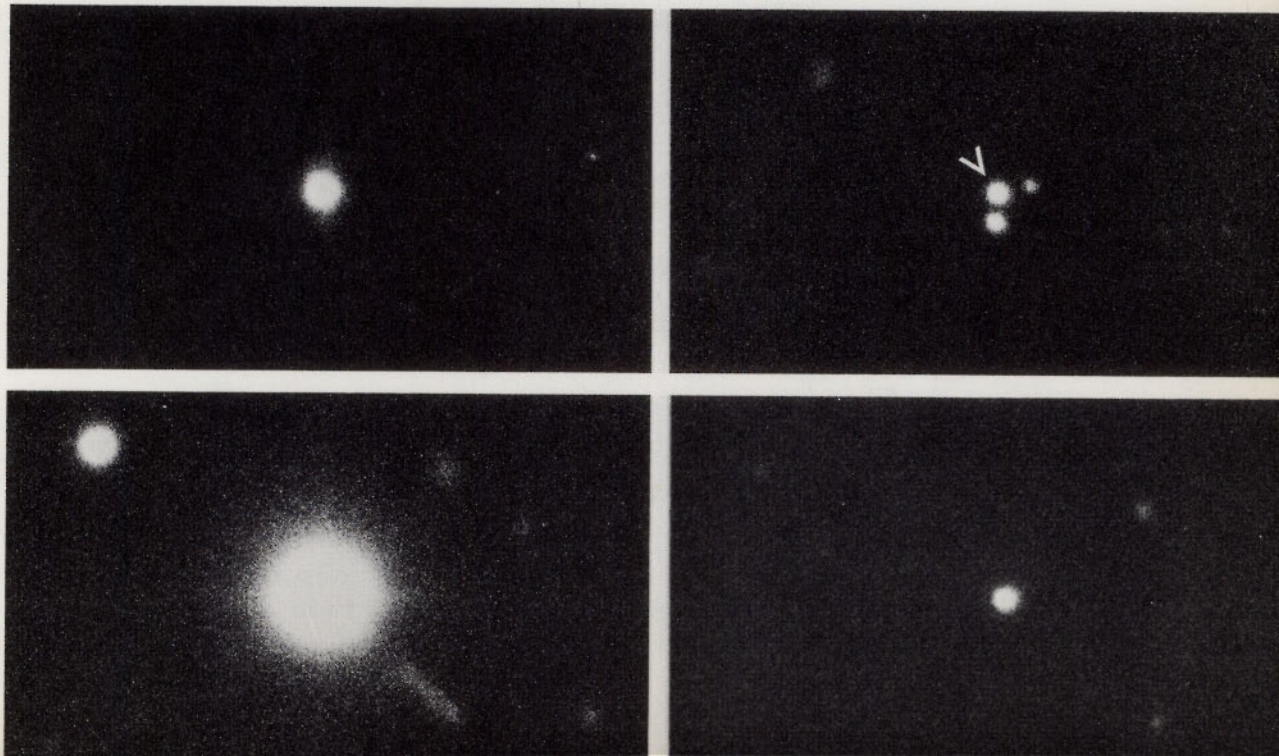
The radio power output of a spiral galaxy is on the order of 10^{38} ergs/sec (the energy emitted in the visual spectrum is on the order of 10^{44} erg/sec) or roughly the total radiated radio power of 100 supernova remnants. If the average lifetime of a supernova remnant is 50,000 years, we can explain the radio output of a normal galaxy by assuming a rate of one supernova per 500 years. Radio galaxies on the other hand have an average power output of 10^{43} ergs/sec (compared to 10^{44} ergs/sec in the visual spectrum) so we must invoke a somehow more energetic mechanism to explain radio galaxies.

The successful identification of Seyfert galaxies led to a diligent effort to identify all radio sources with an optical counterpart. This effort led to the discovery of the *quasi-stellar objects.*

18.23 Quasi-Stellar Objects

Several years after the discovery of these objects and their optical identification, M. Schmidt was able to identify the emission lines in one of them, 3C273 as due to hydrogen and magnesium by assuming a redshift of 0.16. Other identifications followed rapidly for other objects and the redshifts required to fit the emission lines were larger and larger, the

Figure 18.23A
Quasi-stellar radio sources. *(Photograph from the Hale Observatories)*



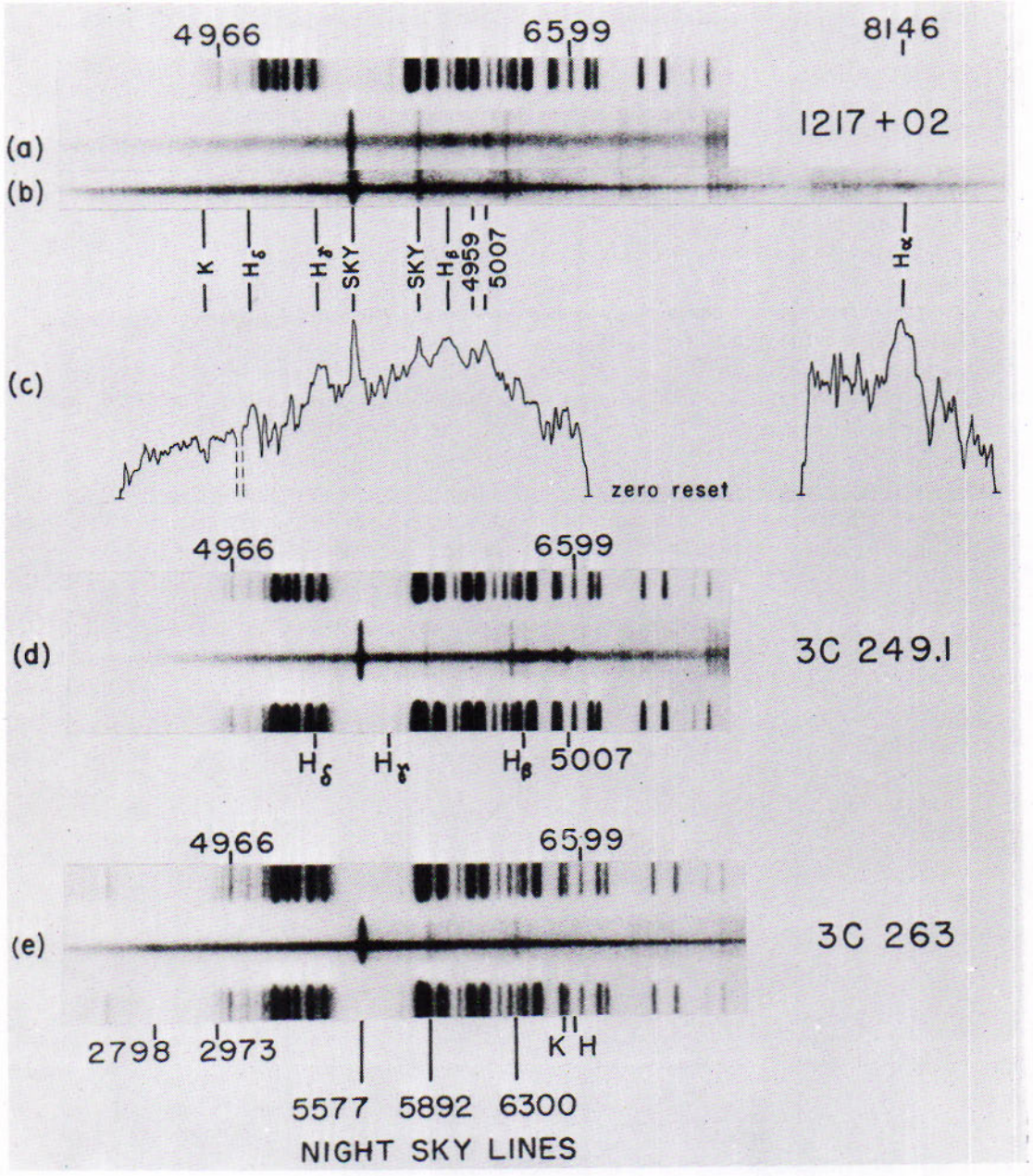


Figure 18.23B
Spectrograms of three quasi-stellar objects. *(Courtesy of K. Ford, Carnegie Institution of Washington, Department of Terrestrial Magnetism)*

largest being on the order of 2.2. Spectra of these objects reveal many of the forbidden lines present in Seyfert galaxies. The high excitation emission lines and great amount of radio energy along with the large redshifts present problems in the interpretation of the objects. We return to this subject in Chapter 19.

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

If the redshift of these objects is interpreted as it is for galaxies then these are the most distant objects known and an important tool in studying and interpreting the Universe. Only about 10% of the quasistellar objects now known are radio emitters. Those that are, have been studied by interferometric techniques using intercontinental baselines. Generally, the observed fringe pattern can be interpreted as a small diffuse halo surrounding a point source so small that it is still unresolved at spacings corresponding to a few parsecs at the assumed distance of the source.

Some optical and radio quasi-stellar objects are variables, and some of the radio variability is attributed to intergallatic scintillation which confirms the small angular diameter cited above. Studies of their redshifts by G. and E. M. Burbidge, K. Ford, M. Schmidt, V. Rubin and many others indicate that the redshift measured from absorption lines may not agree with that of the emission lines. In fact, G. Burbidge has called attention to the fact that a redshift of 1.95 appears quite frequently in those objects that have absorption spectra. He feels that this is an intrinsic property of quasi-stellar objects with large redshifts.

18.24 Rotations Shown by the Spectra

The flattened forms of regular galaxies suggest their rotations. Definite evidence of rotation is found in the spectra of spiral galaxies where the equators are presented nearly edgewise to us. When the slit of the spectroscope is placed along the major axis of the inclined spiral, the spectrum lines slant (Fig. 18.24) at an angle from the vertical that depends on the speed of the rotation. It is the same Doppler effect that appears in the spectrum of a rotating planet (Fig. 8.30A).

Since the pioneer work of V. M. Slipher, the rotations of several spirals have been studied in this way, particularly by N. U. Mayall and associates at Lick Observatory. The inner part of a spiral rotates like a solid, all in the same period, showing the uniform distribution of material there. In the outer parts the period increases with distance from the center, resembling the rotation of our own spiral galaxy (17.17).

Determination of rotational velocities makes it possible to compute the mass of the galaxy in question. The mass of a number of galaxies have been determined by E. M. Burbidge who found that spirals do not differ very greatly in mass. There is, however, a relation between the ratio of the mass of a galaxy and its luminosity, much as with stars. Elliptical galaxies have a larger value for the ratio of their mass to their luminosity than do the spirals, which in turn have a larger value than the irregulars.

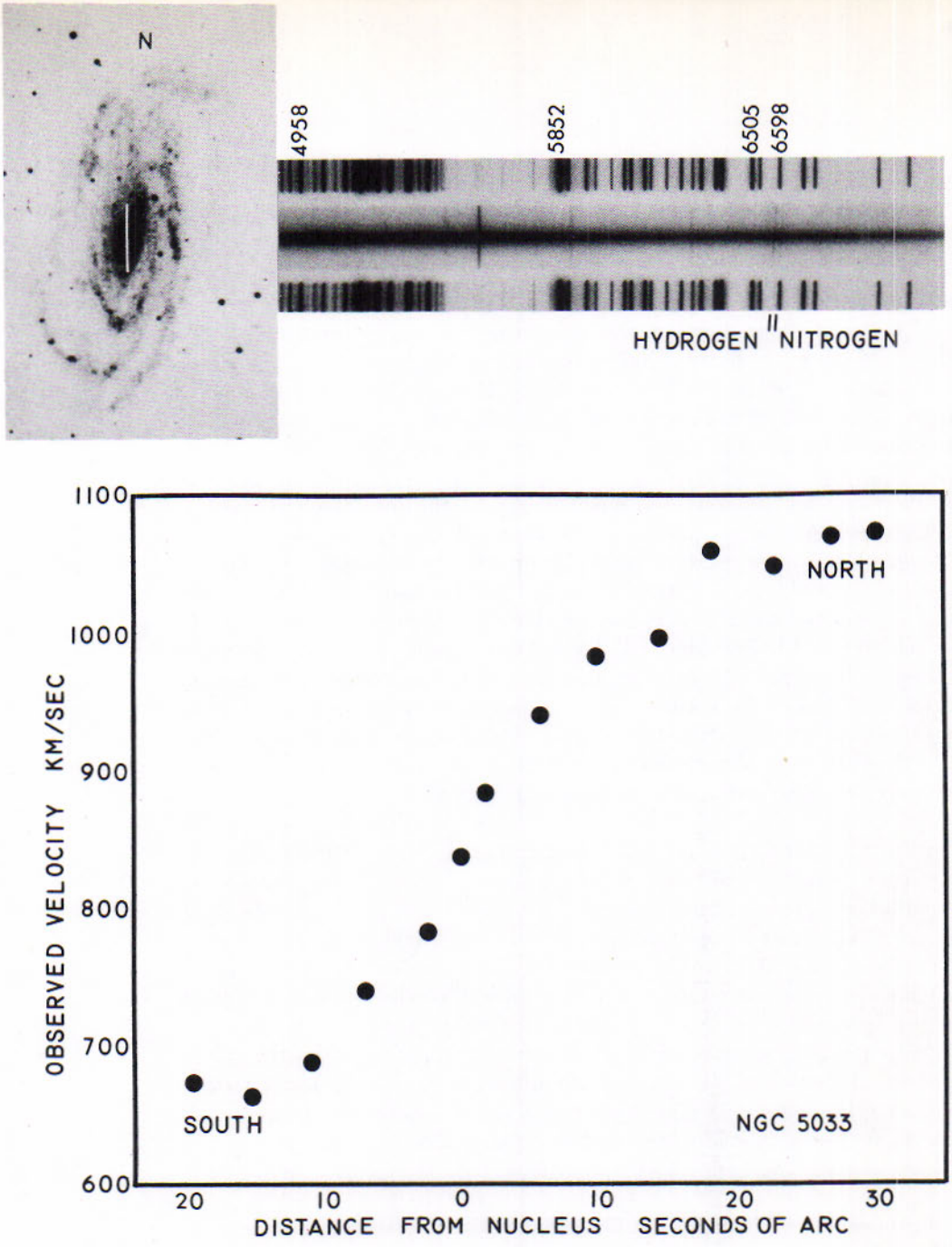
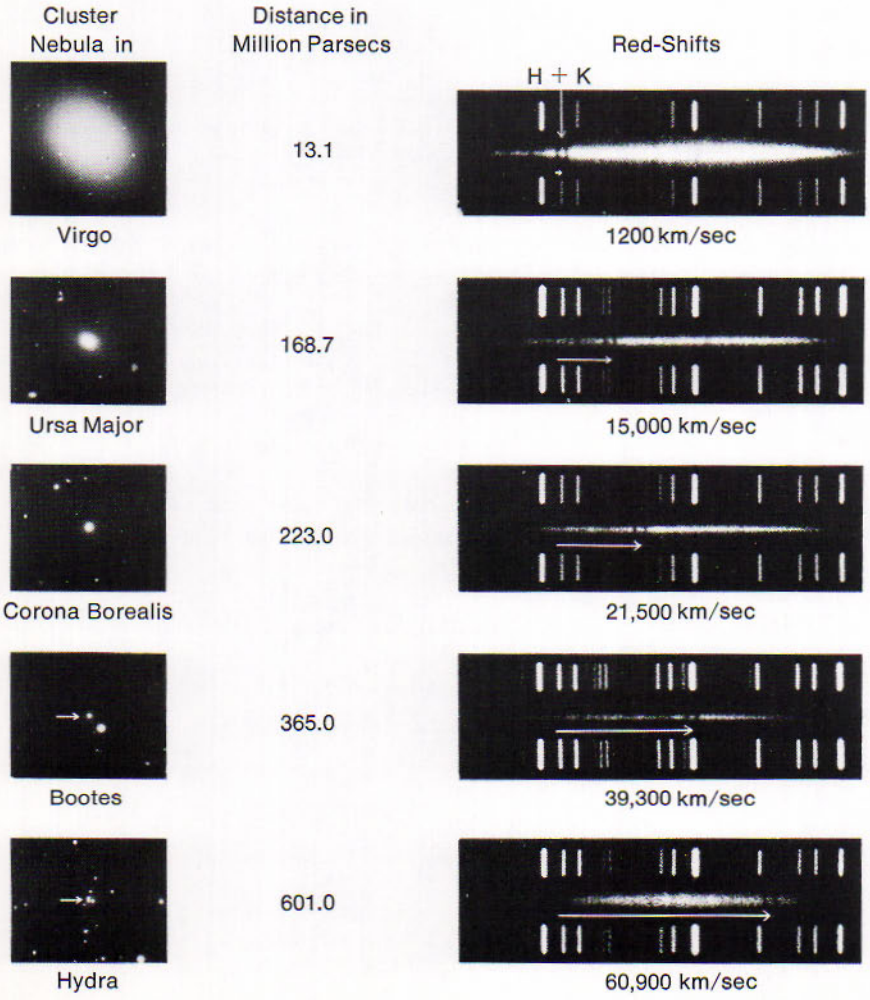


Figure 18.24
Spectrum of Nucleus of NGC 5033, showing emission lines of hydrogen and nitrogen. The white line in the center of the galaxy shows the location of the slit of the spectrograph. The emission lines are inclined, due to the rotation of the galaxy. The measured velocities, as a function of distance from the center, are plotted in the graph. As the galaxy rotates, its north end recedes from the observer, while the south end approaches. *(Spectrum and figure courtesy of the Carnegie Institution of Washington, Department of Terrestrial Magnetism; photograph copyright by the National Geographic Society—Palomar Observatory Sky Survey)*

18.25 Redshifts in the Spectra

E. Hubble, in 1929, announced a remarkable relation between the radial velocities and distances of galaxies. With appropriate corrections to the observed values, he discovered that the redshifts of the spectrum lines increase linearly as the distances of the galaxies are greater. This linear relation persisted when the inquiry was later extended to what was then the limiting distance for the 100-inch telescope. More recently, M. L. Humason extended the distance range of the observations toward the limit for the 200-inch telescope.

Figure 18.25
Redshifts in the spectra of galaxies. The arrows indicate the shifted H and K lines. *(Photograph from the Hale Observatories)*



Humason, Mayall, and Sandage, in 1956, reported the progress on the program to that time. The redshifts had then been observed for galaxies of 26 clusters and for several hundred field galaxies. The clusters range from the Virgo cluster relatively nearby to the remote Hydra cluster. The redshifts of the spectrum lines considered as Doppler effects correspond to velocities of recession varying from 1,100 to 61,000 km/sec, or to one fifth the speed of light. The rate of increase in the velocities, then given as 180 km/sec per 1 million parsecs, or about 35 miles a second per 1 million light years, has been reduced to a little more than half of these values by a revision of the distance scale.

The problem of the expanding universe is not completely solved. Some approaches to the solution by observation and theory are noted in the following chapter.

REVIEW QUESTIONS

1. Describe the basic features of the three major types of galaxies.
2. Elliptical galaxies resemble the globular clusters of the Milky Way. Why are they not globular clusters ejected from spiral galaxies?
3. M31 is often called a "prototype" of the Milky Way. Why?
4. Why are the Magellanic Clouds so important to the morphology of the Universe?
5. The Magellanic Clouds have a distance modulus of about +18. If your telescope allows you to observe to magnitude +21, what type of objects should you be able to see?
6. Why do we feel that the present condition of M82 is the result of a catastrophic event?
7. List some of the distance indicators for determining extra galactic distances. There are radio indicators, have you included them?
8. How bright would a star in the center of the Galaxy have to be to be seen from the earth?
9. According to Abell's classification there are only two classes of clusters. What are they? Have we given any evidence to indicate otherwise?
10. Give three arguments why quasi stellar objects are point sources to present day observing techniques.
11. All faint galaxies have redshifts that correlate with their brightness. What does this imply?
12. Discuss the rotation observed in spirals which are seen nearly edgewise. How is the curve used?

REFERENCES

Coudere, Paul, *The Wider Universe,* New York: Harper and Brothers, 1960
Shapley, Harlow, *The Inner Metagalaxy,* New Haven: Yale Univ. Press, 1957

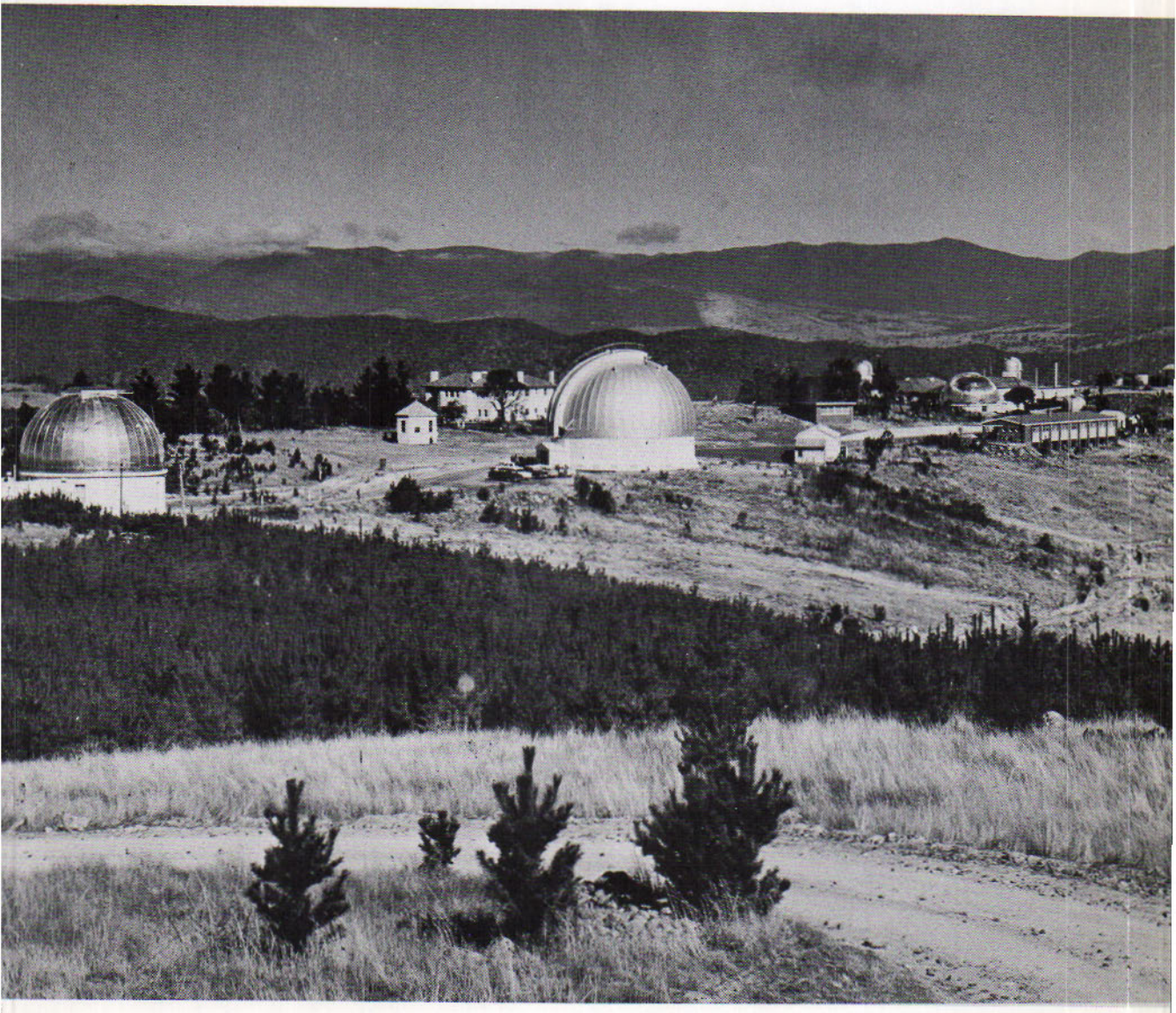
FOR FURTHER STUDY

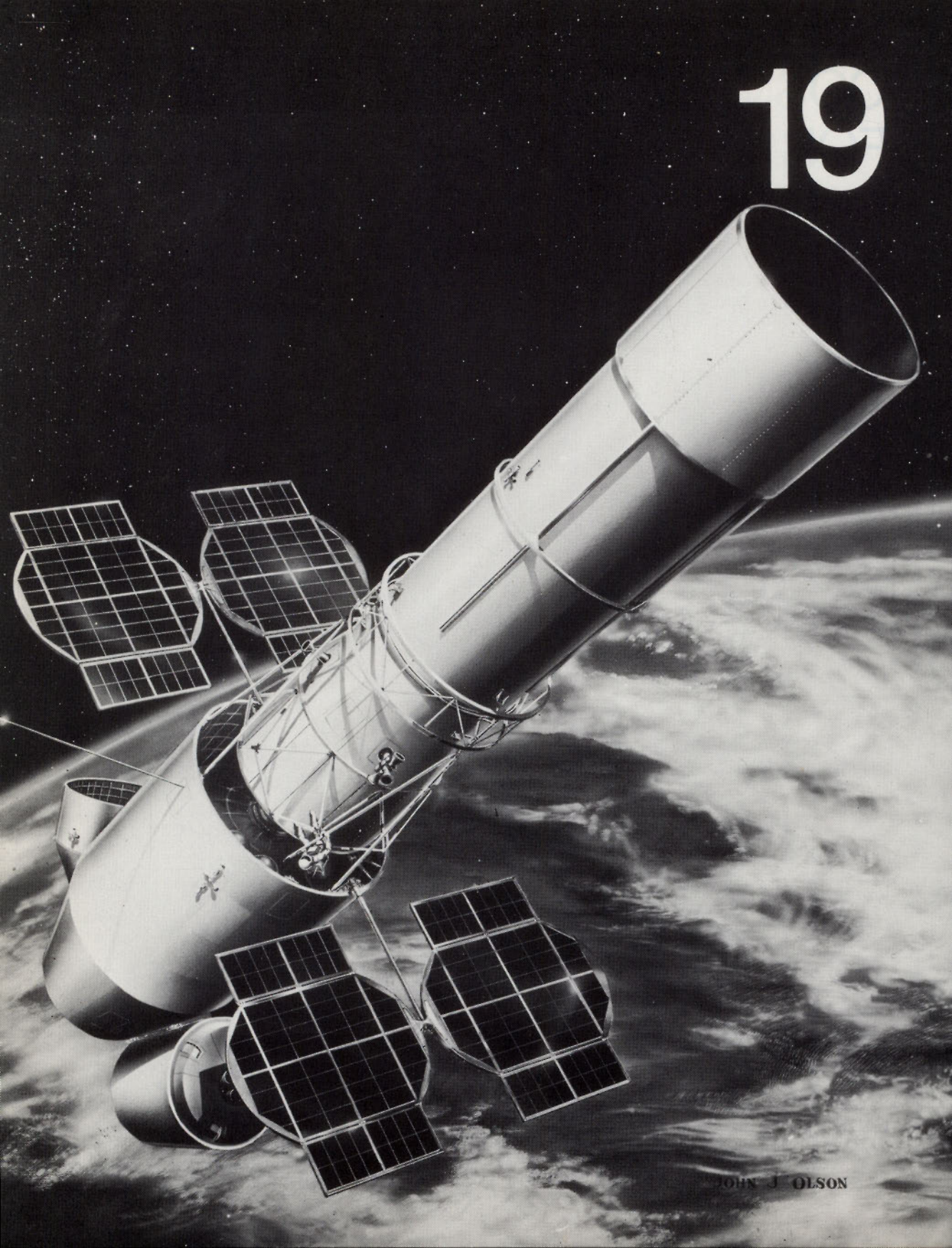
Goldberg, Leo, ed., *Annual Review of Astronomy and Astrophysics,* Vols. 1, 3, 4, 6, 7, Annual Reviews, Inc., Palo Alto, California.

Woltjer, Lodewijk, ed., *Interstellar Matter in Galaxies,* New York: W. A. Benjamin, 1962.

——, *Galaxies and the Universe,* New York: Columbia Univ. Press, 1968.

Panoramic view of the Mount Stromlo Observatory complex, Australia. *(Mount Stromlo Observatory photograph)*




19
JOHN J OLSON

COSMOLOGY—DIRECTIONS FOR THE FUTURE

OBSERVATIONS AND THE EVOLUTION OF GALAXIES—COSMOLOGICAL PRINCIPLE AND PROPOSED THEORIES — THE AGE OF THE UNIVERSE AND LIFE IN THE UNIVERSE — THE COURSE OF ASTRONOMY — THE PHYSICAL UNIVERSE

Through the previous chapters we have presented the *observational* material of astronomy. We will now try to organize the results of all the observations in order to present a unified world picture. Once all the data from the observations is collected, it is possible to present some "world views" or *cosmologies* to explain what has been observed. The next step consists of studying the different tests that support one world view or another. Finally, we will look at the fronts upon which astronomy has been expanding, and at what can we look for in the future.

Artist's conception of a large manned orbiting astronomical observatory. *(NASA)*

OBSERVATIONS AND THE EVOLUTION OF GALAXIES

The progressive redshifts in the spectra of galaxies argue for an expanding Universe. The Universe appears to be devoid of intergalactic material and the three major types of galaxies seem to have the same age.

19.1 The Expanding Universe

The observed redshifts in distant galaxies and clusters of galaxies are considered to be due to the Doppler effect. Ever since V. M. Slipher first observed redshifts in the galaxies the redshift observations have formed the single most fundamental observation to be explained by any theory of the Universe. Considered as Doppler effects, the redshifts of the spectrum lines show that the universe of galaxies is expanding in all directions at a rate that increases very nearly in direct proportion to the distance from us. The systematic expansion, however, does not operate within the galaxies themselves or within the groups and clusters.

Hubble's extensive work established that the rate of recession (v) is proportional to a constant (H) times the distance (r). An additive constant may be involved; some workers include such a constant. If the velocity of recession is given in kilometers per second and H is in kilometers per second per megaparsec (a million parsecs) then the distance is in megaparsecs. An often used value for H is 100km/sec/mpc though the best value may turn out to be as small as 75. S. van den Bergh suggests 95km/sec/mpc.

$v = Hr$

H the Hubble constant

The velocity used in the Hubble relation is the true velocity as derived from the special theory of relativity and not that from the first order Doppler effect given earlier. Einstein's theory of relativity has been shown to predict high velocity effects in the laboratory, so we expect it to hold generally. The very high valued redshifts found for the quasi-stellar objects would imply velocities of more than twice the speed of light, whereas the theory of relativity yields values that continue the uniform expansion in a uniform way.

$$\frac{\lambda'}{\lambda} = \frac{1-(v/c)\cos\theta}{(1-v^2/c^2)^{1/2}}$$

The basic formula from the special theory of relativity relates the observed wavelength (λ') to the wavelength at rest (λ) by means of a relationship involving the velocity of the object (v), the velocity of light

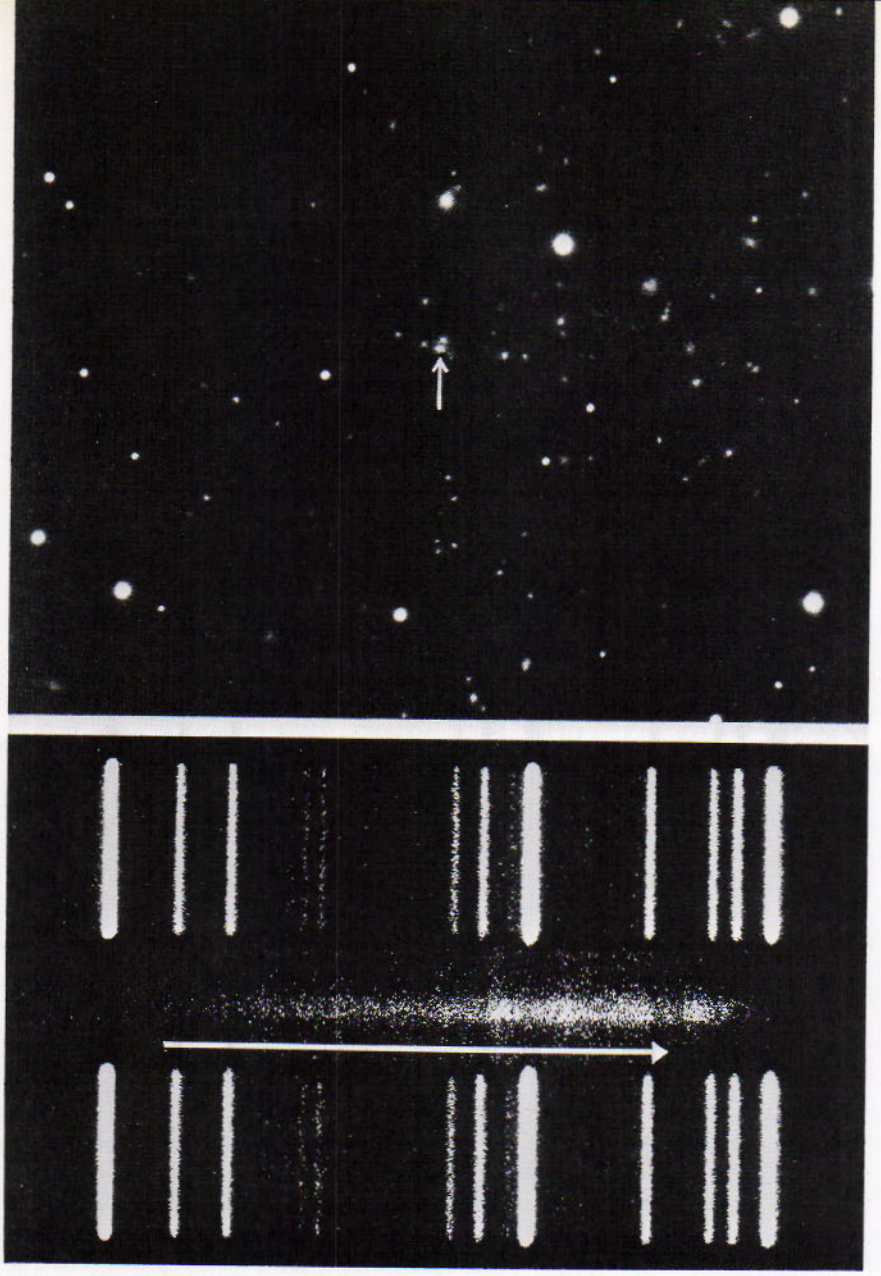


Figure 19.1A
Redshift of the H and K lines in a galaxy in the Hydra cluster. *(Photograph from the Hale Observatories)*

(c) and the angle (θ) the object is moving with respect to the line of sight. This can easily be put into the customary $\frac{\Delta\lambda}{\lambda}$ form in the radial direction, for there $\cos\theta = -1$. Thus a redshift of 2 is quite possible and corresponds to a velocity of recession of eight tenths the velocity of light.

We are rather confident that the redshifts seen in the spectra of galaxies are due to the Doppler effect. Larger redshifts are observed in the spectra of quasi-stellar objects, so large, in fact, that the original identifications took several years (18.23). Quasi-stellar objects with "small" redshifts ($\Delta\lambda/\lambda \approx 0.15$) appear to have a wispy structure around them on direct

$$\frac{\Delta\lambda}{\lambda} = \left(\frac{1+{}^{v}/_{c}}{1-{}^{v}/_{c}}\right)^{1/2} - 1$$

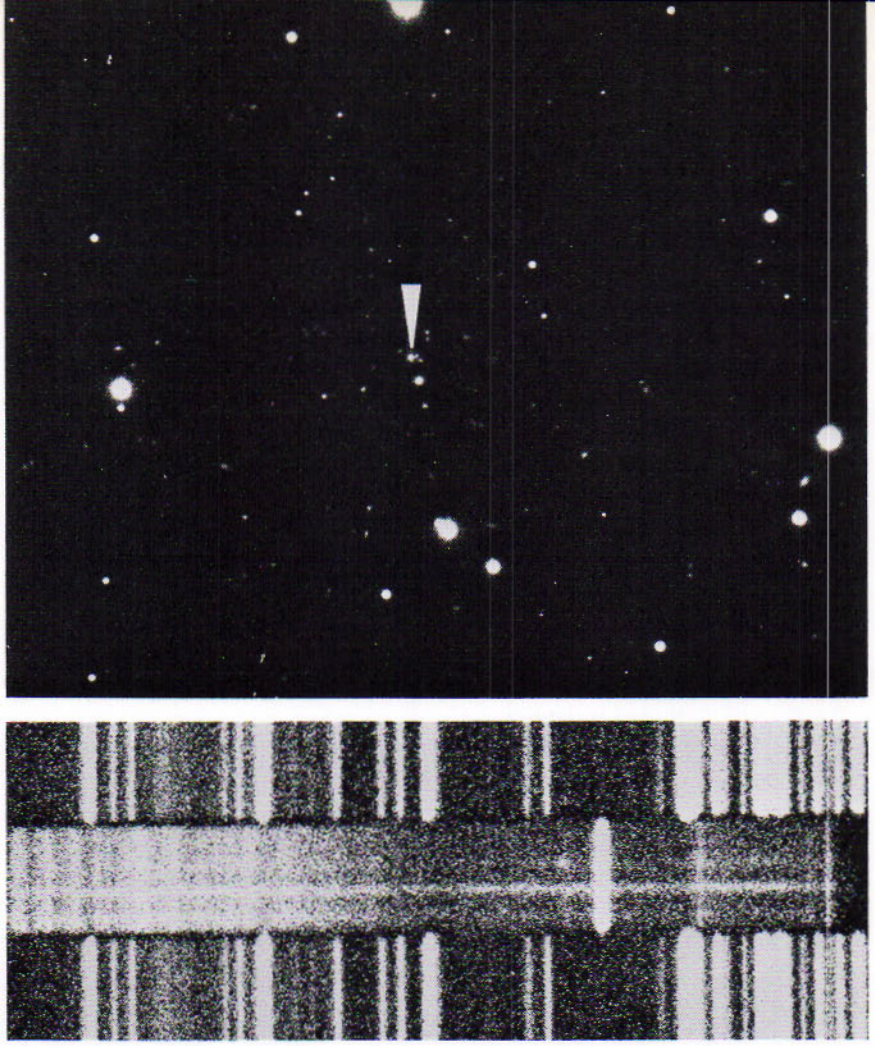


Figure 19.1B
Redshift of 3C295. Bright lines across the spectrum are night sky lines. *(Photograph from the Hale Observatories)*

photographs, those with large redshifts are point-like. The essential question is whether these objects are at the extreme distances implied by the redshift or whether the redshift is due to some other factor, such as light photons getting "tired", intense gravitation, or something not yet determined.

If we assume that the quasi-stellar objects form a homogeneous group, then we should expect a reasonable correlation between their magnitudes and their redshifts. The diagram scatter is extreme. Plotting radio magnitudes against redshift gives a diagram with even greater scatter. Such plots led some astronomers to consider that the quasi-stellar objects are not at great distances from us. N. Terrell has proposed that they are objects ejected from the Galaxy whereas Arp and others have proposed that they are ejected from other galaxies. Since the true nature of these objects is still not known, at the present at least they do not contribute to determining our choice of a world view.

19.2 Progressive Reddening of Galaxies

The Doppler effect of the receding galaxies displaces their spectral energy curves (10.5) toward longer wavelengths and therefore causes the light of these galaxies to become redder as their velocities of recession are greater. This progressive reddening, studied by W. A. Baum, Sandage and others, provides us with another means of observing the expansion of the Universe.

A diagram of the velocities so far determined, primarily with the 200-inch telescope, is shown in Figure 19.2 for galaxies in twenty-four clusters. Here the logarithm of the recession velocity for each cluster is plotted with respect to the visual magnitude of the galaxies, which is a function of the distance of the cluster, after being corrected for instrumental effects, galactic absorption and the reddening effect.

Cluster 1410, the most remote of the twenty-four, is represented by the single galaxy corresponding to the radio source 3C295. Baum's value for the velocity of recession for this galaxy is 44 percent of the speed of light,

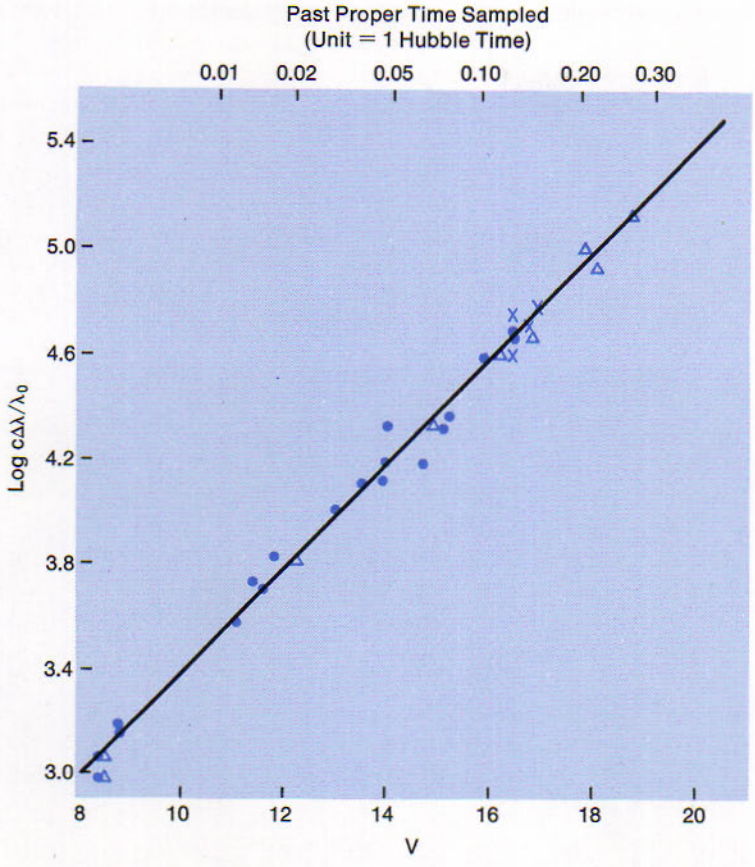


Figure 19.2
Corrected magnitude (V) plotted against velocity for the brightest cluster galaxy. *(Diagram as presented by A. Sandage in* Galaxies and the Universe)

or about 132,000 km/sec. The points for the 24 clusters fall very nearly on a straight line; but their positions are preliminary, requiring certain further corrections not yet evaluated. Note the time scale given at the top of the diagram.

Quasi-stellar objects whose large redshifts displace their ultraviolet continuum into the visible region show an ultraviolet excess. This has been thought to be characteristic of quasi-stellar objects. Observations from the Orbiting Astronomical Observatory studied by A. Code and his colleagues indicate that all galaxies seem to possess an ultraviolet excess. If this is so then the distances assigned to remote galaxies will have to be revised. In addition, this may form the first link indicating that the quasi-stellar objects are at truly cosmological distances.

19.3 The Evolution of Galaxies

Before looking at the possible theories for the Universe we should consider the question of evolving galaxies. The Hubble diagram leads one at first to think that galaxies evolve from irregulars to ellipticals. It is not obvious why this should be so, so some astronomers prefer to consider that galaxies evolve from ellipticals to irregulars.

M. S. Roberts states that the Universe is too young for any galaxy of a given type to have evolved into one of another type. He has shown that the best estimates of the age of various types of galaxies are all on the order of the age of the Universe and in this time, stellar evolution could not produce the large number of stars on the lower main sequence in the Sa spirals and the ellipticals. He believes that the major morphological types of galaxies have existed essentially as they are. Certainly, out to the limit of our ability to distinguish galaxies, i.e. back in time, the relative numbers of the various types do not change greatly.

Figure 19.3
Spiral galaxy NGC 7217. A beautiful face-on view of an *Sa* galaxy. *(Photograph from the Hale Observatories)*

COSMOLOGICAL PRINCIPLE AND PROPOSED THEORIES

With the basic observed facts in hand, we are prepared to propose theories. Two fundamental models are current to explain the observations—a relativistic model and a steady state model. Present observations seem to favor the relativistic model.

19.4 The Cosmological Principle

The basic assumption or ground rule under which scientists build models of the Universe is called the *cosmological principle.* This principle states simply that any observer at rest with his local part of the Universe should see the Universe the same in all directions and that the Universe looks the same to all such observers wherever their location in the Universe. There may, of course, be local anomalies but averaged over significant portions of the Universe, the principle must hold.

An extension of the cosmological principle includes the dimension of time. Not only does the Universe look everywhere the same to all observers, but it must look the same at any time. This extended principle is referred to as the *perfect cosmological principle.*

19.5 The Steady-State Model

A model for the Universe proposed by H. Bondi and T. Gold has led to a great deal of speculation and controversy. Using the perfect cosmological principle Bondi and Gold explain the Universe by assuming a steady state Universe where the density remains constant. Since the redshifts of galaxies show the Universe to be expanding, they require that matter be created at a very small rate to keep the density constant. The geometry of such a model is the familiar Euclidean geometry.

Critical to the model is the creation of matter. This has not been observed.

19.6 The Relativistic Model

The general theory of relativity coupled with the cosmological principle produces a model of the Universe. Since three fundamental constants appear in the equations numerous variations of the model are possible and it remains for observations to determine the values of these constants and hence the correct model.

One of the constants is the geometrical constant. It can have values such that space is curved and finite, curved and infinite, or flat and infinite (similar to the geometry of the steady state model). The second constant is the cosmological constant. This constant can be positive or

negative; if it is positive it has the effect of opposing gravity over large distances, if it is negative it reinforces gravity over large distances (where "large distances" means on the order of the radius of the visible Universe). The third constant is an assumption that the product of universal gravity, the density of the Universe and the cube of its radius is proportional to a constant. A derived constant, called the acceleration parameter, is an observable quantity and sets limiting values on the cosmological constant, hence considerable effort is being expended to determine it.

The relativistic model predicts, essentially, three possibilities: 1) a Universe that starts at some unique time in the past and goes to infinity, 2) a Universe similar to the first that goes to infinity at an accelerating rate, or 3) a Universe similar to the first that is decelerating; it goes to some limiting radius and falls back upon itself. These possibilities are known as the Big Bang Universe, the Pulsating Universe, etc.

Critical to the relativistic models are the density of matter in the Universe and the deceleration parameter. Present observations suggest that the expansion is slowing down, as we see later.

The "classic" relativity models have certain objections in detail. To overcome some of the objections, C. Brans and R. H. Dicke have proposed a modification which we will not explore here. All of the models including the steady state model are subject to delicate observational tests. Models that are proposed but untestable are discarded.

19.7 Critical Observations

The steady state model requires every part of the universe to look the same regardless of time. Counts of galaxies offer a test, but the visible galaxies involve corrections that confuse the statistics. Extrapolating from visible radio galaxies to radio galaxies beyond the range of optical telescopes leads to the conclusion that there were more radio galaxies earlier

Figure 19.7
NGC 5866, an interesting galaxy with bright wings. *(Photograph from the Hale Observatories)*

in time than there are now. Since time is removed from the steady state model this argues heavily against the model. We have already noted the problem of creating matter from nothing.

The relativity models, i.e. those that have a time when the Universe was a primeval ball of energy, predict an isotropic remnant background field with a temperature around 3°K. Recent observations, first conducted by A. A. Penzias and R. W. Wilson, give a value of 2.8°K for the temperature of the cooled fireball—remarkably close to the predicted value. By the same token, the relativity models predict a larger amount of helium in the Universe than is observed (the Brans-Dicke model does not).

All current models predict a mass for the Universe far in excess of that observed.

19.8 Intergalactic Material

It has been argued that the presence of gas filaments as seen drawn out by interacting galaxies demonstrates the existence of material in intergalactic space. One would expect that the preponderance of intergalactic material should be neutral hydrogen or, allowing that conditions are extreme, it must be electrons and ions. Careful experiments have been performed that seem to give an extremely low value for the intergalactic material and, in one case, even for neutral hydrogen in the Virgo cluster.

Precise observations by R. J. Allen failed to detect any λ 21cm absorption due to the Virgo cluster. Observations looking for phase delays between the radio and optical variations in radio galaxies also yield an extremely low value for the number of free electrons in the line of sight. The significance of these observations is that if the expansion of the Universe is decelerating, as some investigations seem to indicate, then a gravitational mass some 30 times greater than that attributed to all the galaxies is required. It has been presumed that this "missing mass" has been the intergalactic material, if this material does not exist then the decelerating force is not gravitational.

If the quasi-stellar objects are at a great distance then another test for intergalactic material is available. Singly ionized magnesium should show as an absorption line in the spectra of these objects. Shklovsky could not detect such an absorption and sets the upper limit of ionized magnesium at 10^{-12} ions per cubic centimeter. He also sets the upper limit of intergalactic hydrogen at 10^{-7}; both values being far too low to account for any missing mass in the Universe.

19.9 The Acceleration of the Universe

The relativistic models can be sorted out by the various values of the *acceleration constant* (q_0) which is an observable quantity as pointed out in 19.6. If this constant is negative an accelerating expansion is indicated, if it is zero then a constant rate of expansion holds while if q_0 is positive the Universe is decelerating. Figure 19.9 is a diagram for testing for the sign of q_0.

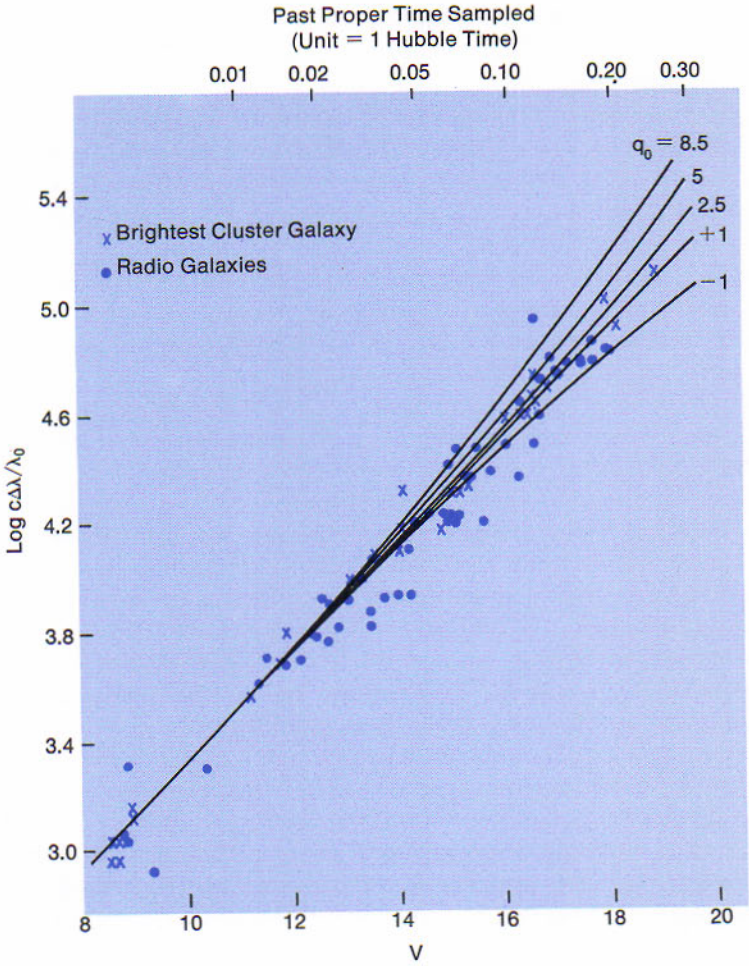


Figure 19.9
Magnitude-velocity diagram similar to the one on Fig. 19.2. *(Diagram by A. Sandage in* Galaxies and the Universe*)*

In Figure 19.9 we have replotted Figure 19.2 as crosses and plotted radio galaxies as dots. The diagram is due to Sandage and has the various expansion lines for various values of the acceleration parameter plotted on it. While the scatter is large the diagram seems to indicate that q_0 is positive and hence that the Universe is decelerating. It also demonstrates that radio galaxies have the same visual luminosity as their non-radio counterparts.

While we do not know which model of the Universe is correct we feel that we are very close to deciding. We should however observe some caution; A. Sandage puts it in the following way: "Although it appears that the problem is now on its last asymptotic approach toward 'certainty', men of all previous ages undoubtedly felt the same".

How do we continue to approach the problem? First we must set up our theoretical model, which must explain the Universe on a grand scale,

at the same time it must also predict what we observe in our "part" of the Universe, then our Local Group, then our Galaxy, then the stars and finally the solar system. Chipping away at this problem observationally and theoretically is the task for astronomy and is its future.

THE AGE OF THE UNIVERSE AND LIFE IN THE UNIVERSE

One fundamental question is, assuming the Universe has an age—what is it? A second fundamental question is, does life recognizable to man exist elsewhere in the Universe?

19.10 The Age of the Universe

Assuming the relativistic models are nearly correct for describing the Universe we can assign a time to the interval since its beginning. Sandage sets the following framework in which we operate:

Time since the creation of the hydrogen atoms is greater than or equal to the beginning of the expansion of the Universe; is greater than or equal to the condensation of the first galaxies; is greater than or equal to the formation of the oldest stars in the Galaxy; is greater than or equal to the manufacture of the heavy elements; is greater than or equal to the contraction of the Sun and isolation of the Solar System, is greater than or equal to the age of the crust of the earth.

Globular clusters contain the oldest stars and indicate an age somewhat greater than 10^{10} years. If we assume that all of the heavy elements were made immediately at some point in evolution we have an age for their origin of almost 10^{10} years. If we assume the heaviest elements are being synthesized we arrive at an age very nearly 10^{10} years. Finally the expansion of the Universe yields a time on the order of 10^{10} years. Based on our present knowledge, we conclude that the Universe is about 10 billion years old.

19.11 Life in the Universe

Ten billion years is a long time compared to the evolutionary time for life on earth, which is measured in millions of years only. The conditions for life, as we know it, are simply a *stable source of energy*—stable over several billion years—and a carbon-oxygen chemistry.

Stable stars, such as the sun, exist in great numbers in the Galaxy

(in obviously greater numbers in the Universe). Proto-suns that form in the absence of a large companion of similar mass may have planets forming at the same time and form a stable system. If the nebula from which this formed contained carbon and oxygen there is a chance for life.

Microwave observations have recently detected water vapor (H_2O), hydroxyl molecules (OH), and formaldehyde (H_2CO, found with normal carbon and isotopic carbon), an organic molecule at various temperatures in discrete clouds. Along with the long recognized cyanogen molecule (CN) there seems to be an abundance of carbon and oxygen for the life chemistry that we know.

The formation of *complex molecules* in interstellar space is an interesting problem. Chemists have successfully formed exotic molecules in ices by irradiating them. Perhaps various molecules are formed in the interstellar ice grains and the grains are then exploded by cosmic rays releasing the molecules to space. While this process takes an extremely long time the dimensions of space are large so that the sum total is effective. Considerable laboratory work is in progress on these processes and more is needed to solve this interesting question. The answers may hold the key to life in the universe.

Considering what we have just presented, we would be more than presumptuous if we asserted that life on planet earth was unique in the Galaxy.

THE COURSE OF ASTRONOMY

The next few decades will see an ever-expanding challenge in astronomy. Certainly the last decade has turned up surprises unparalleled since the early part of this century when V. M. Slipher discovered the redshifts in galaxies and H. Shapley presented the galacto-centric point of view. Unique instrumentation will play the major role.

19.12 Optical Telescopes

We are presently seeing an extremely rapid growth in effective light collecting area of optical telescopes. Some forty telescopes in the 36–60 inch aperture class have been installed in the United States alone. These instruments aided by image tubes become extremely effective research tools.

Larger telescopes are being built as well. Most large telescopes are of a general purpose nature because of their cost. Fortunately, a fair number of large telescopes having apertures close to 240 centimeters are under construction in Chile and Australia (4.23). There is a great disparity in observational capability between the northern and southern hemispheres. A requirement for numerous moderate-sized instruments in the southern hemisphere will still be there even after the larger telescopes are built.

We feel, however, that a trend to the special purpose telescope will develop. Already we have a large aperture reflector designed for and dedicated to astrometric studies. This instrument designed and built for the U. S. Naval Observatory can be used for other applications, but it is far from optimum for them. We have also seen telescopes called "collectors" built for the infrared sector of the spectrum and used as photometric telescopes. The images are poor, but in photometry it is only necessary to collect the greatest possible *quantity* of light, and image quality is only of secondary importance. We have had a telescope built where the tube can be rotated and hence suited especially for polarization measures.

In the future we expect to see large alt-azimuth mounted telescopes applied where the rotating field is not a factor. Modern computer technology makes such telescopes feasible. Indeed, we are now seeing computers controlling an ever increasing number of conventional telescopes just as they have controlled radio telescopes for a number of years. We will see many more space telescopes and we discuss this later (19.15).

Finally, we expect to see very large telescopes designed to study special objects. As an example, a large telescope designed to move only over the meridian, say $\pm 1^h$, and between $-65°$ and $-75°$ is all that is required to study the Magellanic Clouds. The limited motions required allow a very large, stable instrument to be constructed.

In instrumentation for telescopes there will be more on line data reduction, more general use of image tubes, the general use of Fourier transform spectroscopy and the application of Josephson detectors. A long term but ultimately successful effort will be made to make two widely separated telescopes into a long baseline optical interferometer. This will be accomplished by using the coherent beam from a laser as the connecting "cable."

19.13 Radio Astronomy

The present state of world wide radio astronomy is good. The very long baseline (intercontinental) interferometers yield resolutions approaching the theoretical limit and the use of aperture synthesis gives enormous effective collecting areas. In the United States the over-all picture in 1970 is not so good. However, this should change.

We expect several large steerable telescopes to be built with surfaces yielding good definition into the submillimeter region. We also expect

the United States to undertake a huge array of steerable telescopes presently referred to as the Very Large Array (VLA). The system is technically feasible and would serve as an aperture synthesis telescope as well as an interferometer.

Special purpose radar telescopes for probing the far planets are in the offing. Special instrumentation with increasing sensitivity for detection and studying interstellar molecules is also under development. Finally, we will even see an ultralong-baseline interferometer using the moon for the second leg.

19.14 Other Branches of Astronomy

We do not wish to slight developments in any branch of astronomy, but can cite here only briefly a few additional areas of development. X-ray astronomy is, by its nature, strongly tied to a strong space program. We expect to see better X-ray telescopes developed and flown. We even hope to see a γ-ray "telescope" emplaced on the moon.

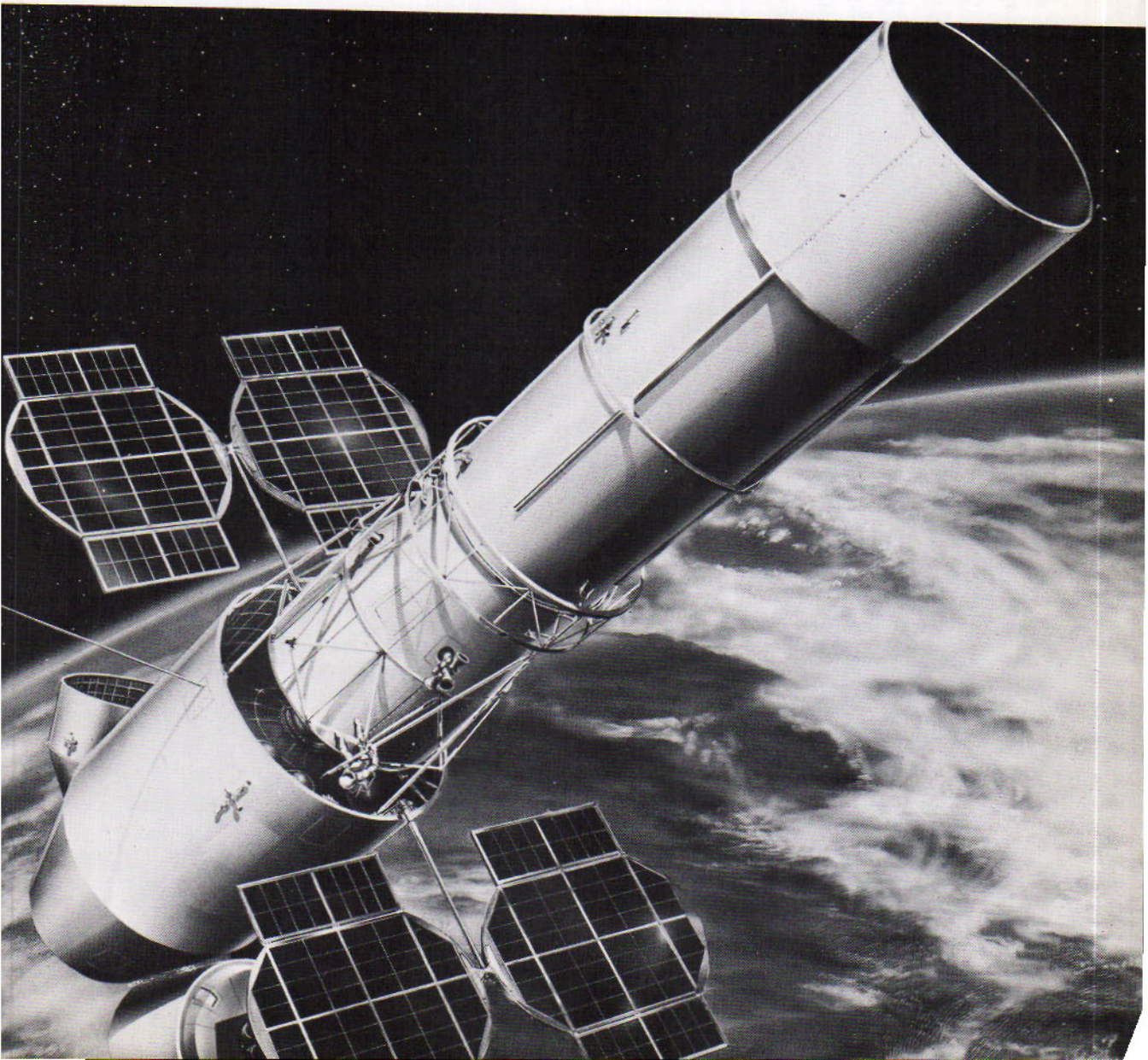
Theoretical branches will continue to grow as larger and faster computers are developed. Indeed, the modern computer has been a major breakthrough for theoreticians working in stellar structure, stellar atmospheres, galactic dynamics, etc.

Figure 19.14
A large modern computer. This computer, an IBM System 360/40 is extensively used to make model atmosphere and interior calculations. *(Courtesy of the International Business Machines Corporation)*

19.15 Space Astronomy

Numerous special-purpose satellites have been flown. Included in this category are the various Orbiting Solar Observatories, the Pioneers and many others dedicated to X-ray studies, particles and fields in the solar system, etc. All have advanced our knowledge greatly. The crowning instrument was the first successful Orbiting Astronomical Observatory (OAO).

Figure 19.15A
Artist's version of a large orbiting astronomical telescope. *(NASA Photograph)*



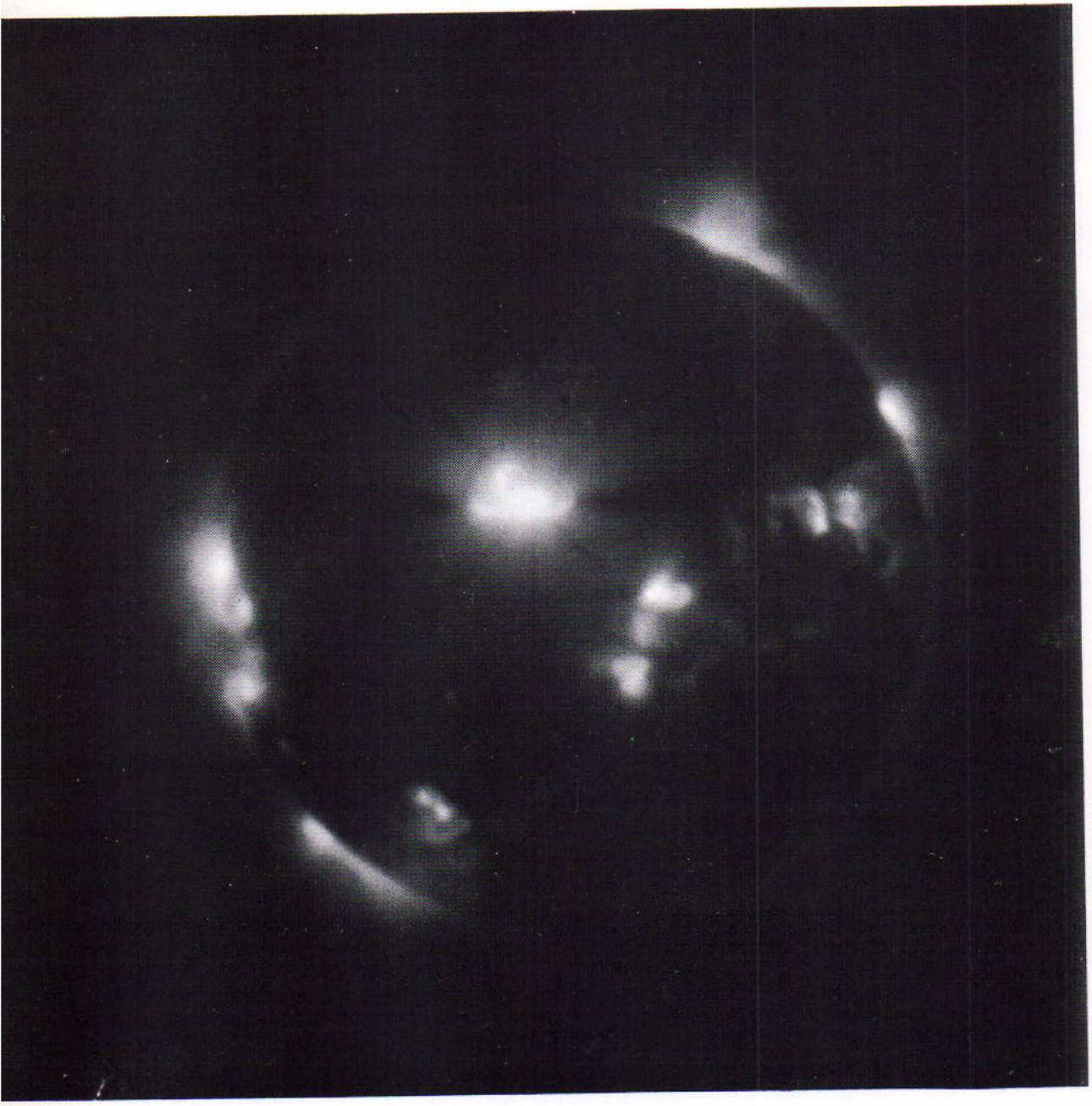


Figure 19.15B
The Sun in the X-ray region. X-ray photograph taken by a rocket-borne x-ray telescope on June 8, 1968. *(American Science & Engineering, Inc.)*

We expect the OAO series to be extended by adding several instruments covering the next 15 years. We then expect to see a large space telescope launched that is capable of tackling the cosmological problems. This instrument may not be in the cards until the late 1980's, but once in operation it will yield a quantum jump in our knowledge.

The moon may become an operational astronomy base. It appears to be a very good place to do γ-ray and X-ray astronomy. Certainly the far side of the moon, shielded from earth-caused radio contamination, would be an ideal location for radio astronomy.

Concerning the planetary system we expect to reap a harvest of knowledge on the origin and evolution of life and the origin and evolution of the solar system. Mars, Venus, and Jupiter will receive most of the attention, for knowledge about these three planets offers the biggest rewards for answering our questions in detail. We hope, however, that it will be possible to probe an asteroid and to fly a spacecraft into a comet.

Planetary astronomers are especially interested in the preparations to send a Mariner type spacecraft past Jupiter, Saturn, Uranus and Neptune. The four major planets will be positioned in such a way during 1977–1978 that each planet in turn imparts energy to the spacecraft and "slings" it out to the next planet. Because of this, the mission is called "the Grand Tour." It is expected that such a mission will reveal information on the interiors of the major planets, examine their atmospheres in detail and might even provide us with a close-up view of some of their satellites.

THE PHYSICAL UNIVERSE

The unfolding of knowledge of the universe has revealed a scene of ever-increasing grandeur. It began with the anthropocentric view in which the observer was the central figure. Around him was a flat earth and over him bent a sky full of stars, which rose and set daily and marched westward with the changing seasons.

In the geocentric view of the early Greek scholars the stationary globe of the earth was the dominant feature. The sun, moon, and planets revolved around the earth within the rotating sphere of the "fixed stars." Next, the heliocentric view, dating formally from the time of Copernicus, established the planetary system on a more nearly correct basis and brought the sphere of the stars to rest. Then emerged the idea that the stars are other suns, many of them perhaps attended by planetary systems.

The invention and development of the telescope extended the inquiries into the star fields and promoted the first attempts to determine the structure of the universe. The discovery of the law of gravitation inaugurated dynamical interpretations of what goes on in the heavens. The recognition of physically related pairs of stars, of star clusters, of nebulae that seemed not of a starry nature, and of other nebulae that seemed to avoid the Milky Way—these and other features of the sidereal scene presented problems for study with the eye at the telescope and later in photographs made with telescopes.

Astrophysics, the "new astronomy," extended techniques of the terrestrial physical laboratories to the laboratory of space, with benefit to the findings of both. It employs the spectroscope, the photographic plate, the photoelectric cell, filters for transmitting celestial radiations at various wavelengths, the radio telescope, and other devices. Equipment sent above the earth record and relay to the ground celestial information not previously available to us under the blanket of the atmosphere.

In recent years, as we have seen, the picture of the universe is unfolding with spectacular rapidity. Our galaxy, having the sun in its suburbs, now stands out clearly as a spiral structure in the foreground of the celestial scene. The formerly mysterious spirals and associated "extragalactic nebulae" are established as galaxies, often gathered into groups and larger clusters. Attractive theories are available as to how stars are born in the cosmic clouds and how they continue to shine and to build up heavier chemical elements in their interiors, until they end in the celestial cinder heap.

Evidence is also available that unusual processes are at work in the cosmos. From the pulsars and quasars we deduce that in these objects matter is in a condition impossible to achieve in the laboratory and that sources of energy are being tapped that we do not know about. It is almost certain that the laboratory of space is about to yield new basic laws of physics.

REVIEW QUESTIONS

1. What are three possible explanations for the redshift? Which explanation do the majority of astronomers subscribe to?
2. What is the limiting value for the Doppler effect if an object is moving directly at us?
3. Why must radio galaxies have unusual processes going on in them?
4. What new observation might link quasi stellar objects to galaxies?
5. Can you advance an argument why quasi stellar objects do not form a homogeneous group of objects? While this is an answerable question why may it not be a valid question?
6. What arguments show that the progressive reddening of the galaxies is not an evolutionary effect?
7. State the cosmological principle.
8. Give two arguments that indicate the steady state model does not describe the Universe as observed.

9. Give some recent observations that lend credence to using a relativistic model of the Universe.
10. We tend to favor a decelerating relativistic model for the Universe. What major piece of observational information must be circumvented to make this model stand up?
11. Why do the authors feel that life as we know it is not unique to the earth? Can you give other arguments?
12. Why would the moon be an excellent place for an observatory? Can you think of any scientific reasons why it would not be a good site (or why certain areas of the moon would not make good sites)?

REFERENCES

Alfven, H., *Worlds-Antiworlds,* San Francisco: Freeman Co., 1966.
Bonnor, H., *The Mystery of the Expanding Universe,* New York: The Macmillan Co., 1964.
Gamow, G., *The Creation of the Universe,* New York: Viking Press, 1952.

FOR FURTHER STUDY

Bondi, H., *Cosmology,* Cambridge Univ. Press, 2nd ed., 1960.
McVittie, G. C., *General Relativity and Cosmology,* 2nd ed. Urbana, Ill.: Univ. of Illinois Press, 1965.
North, J. D., *The Measure of the Universe,* New York: Oxford Univ. Press, 1965.
Schatzman, E., *The Origin and Evolution of the Universe,* New York: Basic Books, 1966.

Man's bootprint on the lunar dust. *(NASA photograph)*

GENERAL REFERENCES

The junior author has resorted quite often to the following works, and considers them essential on a good reference shelf:

Beer, Arthur, ed., *Vistas in Astronomy,* Oxford: Pergamon Press, 1955 to the present. (a continuing series).

Burham, Robert, *Celestial Handbook,* Flagstaff, Arizona, Northland Press, 1965.

Feynman, Richard, R. Leighton and M. Sands, *The Feynman Lectures on Physics,* 3 vols., Reading, Mass.: Addison-Wesley Publishing Co., 1963–64.

Goldberg, Leo, ed., *Annual Review of Astronomy and Astrophysics,* Palo Alto, Cal., Annual Reviews, Inc., 1963 to the present.

Kuiper, Gerald and Barbara Middlehurst, eds., *Stars and Stellar Systems,* Chicago: University of Chicago Press, 1960 to the present; work still in progress.

ENCYCLOPEDIAS

Encyclopedia Britannica, Chicago: William Benton (yearly editions)

Van Nostrand's Scientific Encyclopedia, 4th ed., New York: Van Nostrand Reinhold, 1968.

JOURNALS

Some continuing sources of the results of astronomical research are listed here for reference. Relatively popular and reportorial type references are listed first. Research journals readily available internationally are listed last.

The Griffith Observer, monthly, Griffith Observatory, P. O. Box 27787 Los Angeles, California 90027.

Natural History, monthly, The American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024.

The Observatory, monthly, The Editors, Royal Greenwich Observatory, Herstmonceaux Castle, Hailshorn, Sussex, England.

Science News, weekly, Science News, 1719 N Street, N. W., Washington, D.C., 20036.

Scientific American, monthly, Scientific American, 415 Madison Avenue, New York, New York.

Sky and Telescope, monthly, Sky Publishing Corporation, 49–51 Bay State Rd. Cambridge, Massachusetts 02138.

The Strolling Astronomer, bi-monthly, The Strolling Astronomer, Box 3AZ, University Park, New Mexico 88001.

The Astronomical Journal, monthly, American Institute of Physics, 335 East 45th Street, New York, New York 10017.

Astronomy and Astrophysics, monthly, Springer-Verlag New York Inc., 175 Fifth Avenue, New York, New York 10010.

The Astrophysical Journal, bi-monthly, University of Chicago Press, 5750 Ellis Avenue, Chicago, Illinois 60637.

Astrophysical Letters, monthly, Gordon and Breach Science Publishers, Inc., 150 Fifth Avenue, New York, New York 10011.

The Australian Journal of Physics, monthly, Commonwealth Scientific and Industrial Research Organization, 372 Albert Street, East Melbourne, Victoria 3002, Australia.

The Journal of the British Astronomical Association, monthly, 303 Bath Road, Hounslow West, Middlesex England.

The Journal of the Royal Astronomical Society of Canada, monthly, the Royal Astronomical Society of Canada, 252 College Street, Toronto 130, Canada.

Monthly Notices of the Royal Astronomical Society, monthly, Blackwell Scientific Publications Limited, 5 Alfred Street, Oxford England.

Nature, weekly, Macmillan Journals, Limited, Brunel Road Basingstoke, Hampshire, England.

Proceedings of the Astronomical Society of Australia, monthly, Treasurer of the School of Physics, University of Sydney, N.S.W. 2006 Australia.

Publications of the Astronomical Society of the Pacific, monthly, The Astronomical Society of the Pacific, California Academy of Science, Golden Gate Park, San Francisco, California 94118.

The Quarterly Journal of the Royal Astronomical Society, quarterly, Blackwell Scientific Publications Limited, 5 Alfred Street, Oxford England.

Science, weekly, American Association for the Advancement of Science, 1515 Massachusetts Avenue, N.W., Washington, D.C. 20005.

For additional listings as well as a brochure entitled "A Career In Astronomy" the reader may write to:

H. M. Gurin, Executive Officer
The American Astronomical Society
211 Fitz Randolph Road
Princeton, New Jersey 08540

The *American Astronomical Society* is the professional society covering Canada, Mexico and the United States with the primary goals of advancing astronomical research, experimental tools and techniques (for carrying out this research) and education in astronomy. To accomplish these goals, the American Astronomical Society holds several three-day meetings each year and sponsors colloquia, certain professional journals, films, lecture programs, etc. Information concerning the American Astronomical Society can be obtained by writing to its Executive Officer.

APPENDIX

Angular Measure and Temperature Scales

The unit of angular measure used mainly in this book is the sexagesimal degree (°) and its subdivisions the minute (′) and second (″) of arc. The circumference of a circle is divided into 360 degrees, the degree into 60 minutes, and the minute into 60 seconds. Another unit of angular measure is the *radian* (11•3), which is the angle subtended by the radius of a circle laid off along its circumference. The radian equals 57°.3, or 3437′.7, or 206,264″.8.

Two temperature scales are *Fahrenheit* (F), which is still commonly used in civil life in the United States, and the *centigrade* (C), which is used for scientific purposes in all countries. In the familiar mercury thermometer the heights of the mercury column are marked on the glass tube for the ordinary freezing and boiling points of water. Between these two marks the tube is divided evenly into 180 parts for the Fahrenheit scale and into 100 parts for the centigrade scale. The space between two adjacent divisions represents 1° change in temperature. Hence 1°F = 5/9 of 1°C. Water freezes at 0°C, or 32°F, and boils at 100°C, or 212°F. Thus to convert a temperature reading from the Fahrenheit to the centigrade scale subtract 32° and multiply by 5/9.

Used exclusively in scientific measurements, the absolute centigrade, or *Kelvin* (K), scale of temperature starts at −273°C as absolute 0°K. Water freezes at 273°K and boils at 373°K.

Basic Astronomical Data

	METRIC	ENGLISH	IN SECTION
Earth's equatorial radius	6,378 km	= 3963 miles	2.3
Earth's mass	5.98×10^{21} metric tons	= 6.6×10^{21} tons	2.4
Velocity of light	299,792 km/sec	= 186,300 mi/sec	4.1
Moon's Mass	7.35×10^{9} metric tons	= $\frac{1}{81}$ × earth's mass	5.1
Moon's Radius	1,740 km	= 1080 miles	5.1
Moon's Mean Distance	384,400 km	= 238,857 miles	5.4
Astronomical unit (AU)	149,598,000 km	= 92,956,000 miles	7.16
Sun's mass	1.99×10^{27} metric tons	= $\frac{1}{3}$ million × earth's	10.1
Sun's radius	696,000 km	= 432,000 miles	10.1
Light year (LY)	9.46×10^{12} km	= 5.88×10^{12} miles	11.3
Parsec (pc)	308×10^{13} km	= 1.92×10^{13} miles	11.3

TABLE 1 The Planets

	NAME	SYMBOL	MEAN DISTANCE FROM SUN		PERIOD OF REVOLUTION		ECCENTRICITY OF ORBIT	ORBITAL INCLINATION TO ECLIPTIC
			ASTRON. UNITS	KILOMETERS ×10⁶	*Sidereal*	*Synodic*		
					days	days		
Inner	Mercury	☿	0.3871	57.91	87.969	115.88	0.206	7° 0′
	Venus	♀	0.7233	108.20	224.701	583.92	0.007	3 24
	Earth	⊕	1.0000	149.60	365.256	.....	0.017	0 0
	Mars	♂	1.5237	227.94	686.980	779.94	0.093	1 51
					years			
	Ceres	①	2.7673	413.98	4.604	466.60	0.077	10 37
Outer	Jupiter	♃	5.2028	778.33	11.862	398.88	0.048	1 18
	Saturn	♄	9.5388	1429.99	29.458	378.09	0.056	2 29
	Uranus	♅	19.1819	2869.57	84.013	369.66	0.047	0 46
	Neptune	♆	30.0579	4496.60	164.794	367.49	0.009	1 46
	Pluto	♇	39.5177	5911.77	248.430	366.74	0.249	17 9

NAME	EQUATORIAL DIAMETER IN KILOMETERS	MASS ⊕ = 1	DENSITY WATER = 1	PERIOD OF ROTATION IN DAYS	INCLINATION OF EQUATOR TO ORBIT	OBLATENESS	STELLAR MAGNITUDE AT GREATEST BRILLIANCY	ALBEDO
Sun ☉	1,392,000	332,960	1.41	25.38	7° 10′	0	−26.8	.....
Moon ☾	3,476	0.012	3.34	27.322	6 41	0	−12.6	0.07
Mercury	4,868	0.05	5.49	59:	<28°	0	−1.9	0.06
Venus	12,112	0.82	5.26	244	3° 18′	0	−4.4	0.76
Earth	12,756	1.00	5.52	0.997	23 27	.003	.....	0.36
Mars	6,787	0.11	3.94	1.026	24 59	.005	−2.8	0.16
Jupiter	142,830	317.9	1.33	0.410	3 4	.067	−2.5	0.73
Saturn	119,330	95.1	0.71	0.426	26 44	.105	−0.4	0.76
Uranus	47,200	14.5	1.70	0.451	97 53	.071	+5.7	0.93
Neptune	50,000	17.5	1.63	0.658:	28 48	.025	+7.6	0.84
Pluto	5,800	0.10:	5.5	6.39	?	?	+14.9	0.14:

TABLE 2 The Moon and Other Satellites

NAME	DISCOVERY	MEAN DISTANCE FROM PRIMARY IN KILOMETERS	MEAN SIDEREAL PERIOD OF REVOLUTION IN DAYS	DIAMETER IN KILOMETERS	STELLAR MAGNITUDE AT MEAN OPPOSITION
Moon		384,397	27.322	3476	−12.7
	SATELLITES OF MARS				
Phobos	Hall, 1877	9,350	0.319	13:	+11.6
Deimos	Hall, 1877	23,480	1.261	8:	12.8
	SATELLITES OF JUPITER				
Fifth	Barnard, 1892	181,000	0.498	193	13.0
I Io	Galileo, 1610	421,800	1.769	3220	4.8
II Europa	Galileo, 1610	671,100	3.551	2900	5.2
III Ganymede	Galileo, 1610	1,070,000	7.155	4990	4.5
IV Callisto	Galileo, 1610	1,883,000	16.689	4500	5.5
Sixth	Perrine, 1904	11,478,000	250.583	96:	13.7
Seventh	Perrine, 1905	11,737,000	259.667	30:	16.0
Tenth	Nicholson, 1938	11,721,000	259.208	19:	18.6
Twelfth	Nicholson, 1951	21,243,000	631	19:	18.8
Eleventh	Nicholson, 1938	22,531,000	692	19:	18.1
Eighth	Melotte, 1908	23,496,000	744	19:	18.8
Ninth	Nicholson, 1914	23,657,000	758	19:	18.3
	SATELLITES OF SATURN				
Janus	Dollfus, 1967	157,700	0.749	300	14.0
Mimas	Herschel, 1789	185,600	0.942	480	12.1
Enceladus	Herschel, 1789	238,200	1.370	560	11.8
Tethys	Cassini, 1684	294,800	1.888	960	10.3
Dione	Cassini, 1684	377,500	2.737	960	10.4
Rhea	Cassini, 1672	527,200	4.517	1287	9.8
Titan	Huygens, 1655	1,221,000	15.945	4830	8.4
Hyperion	Bond, 1848	1,483,000	21.276	480:	14.2
Iapetus	Cassini, 1671	3,560,000	79.331	1090:	11.0
Phoebe	Pickering, 1898	12,952,000	550.333	190:	16.5

The Moon and Other Satellites (contd.)

NAME	DISCOVERY	MEAN DISTANCE FROM PRIMARY IN KILOMETERS	MEAN SIDEREAL PERIOD OF REVOLUTION IN DAYS	DIAMETER IN KILOMETERS	STELLAR MAGNITUDE AT MEAN OPPOSITION
		SATELLITES OF URANUS			
Miranda	Kuiper, 1948	130,500	1.413	300:	16.5
Ariel	Lassell, 1851	191,800	2.520	800:	14.4
Umbriel	Lassell, 1851	267,200	4.144	650:	15.3
Titania	Herschel, 1787	438,400	8.706	1130:	14.0
Oberon	Herschel, 1787	586,300	13.463	960:	14.2
		SATELLITES OF NEPTUNE			
Triton	Lassell, 1846	355,200	5.876	3700	13.5
Nereid	Kuiper, 1949	5,562,000	359.875	320:	18.7

TABLE 3 Names of the Constellations

LATIN NAME	POSSESSIVE	ENGLISH EQUIVALENT	MAP
*Androm'eda	Androm'edae	Andromeda	4, 5
Ant'lia	Ant'liae	Air Pump	
A'pus	A'podis	Bird of Paradise	
*Aqua'rius	Aqua'rii	Water Carrier	4
*Aq'uila	Aq'uilae	Eagle	3, 4
*A'ra	A'rae	Altar	6
*A'ries	Ari'etis	Ram	4, 5
*Auri'ga	Auri'gae	Charioteer	5
*Boö'tes	Boö'tis	Herdsman	2, 3
Cae'lum	Cae'li	Graving Tool	
Camelopar'dalis	Camelopar'dalis	Giraffe	
*Can'cer	Can'cri	Crab	2, 5
Ca'nes Vena'tici	Ca'num Venatico'rum	Hunting Dogs	2
*Ca'nis Ma'jor	Ca'nis Majo'ris	Larger Dog	5
*Ca'nis Mi'nor	Ca'nis Mino'ris	Smaller Dog	5
*Capricor'nus	Capricor'ni	Sea-Goat	4
†Cari'na	Cari'nae	Keel	6
*Cassiope'ia	Cassiope'iae	Cassiopeia	1, 4
*Centau'rus	Centau'ri	Centaur	2, 6
*Ce'pheus	Ce'phei	Cepheus	1, 4
*Ce'tus	Ce'ti	Whale	4, 5
Chamae'leon	Chamaeleon'tis	Chameleon	
Cir'cinus	Cir'cini	Compasses	
Colum'ba	Colum'bae	Dove	5
Co'ma Bereni'ces	Co'mae Bereni'ces	Berenice's Hair	2
*Coro'na Austra'lis	Coro'nae Austra'lis	Southern Crown	
*Coro'na Borea'lis	Coro'nae Borea'lis	Northern Crown	3
*Cor'vus	Cor'vi	Crow	2
*Cra'ter	Crater'is	Cup	2
Crux	Cru'cis	Cross	6
*Cyg'nus	Cyg'ni	Swan	3, 4
*Delphi'nus	Delphi'ni	Dolphin	4
Dora'do	Dora'dus	Dorado	
*Dra'co	Draco'nis	Dragon	1, 3
*Equu'leus	Equu'lei	Little Horse	
*Erid'anus	Erid'ani	River	5, 6
For'nax	Forna'cis	Furnace	
*Gem'ini	Gemino'rum	Twins	5
Grus	Gru'is	Crane	4
*Her'cules	Her'culis	Hercules	3
Horolo'gium	Horolo'gii	Clock	
*Hy'dra	Hy'drae	Sea Serpent	2
Hy'drus	Hy'dri	Water Snake	6
In'dus	In'di	Indian	
Lacer'ta	Lacer'tae	Lizard	
*Le'o	Leo'nis	Lion	2

TABLE 3 Names of the Constellations—*Continued*

LATIN NAME	POSSESSIVE	ENGLISH EQUIVALENT	MAP
Le'o Mi'nor	Leo'nis Mino'ris	Smaller Lion	
*Le'pus	Le'poris	Hare	5
*Li'bra	Li'brae	Scales	3
*Lu'pus	Lu'pi	Wolf	3
Lynx	Lyn'cis	Lynx	
*Ly'ra	Ly'rae	Lyre	3, 4
Men'sa	Men'sae	Table Mountain	
Microsco'pium	Microsco'pii	Microscope	
Monoc'eros	Monocero'tis	Unicorn	
Mus'ca	Mus'cae	Fly	6
Nor'ma	Nor'mae	Level	
Oc'tans	Octan'tis	Octant	
*Ophiu'chus	Ophiu'chi	Serpent Holder	3
*Ori'on	Orio'nis	Orion	5
Pa'vo	Pavo'nis	Peacock	6
*Peg'asus	Peg'asi	Pegasus	4
*Per'seus	Per'sei	Perseus	4, 5
Phoe'nix	Phoeni'cis	Phoenix	4
Pic'tor	Picto'ris	Easel	
*Pis'ces	Pis'cium	Fishes	4
*Pis'cis Austri'nus	Pis'cis Austri'ni	Southern Fish	4
†Pup'pis	Pup'pis	Stern	5
†Pyx'is	Pyx'idis	Mariner's Compass	
Retic'ulum	Retic'uli	Net	
*Sagit'ta	Sagit'tae	Arrow	3, 4
*Sagitta'rius	Sagitta'rii	Archer	3
*Scor'pius	Scor'pii	Scorpion	3
Sculp'tor	Sculpto'ris	Sculptor's Apparatus	4
Scu'tum	Scu'ti	Shield	
*Ser'pens	Serpen'tis	Serpent	3
Sex'tans	Sextan'tis	Sextant	
*Tau'rus	Tau'ri	Bull	5
Telesco'pium	Telesco'pii	Telescope	
*Trian'gulum	Trian'guli	Triangle	4, 5
Trian'gulum Austra'le	Trian'guli Austra'lis	Southern Triangle	6
Tuca'na	Tuca'nae	Toucan	6
*Ur'sa Ma'jor	Ur'sae Majo'ris	Larger Bear	1, 2
*Ur'sa Mi'nor	Ur'sae Mino'ris	Smaller Bear	1, 3
†Ve'la	Velo'rum	Sails	2, 6
*Vir'go	Vir'ginis	Virgin	2
Vo'lans	Volan'tis	Flying Fish	
Vulpec'ula	Vulpec'ulae	Fox	

* One of the 48 constellations recognized by Ptolemy.
† Carina, Puppis, Pyxis, and Vela once formed the single Ptolemaic constellation Argo Navis.

TABLE 4 Oppositions of Jupiter

DATE		DISTANCE	MAGNITUDE
1970	Apr 21	4.44 A.U.	−2.0
1971	May 23	4.35	−2.1
1972	Jun 24	4.22	−2.21
1973	Jul 30	4.08	−2.3
1974	Sep 5	4.00	−2.4
1975	Oct 13	3.96	−2.5
1976	Nov 18	4.02	−2.4
1977	Dec 23	4.15	−2.3
1979	Jan 24	4.29	−2.2
1980	Feb 24	4.40	−2.1

Distance is from earth at opposition

TABLE 5 Clusters of Galaxies

CLUSTER	DISTANCE (Mpc)	RADIAL VELOCITY (km/sec)	NUMBER OF GALAXIES
Local Group	—	—	17
Virgo	11	1200	2500
Perseus	58	5500	500
Coma	68	6700	1500
Ursa Maj. I	160	15400	300
Corona Bor.	190	21600	400
Ursa Maj. II	380	42000	400

GLOSSARY

aberration 1) of starlight—the apparent shift in direction of the observed object due to the tangential motion of the observer. 2) of optics—defects in an optical image more or less correctable by careful design.

absolute magnitude the brightness a star or object would have at the arbitrary distance of 10 parsecs; designated by M and a subscript to indicate the region of the spectrum referred to.

absolute zero by definition the lowest possible temperature; 0°K or −273°C, almost all molecular motion ceases at this temperature.

absorption coefficient a measure of the rate of decrease in intensity of radiation passing through a substance (such as a gas cloud).

absorption spectrum a dark line spectrum superposed on a continuous spectrum and caused by a cool gas between the observer and the continuous source.

acceleration change in velocity or direction.

AGK2 Kataloge der Astronomischen Gesellschaft 2. A catalogue of almost 200,000 star positions north of −2°. Soon to be superceded by the AGK3. Both catalogues are in the system of the FK4 and for epoch 1950.

albedo the ratio of the reflected light from an object to that illuminating the object.

almanac a compilation of tables of astronomical events and published in advance.

altitude angle of an object, above or below the horizon measured along its vertical circle.

amplitude the range of variability.

angstrom a convenient unit of length for expressing wavelengths of visible and ultraviolet light; 1 angstrom unit equals 10^{-8}cm; the symbol is Å or with increasingly common use, just A.

angular diameter the angle subtended by the diameter of an object.

angular distance the angle on the celestial sphere along a great circle between two objects.

angular momentum the momentum of a body derived by its motion about an axis or point.

annular eclipse an eclipse of the sun that is central or almost central and showing a ring of the disk of the sun.

anomalistic month the month measured by the moon's motion from perigee to perigee.

anomalistic year the year measured by the earth's motion from perihelion to perihelion.

anomaly 1) non-uniform motion 2) the angle between the planet, sun and perihelion.

antenna temperature the temperature which a resistor must have to equal the noise received by a radio antenna.

aperture the effective collecting area of a telescope.

apo a prefix meaning the farthest point in an orbit, thus apogee is the farthest point in the orbit of an object orbiting the earth. Occasionally reduced to ap- (as in aphelion) for simpler phonetics.

apparent magnitude the brightness of an object in magnitudes as observed; designated by m and a subscript denoting the spectral region referred to. The apparent magnitude equals the absolute magnitude at a distance of 10 parsecs.

apparent solar time time determined by the hour angle of the true sun; apparent noon is when the center of the sun crosses the meridian.

apparent sun the true sun.

appulse a penumbral lunar eclipse.

apsidal motion rotation of the major axis of an orbit in the plane of the orbit.

ascending node the point where the motion of a planet carries it from below to above the ecliptic plane or, in the case of stars, out of the plane of the sky towards the observer.

association a loose grouping of stars with characteristics such that one can say they had a common origin.

Astrographic Catalogue see *Carte du Ciel.*

astrology 1) a primitive religion originating in the near East. 2) a pseudo science.

astrometric binary a binary star where only one star is seen; the other reveals itself by disturbing the observed proper motion.

astrometry the branch of astronomy dealing with the positions, distances and motions of the planets and stars; it includes the determination of time and position.

astronautics application of the physical laws of motion to space flight.

astronomical unit the mean distance of the earth from the sun; designated by AU its current value is 1.495985×10^8km.

astronomy the branch of science dealing with the Universe beyond the earth's atmosphere.

astrophysics the branch of astronomy that applies the theories and laws of physics to celestial bodies.

autumnal equinox that point on the celestial equator where the sun passes from above to below the equator; approximately September 21.

azimuth the angle from the north point eastward along the hori-

zon to the vertical circle passing through the object in question.

Bailey's Beads in a total solar eclipse the sun's chromosphere shows between the mountains and ridges of the moon giving a beaded appearance.

Balmer discontinuity where the Balmer lines fall so close together a general depression of the continuum occurs.

Balmer lines electronic transitions in the hydrogen atom from upper levels to the second level (emission lines) or from the second level to upper levels (absorption lines).

barred spiral a spiral galaxy the arms of which start from the ends of a bar passing through the center of the galaxy.

barycenter center of mass in a two-body system.

BD The *Bonner Durchmusterung*, a valuable catalogue and series of charts showing positions of stars (epoch 1855) down to about apparent magnitude 10.

"big bang" theory a cosmological theory having the Universe originate from a primeval explosion.

binary star a pair of stars, gravitationally bound to each other.

black body a hypothetically perfect radiator that absorbs and reemits all energy falling on it.

black dwarf the end product of stellar evolution.

blink comparator a measuring machine which alternately allows the measurer to see first one plate and then a second. Used to discover moving objects or variable stars. It is often referred to simply as the blink.

Bode's relation an empirical relationship of the form $D = A + BC^n$ giving the distances of the planets in astronomical units. Here $A = 0.4$, $B = 0.3$, $C = 2$ and $n = -\infty, 0, 1, 2, 3 \ldots$.

bolide a meteor of extreme brightness, a fireball that breaks up.

bolometric correction a factor required to correct magnitudes on one system to bolometric magnitudes.

bolometric magnitude the total flux received from an object, the measurement covers the entire electromagnetic spectrum and is corrected to outside the earth's atmosphere. The symbols M_{bol} and m_{bol} are used.

bremsstrahlung energy radiated by an electron being accelerated usually by a nucleus or another charged particle.

brightness loosely and variously used by astronomers to indicate luminosity of a body or intensity of radiation.

brightness temperature the temperature that a black body would have to radiate the observed power.

carbon cycle a nuclear process converting hydrogen to helium using carbon in the role of a catalyst.

candle the measure of luminous intensity.

cardinal points the four primary points of the compass, North, East, South, West.

Cassini's division a major gap in the ring system around Saturn.

Carte du Ciel a major astrographic catalogue mapping the entire sky with identical telescopes. Its epoch is 1900 but all of the zones are not finished (as of 1970).

CD the Cordoba Durchmusterung, a southern extension of the BD carried out by the Cordoba Observatory.

celestial equator the intersection of the plane of the earth's equator with the celestial sphere.

celestial horizon the ideal horizon, 90° from the zenith; the horizon as defined by the eye at sea level.

celestial mechanics the branch of astronomy dealing with the motions of multiple bodied systems.

celestial meridian the great circle on the celestial sphere passing through the celestial poles and the observer's zenith.

celestial sphere the infinitely distant apparent sphere of the sky.

cepheid a type of pulsating variable star with periods between 1 day and about 60 days.

CerVit trade name by Owens-Illinois for a material that can be finished as mirrors and has an essentially zero temperature coefficient.

chromatic aberration one of the major aberrations of an optical system where the system fails to bring all wavelengths to the same focus.

chronograph an accurate recording timekeeping device.

chronometer an exceedingly accurate clock.

chromosphere the portion of a stellar (solar) atmosphere just above the photosphere. It receives its name from its bright red color.

circumpolar stars near the pole that never go below the observer's horizon.

cluster a grouping of stars (see open cluster and globular cluster) or galaxies.

cluster variable a class of pulsating stars with periods less than a day.

coelostat a stationary telescope looking at a flat mirror in a polar mount. The drive is arranged to keep the sun or a star field centered in the telescope.

collimator an optical system designed to render light parallel. A telescope is a reverse collimator.

color excess the increased reddening of a star due to a scattering intervening material.

color index strictly, the difference between the photovisual and the photographic magnitudes of

a star; it is sometimes loosely used to mean the difference between any two color measurements of a star.

color temperature temperature reading obtained by taking measurements of photographs in two colors and fitting them to a given radiation law.

coma 1) one of the major aberrations of an optical system
2) the hazy material around the nucleus of a comet
3) when capitalized, the great cluster of galaxies in the constellation Coma Berenices.

comparison spectrum the spectrum of emission lines placed beside the observed objects' spectrum for identification and standardization purposes.

conjunction originally a planet having the same longitude as the sun, more recently also when two bodies have the same longitude or right ascension.

constellation a configuration of stars named by the ancients to delineate areas of the sky. This has been extended as a convenience to the southern hemisphere.

contacts the four distinct moments in an eclipse.

corona the pearly green light around the sun during total eclipse—the outermost atmosphere of the sun.

coronagraph an instrument designed for studying the corona of the sun without a total eclipse.

cosmic rays particles impinging on the earth's atmosphere at extremely high energies, mostly high energy protons.

cosmogony the study of the origin and evolution of the Universe.

cosmological principle the basic precept stating that the Universe is the same from any given place at a given time.

coudé more correctly the coudé focus where the image formed by the telescope may rotate but is at a fixed position with respect to the earth.

crescent a phase of the moon or planets where the elongation from the sun is less than 90°.

curve of growth a plot of the equivalent widths of spectral lines (as the ordinate) against the number of atoms required to form the line (as the abscissa).

declination one of the coordinates of the equatorial system, it is the angular distance from the celestial equator on a great circle through the poles.

deceleration a change in velocity to lower values, a negative acceleration.

deferent a circle on which another circle containing an object moves.

degenerate gas a condition where most or all of the lower energy levels are filled and hence requiring an external energy

source for physical processes to take place.

differential rotation motion where the outer parts of a system rotate at a different velocity than those nearer the center.

diffraction in physical optics the spreading of light as it passes an opaque edge.

diffraction grating a ruled grating such that upon reflection from an edge (or transmission past an edge) the light path is different for different wavelengths leading to an angle of reflection (or transmission) dependent upon wavelength.

diffuse nebula a reflection or emission nebula caused by a concentration of interstellar matter near a bright star.

dispersion the separation of electromagnetic radiation by wavelength due to a characteristic of the medium.

distance modulus the difference between the apparent magnitude and the absolute magnitude, related to the distance in parsecs by $m - M = 5 \log D - 5$.

diurnal daily or happening each day.

diurnal circle the path of a star on the celestial sphere due to the daily rotation of the earth.

diurnal parallax the apparent change in direction to an object due to the baseline created for an observer by the daily rotation of the earth.

Doppler effect the change in frequency due to the relative motion between the source and observer.

double star any two stars in close proximity to each other as seen on the celestial sphere, such a pair may be a chance occurance (an optical double) or a pair gravitationally bound to each other (a binary star).

draconitic month a lunar month as measured from a given node of the moon's orbit and back to the same node.

dwarf a star on the main sequence, for example: the sun.

dynamical parallax a method for deriving the distance to a binary star system using the laws of motion and the mass-luminosity relation.

earthlight sunlight reflected from the earth.

eccentric motion on a circle whose axis of rotation does not coincide with its center.

eclipse cutting off from view, either partially or totally, one body by the interposition of another.

eclipse season occurs twice a year when the sun passes near the nodes of the moon's orbit.

eclipsing binary a binary, the plane of whose orbit is such as to contain or pass close to the solar system so that the stars alternately eclipse each other.

ecliptic the path of the sun on the celestial sphere.

ecliptic limit the angular distance before and after a node in which the moon must be to have an eclipse.

effective temperature the temperature of a body as deduced from the Stefan-Boltzmann Law.

Einstein effect a shift in the frequency of radiation due to the mass of the emitting body, $\frac{\Delta\lambda}{\lambda} = \frac{GM}{c^2R}$, where M is the mass of the body and R is its radius.

electromagnetic radiation radiant energy from γ rays through visible light to the longest radio waves and caused by an oscillating electric or magnetic charge.

elements the seven parameters required to specify the size, shape and orientation of an orbit.

elliptical galaxy a galaxy having an ellipsoidal distribution of stars and characterized by a lack of any appreciable gas or dust.

elongation the difference between the longitudes of the sun and a planet or the moon.

encounter a close chance passing resulting in a gravitational perturbation of the motions.

ephemeris tables of astronomical data and positions published in advance and forming the basis of an almanac.

ephemeris time time based upon a definite year (1900) and used to calculate times of astronomical phenomena. It is necessitated by the slowing and non-uniform rotation of the earth.

epicycle a circle moving on a circle, used in early cosmologies to explain the retrograde motion of the planets.

epoch an arbitrary date to which observations are referred or reduced.

equant a point in the plane of an orbit about which an epicycle or body revolves with uniform angular velocity.

equation of time the difference between apparent and mean solar time.

equator the great circle 90° from the poles of the earth or the celestial sphere.

equatorial mount a telescope mounting placing the telescope motions in the equatorial system necessitating a driving motion in only one coordinate.

equatorial system the coordinate system based upon the celestial equator as its fundamental plane.

equinox any of the two intersections of the equatorial and ecliptic planes on the celestial sphere.

equivalent width the width of a rectangular absorption line with zero intensity and whose area is equivalent to that of the true absorption line.

eruptive variable a star whose light output varies abruptly and erratically.

ether the medium once believed to be required to transmit electromagnetic radiation through space.

extinction the dimming of starlight by the earth's atmosphere or interstellar clouds, also the attenuation of radio radiation by the interstellar medium and the disk of the Galaxy.

extragalactic outside the Galaxy, i.e. the Milky Way.

F number the ratio of the aperture to the focal length of a telescope or lens. A measure of the system's photographic "speed".

faculae bright regions near the limb of the sun.

filar micrometer a micrometric measuring device used to measure small angular separations or diameters.

fireball 1) a meteor brighter than the planet Jupiter.
2) the explosion of the primeval atom in certain cosmologies.

FK3 Fundamental Katalog Number Three, a catalogue of about 4000 stars spread over the entire celestial sphere and used to determine positions of other stars. Being superseded by the FK4.

flare a sudden, temporary brightening of a small region on the sun.

flare star a group of late type dwarfs that show flare like brightenings.

flash spectrum the spectrum of the sun's reversing layer formed by a slitless spectrograph using the thin crescent of the sun just before or just after totality as its slit.

flocculi see *plage.*

fluorescence the absorption of radiation of one wavelength and its reemission at another wavelength which ceases when the source radiation ceases.

flux energy passing through a surface area in unit time.

flux unit 10^{-26} watts per square meter per Hertz.

focal length the distance from the objective to the focus of a telescope. In a compound telescope it is the distance from the entrance pupil to the focus and is called the effective focal length.

focal ratio see F number.

focus the point where rays from a very distant object come together in an optical system.

forbidden lines spectral lines not normally observed in the laboratory because the transitions causing them are improbable.

Foucault pendulum a massive pendulum bob on a long cord used to demonstrate the rotation of the earth.

free-free transition a change in energy level due to an encounter (but not capture) between an

atom (or ion) and an electron; bremsstrahlung arises from such transitions.

fringes the alternate dark and light lines caused by interference of electromagnetic radiation.

galactic cluster an open cluster located in the Galaxy.

galactic coordinates latitude and longitude using the Milky Way as the great circle forming the galactic equator.

galaxy an independent assemblage of billions of stars, an island universe.

Galaxy the Milky Way galaxy, the galaxy to which our sun and all the lucid stars belong.

GC—the General Catalogue a monumental catalogue of the positions and motions of the brighter stars compiled by Lewis Boss.

gegenschein "counterglow," a hazy diffuse patch of light, several degrees in extent, opposite the sun in the sky; perhaps similar in cause to the heilegenshein.

giant a star of large size and luminosity.

gibbous phase the phases of the moon or planets where more than half the surface is illuminated but not all.

globular cluster a spheroidal distribution of stars located in the halo regions of galaxies.

globule a very small dark nebula.

granules the small scale cells in the solar photosphere giving it the mottled appearance.

grating a ruled optical surface that separates light into its component frequencies by diffraction. It may be a reflection or transmission device. If it is very coarse and placed in front of the objective it becomes an objective grating.

gravitation the property of matter to attract other matter.

gray atmosphere a model atmosphere constructed under the assumption that absorption is independent of wavelength.

great circle the trace formed on a sphere by its intersection with a plane passing through its center.

greatest elongation generally the largest difference in longitude between the sun and an interior planet.

Greenwich meridian the origin for longitude by common usage and agreement.

Gregorian calendar the present calendar, it was introduced by Pope Gregory XIII, hence its name.

H line the intense Fraunhofer line of ionized calcium at λ3968A.

H-I region a cloud of neutral hydrogen in space.

H-II region a cloud of ionized hydrogen in space.

halo 1) a ring around the sun or moon caused by refraction by

water droplets or ice crystals. 2) the large spherical distribution of stars encompassing a spiral galaxy, usually the regions above and below the disk of a galaxy.

Harmonic law Kepler's third law, $a^3 = P^2$ where a is the distance from the sun in AUs and P is the sidereal period in years.

harvest moon the rising of the full moon with least delay. This occurs in late September or early October, hence the name.

HD the Henry Draper Catalogue, a monumental catalogue of star positions, magnitudes and rough spectral types. The HR is a revision to the HD.

heilegenschein a diffraction-reflection effect seen directly opposite the sun around the shadow of an airplane or in early morning or late evening on a moist putting green.

heliacal rising the first rising of a star after being invisible due to its proximity to the sun. The similar condition for setting is referred to as the heliacal setting of a star.

helio prefix referring to the sun.

heliostat a coelostat used to observe the sun.

Hertz a cycle per second, a megahertz is thus a million cycles per second.

Hertzsprung gap the region in an H-R diagram of a globular cluster around zero absolute magnitude where no stars seemed to be present.

H-R diagram the Hertzsprung-Russell Diagram, a plot of luminosity against spectral type (color).

high velocity star a star with a high velocity with respect to the sun, these stars and other high velocity objects actually reflect the sun's rather high velocity.

horizon the place where the sky and earth seem to meet; see celestial horizon.

horizontal branch stars in an H-R diagram of a globular cluster with a range in spectral types but all having zero absolute magnitude.

horizontal parallax the difference in the viewing angle of an object as seen from the center of the earth and from the equator when the object is on the horizon.

hour angle the angle from the meridian to an object's hour circle given in units of time, designated H.A.

hour circle the great circle in the equatorial coordinate system containing the poles and an object in question.

Hubble constant the constant relating the receding galaxies to their distance, designated H and having a value of nearly 100 km/sec/megaparsec.

Hunter's moon the first full moon after the Harvest moon, it also rises with little delay hence the name.

IC the Index Catalogue, a catalogue of clusters, nebulae and galaxies supplementing the NGC.

image optical formation of an object as seen through an optical system.

image tube the name given to a whole group of electronic imaging devices using a photocathode and producing an image in one form or another. A television camera is such a tube. An image tube having one or more electron multiplication stages is called an image intensifier, a tube having a cathode sensitive to certain wavelengths but producing an image at a different wavelength is called an image converter.

inclination the angle between the plane of an orbit and some fundamental plane such as the plane of the sky or the eliptic plane.

inferior conjunction when an inferior planet has the same longitude as the sun and is between the sun and the earth.

inferior planet a planet whose mean distance from the sun is less than one AU.

inner planet planets nearer the sun than Jupiter.

insolation a contraction of incoming solar radiation, the sun's radiation received at the surface of the earth.

intensity energy flowing through unit area in unit time.

interferometer a device using the principle of interference to extract information from incoming radiation.

International Date Line by common usage and consent an arbitrary line roughly 180° from the Greenwich meridian across which the date changes by a day.

international magnitudes an early system of magnitudes based upon the photographic plate and adopted by international agreement. Now largely superceded by the U, B, V and u, v, b, y systems.

interplanetary medium the distribution of dust and gas in interplanetary space.

interstellar medium the distribution of dust and gas in interstellar space.

interstellar reddening reddening of starlight due to scattering by dust particles in the interstellar medium.

irregular galaxy a galaxy lacking a reasonably symmetrical shape.

irregular variable a star or object whose varying energy output is not periodic.

island universe see galaxy

isostasy universal equilibrium in the earth's crust.

isotropic the same in all directions.

Jovian planet one of the large gaseous planets of the solar system.

Julian Calendar a solar calendar introduced by Julius Caesar and a big improvement over its predecessor.

Julian Day a running number for days beginning in 4713 BC. The starting point is arbitrary and was chosen to be early enough that all astronomical events can be given a Julian day number regardless of the calendar in use.

K line one of the Fraunhofer lines due to ionized calcium at λ3933A.

kinetic temperature temperature required to reproduce the velocity distribution observed in a gas.

Kirkwood's gaps gaps in the spacing of the asteroids due to gravitational perturbations primarily by Jupiter.

Lagrangian points five points in the plane of a binary system where the forces on an infinitesimal test mass balance out to zero.

latitude the angular distance above (+) or below (−) the equator.

leap year that year (divisible by four) every four years when an extra day is added to February to keep the civil calendar in step with the sun. Century years are not leap years.

light curve a plot of the variation of light (intensity or magnitudes) against time; also, the composite graph of intensity against the period.

light year the distance light travels in one year; 9.46×10^{12} kilometers or 6.324×10^{4} AUs.

limb the edge of the sun, moon, or planet.

limb darkening the limb or edge of the sun or a star is darkened due to the line of sight passing through thicker cooler layers of the atmosphere.

limiting magnitude the faintest magnitude observable with a given instrument under given conditions.

line broadening spectral lines having widths greater than their natural width due to various causes.

line profile a tracing of a spectral line giving intensity as a function of wavelength.

Local Group the seventeen or so nearest galaxies (including the Milky Way Galaxy) seemingly forming a group.

local apparent time the hour angle of the true sun.

local mean time the hour angle of the mean sun.

long period variable periodic variable stars with periods longer than 70 days.

LTE local thermodynamic equilibrium, an approximation to stellar conditions used in the construction of model atmospheres.

luminescence a visible glow from a material induced by radiation falling on it and persisting after the radiation is removed, a phosphor glows by luminescence.

luminosity class a classification by brightness within a spectral type.

luminosity function the number of stars per unit volume for various absolute magnitudes.

lunar eclipse an eclipse of the moon

magnifying power the apparent size of an object seen through the telescope compared to its size with the unaided eye.

magnitude astronomer's term for a measure of the flux received from a star or other object.

main sequence the diagonal concentration of stars in the H-R diagram.

mare term applied to the sea-like features on the moon and Mars.

mass function for single line spectroscopic binaries the ratio of the cube of the product of the mass of the second body times the sine of the inclination of the orbit to the sum of the masses squared.

mass-luminosity relation an empirical relationship between mass and luminosity that holds primarily for main sequence, giant and subgiant stars.

mass-radius relation a relation between mass and radius holding for main sequence stars.

mean solar day the day determined by successive meridian crossings of the mean sun.

mean sun the hypothetical sun that moves uniformly along the celestial equator.

megaparsec a million parsecs.

meridian 1) the great circle on the surface of the earth containing the observer and the poles.
2) the great circle on the celestial sphere containing the observer's zenith and the celestial poles.

Messier Catalogue an early catalogue of nebulae and galaxies designed to aid comet hunters; objects in the catalogue are designated by M and a running number such as M31.

micrometeorite an extremely small solid particle in space.

microphotometer a device for measuring the photographic density on a plate; often called a microdensitometer.

microwave radio radiation having wavelengths between 1 mm and 30 cm.

Milky Way the hazy band of light visible on a dark night sky.

Mira variable a red giant long period variable.

MKK classification a method of spectral classification devised by

W. W. Morgan, P. C. Keenan and E. Kellman.

modulus see distance modulus.

model atmosphere a juggling of temperature, surface gravity and chemical composition to reproduce the observed radiation of a given star.

monochromatic literally, a single color.

N30 a catalogue of star positions observed by the U.S. Naval Observatory, epoch 1930.

nadir the point opposite the zenith.

nautical mile the mean length along the meridian subtended by one minute of arc as seen from the center of the earth.

neap tide the lowest tide in a given month.

nebula a cloud of interstellar gas or dust or both.

new moon when the moon and sun are lined up on the same side of the earth; if this occurs near a node of the moon's orbit a solar eclipse can occur.

NFK—Neue Fundamental Kalalog one of the earliest modern "fundamental" catalogues, epoch 1870 and 1900.

NGC New General Catalogue, a catalogue of clusters, nebulae and galaxies with epoch 1888.

nodal month the period of the moon around the earth from one node back to the same node.

nodal year the period of the earth around the sun with respect to the line of the node's in the moon's orbit.

node the intersection points of an orbit where the orbital plane intersects some other plane.

non-gray atmosphere a model atmosphere constructed by letting the absorption coefficient vary with wavelength.

north polar sequence stars near the north pole having calibrated and standardized magnitudes.

nova a star that undergoes a sudden violent event increasing its brightness temporarily by many magnitudes.

north celestial pole the point on the celestial sphere in the zenith of the earth's north pole.

north galactic pole the pole of the galaxy, 90° from the galactic equator, nearest the north celestial pole.

nucleus the central part of an atom, a comet, or a galaxy.

nutation the nodding of the earth's axis due to the precession of the axis and the lunar attraction on the earth's tidal bulge.

objective the main lens of a refracting telescope or the primary mirror of a reflecting telescope.

objective grating see grating.

objective prism a large diameter small angle prism placed in front of the objective to give low dispersion spectra of all the stars in the field.

oblateness a measure of the flat-

tening of a spheroid into an oblate spheroid.

obliquity of the ecliptic the inclination of the earth's equator to the plane of its orbit.

obscuration absorption by interstellar dust.

occular an eyepiece for a telescope.

occultation eclipse of a smaller object by a larger object (linear or angular as seen by the observer).

Oort's constants constants that characterize the rotation of the galaxy as seen from the sun.

opacity the property of stopping the passage of light rays.

open cluster a loose clustering of tens to thousands of stars lacking a symmetrical form.

opposition the moon or a planet opposite the sun as seen from the earth.

optical depth a measure of the absorption of radiation in a given layer of material.

optical double star a chance alignment of two stars not physically connected.

Palomar Sky Survey more correctly The National Geographic-Palomar Sky Survey, a monumental photographic atlas of the heavens in two different colors covering the celestial sphere from the north pole to about −30° declination using the 48-inch (121.92 cm) Schmidt.

parallactic ellipse the reflex of the earth's orbital motion in the position of a relatively nearby star.

parallax the apparent displacement of an object as seen from the two ends of a baseline.

parsec a parallactic second; the distance at which one AU subtends 1 second of arc; 3.26 light years.

partial eclipse an eclipse in which the eclipsing body does not completely cover the body being eclipsed.

peculiar motion the residual motion, peculiar to the object being studied, after the common motion has been removed.

peculiar spectrum a spectrum that does not fit conveniently into the standard spectral classes.

penumbra the partial shadow region of an opaque body.

penumbral eclipse an eclipse of the moon when the moon only passes through the earth's penumbral shadow.

peri the nearest point in an orbit to the primary, for example, perihelion is the closest point to the sun for a body orbiting the sun.

period-luminosity relation an extremely useful empirical relation between the periods and brightness of cepheid variable stars.

periodic comet a comet that returns to the vicinity of the sun on a more or less periodic basis.

perturbation the disturbance of

the motion of one body by another.

phase the particular aspect of the moon, the fractional part of a periodically occurring phenomena.

phase angle the angle measured from some zero point of a periodically occurring phenomena.

photoelectric photometer a device for measuring incident radiation using a photocell or a photomultiplier tube.

photographic magnitude the magnitude of an object as measured on a normal panchromatic plate.

photometer a device for measuring the brightness of a celestial object.

photosphere the bright surface of the sun.

photovisual magnitude the magnitude of an object as measured on a photographic plate dyed to have a response nearly reproducing that of the average eye.

PKS a prefix followed by a set of numbers indicating an object in the Parkes Catalogue of Radio Sources.

plage bright regions on the surface of the sun as seen in monochromatic light.

planetarium a complex device for projecting a representation of the night sky and its phenomenon onto a dome; loosely, the institution housing and using such a projector.

planetoid a minor planet (asteroid).

planetesimal hypothesis a theory holding that the planets formed from material pulled out of the sun.

plasma a gas consisting of ionized atoms and/or electrons.

polar axis the axis of a telescope pointing parallel to the earth's axis.

population to astronomers, a graded classification of stars by their chemical abundances, velocities and locations in the Galaxy.

position angle the angle measured from north to a line joining the primary and secondary star in a binary system.

precession 1) the slow conical motion of the rotation axis due to external torques acting on the rotating body.
2) the precession of the equinoxes.

primary minimum the deepest minimum in the light curve of an eclipsing binary.

prime focus the focal point of the primary mirror of a reflecting telescope.

prime meridian see Greenwich meridian.

primeval atom a single ball of energy that exploded to become the Universe.

prominence a solar flare occuring

on the limb of the sun and appearing as a tongue of flame.

proper motion the angular motion per year of an object across the celestial sphere.

protostar condensed material forming a star, but not yet clearly a star.

pulsating variable a periodic variable star changing brightness primarily due to pulsation.

pyrheliometer a device for measuring incident solar rediation.

quadrature two bodies 90° apart as seen from the earth.

quarks hypothetical elementary particles having fractional electric charges.

quasar a sort of acronym for quasi stellar objects.

RR Lyrae variable see cluster variable.

radial velocity relative velocity along the line of sight.

radiant the point on the celestial sphere from which a meteor shower seems to radiate.

radiation pressure pressure exerted by light or other radiation upon an object.

recurrent nova nova that erupt more than once.

red giant a large low temperature star having high luminosity.

redshift 1) the Doppler effect due to a receding object.
2) the Einstein effect due to a massive body.

reddening reddening of starlight caused by interstellar scattering.

regression of the nodes motion of the nodes of an orbit along a fundamental plane due to a perturbation of the orbiting body.

resolution the ability of an optical system to resolve details.

retrograde motion the apparent backward motion of the planets with respect to the stars.

reversing layer the region in the solar atmosphere where the absorption lines appear.

Roche's limit the smallest distance at which a satellite can hold together under the tidal forces created by the primary.

Russell-Vogt theorem a statement that the composition and mass of a star uniquely determine its entire structure.

s-process a slow time scale neutron capture process for building higher mass elements.

saros a cycle of similar eclipses, recurring at 18 year intervals.

scale for a telescope the angle encompassed by one millimeter at the focus.

scale height the height at which a given parameter falls to a certain value, for example one scale height in density is when the density in the atmosphere falls to 1/2.72 of its value at the surface.

secondary minimum the shallow-

est of the two minima in an eclipsing binary light curve.

secular nonperiodic

secular parallax change in the apparent direction to a nearby object due to the motion of the sun through space.

seeing blurring of an image due to the unsteadiness of the earth's atmosphere.

selected areas areas on the celestial sphere in a plan proposed by Kapteyn for studying galactic structure.

seleno prefix from the Greek, meaning the moon.

semiregular variable a variable star having a periodic light curve of varying amplitude or having an almost regular period.

separation the angular distance between two objects.

shell star a star having a thin detached shell of gas.

sidereal period time of intervals with respect to the stars.

small circle any circle on a sphere that is not a great circle.

solar antapex the direction to which the stars are moving, or from which the sun is moving.

solar apex the direction from which the stars are moving, or to which the sun is moving.

solar constant the amount of radiation from the sun falling upon unit area per unit time at the distance of the earth.

solar motion the motion of the sun with respect to the solar neighborhood.

solar parallax ½ the mean angle subtended by the earth as seen from the sun.

solar wind the outward flow of particles from the sun.

solstice the points on the ecliptic farthest from the equator.

space motion the velocity and direction of motion of a star as seen from the sun.

spectral type the classification of a star according to certain spectral features.

spectral sequence arranging spectral types by temperature.

spectrograph a device for recording the spectrum of an object.

spectroheliograph a device for taking pictures of the sun in monochromatic light.

spectroscopic binary a binary system revealed by the periodic shifting of its spectral lines.

spectroscopic parallax assigning a star an absolute magnitude according to its spectral type and measuring its apparent magnitude allowing one to compute the distance to the star.

spectrum the energy output of an object displayed as a function of wavelength (or frequency).

specular reflection reflection from a polished surface such as a mirror.

spectrum-luminosity diagram see *H-R Diagram.*

spectrum variable a star having a variable spectrum due to intrinsic properties of the star, such as a changing magnetic field.

spherical aberration one of the major defects in an optical system caused by the failure of light rays from different radii from the optical axis to come to a common focus.

spicule a narrow jet in the solar chromosphere; the jets give a grass like appearance to the limb of the sun.

spiral an abbreviated term for spiral galaxy, a galaxy having spiral arms.

spiral arms the concentration of gas, dust and bright, hot stars in the disk of a galaxy having a spiral form.

spring tide the highest tide of the month.

standard time the local mean solar time at a standard meridian, adopted over a zone for civil convenience.

star cluster a grouping of stars more or less gravitationally bound.

starquake hypothetical surface explosions causing a neutron star to quiver.

steady state 1) a state of equilib- 2) a cosmological theory.

stellar model theoretical calculations varying the chemical composition of an ideal star to produce a model whose properties closely approximate those of a star.

stellar parallax the angle subtended by one AU at the distance of the star being studied.

Strömgren sphere the spherical region of ionized gas around a bright, hot star.

subdwarf a star whose luminosity is less than that of a main sequence star of the same spectral type.

subgiant a star lying between the normal giants and the main sequence in the H-R diagram.

summer solstice the point on the ecliptic where the sun is farthest north.

supergiant an extraordinary bright star of immense size.

superior conjunction when a planet and the sun have the same longitude with the sun being between the planet and earth.

superior planet a planet whose semimajor axis is greater than one AU.

surface gravity the acceleration due to gravity at the surface of a body; $g = GM/R^2$ where M is the mass of the body, R is its radius and G is the universal constant of gravity.

symbolic velocity the velocity of a galaxy as given by the redshift.

synchrotron radiation radiation from high velocity electrons (velocities approaching the velocity of light) spiraling in a magnetic field.

synodic month the revolution of the moon with respect to the sun.

synodic period the time between successive planetary oppositions or conjunctions of the same kind (e.g. superior conjunction to superior conjunction).

syzygy when the earth, moon and sun lie on the same line; more correctly stated when the centers of the earth, moon and sun lie in a plane vertical to the plane of the earth's orbit. An excellent word-game word.

T Tauri stars variable stars whose variability has been associated with the interstellar medium. Their variations are erratic and may be quite rapid. The stars are believed to be premain sequence stars.

telluric terrestrial in origin.

temperature a measure of the internal energy of a body. Astronomers use many different measures of temperature, e.g. color temperature, brightness temperature, ionization temperature, etc.

temporary star term given to a nova or supernova which appears visible to the naked eye for a brief period of time: a misnomer.

terminator the line of sunset or sunrise on the earth, moon or planets.

theory one or more hypotheses which along with established laws explains a phenomena or class of phenomena.

thermodynamics a branch of physics dealing with the study of heat and heat transfer.

tidal hypothesis a theory explaining the origin of the planetary system, currently rejected.

Titius-Bode Relation *see* Bode Relation.

total eclipse an eclipse where the light of one body is completely cut off by an intervening body.

totality the period of complete light cut off during an eclipse.

transit 1) a telescope for measuring the meridian passage of a star or planet.
2) passage of a smaller body across the face of a larger body.

transverse velocity motion at right angles to the line of sight usually given in kilometers per second.

3C catalogue the third Cambridge catalogue of radio sources.

Trojan the minor planets (asteroids) located approximately 60° ahead and behind Jupiter as seen from the sun.

tropical year the interval of the earth's revolution about the sun with respect to the vernal equinox.

Tychonic system a cosmology proposed by Tycho Brahe.

U, B, V system a photometric system defined by three specific filters (roughly-ultraviolet, blue and visual)

u, v, b, y system a photometric system defined by four specific filters having narrower passbands than the U, B, V system filters.

U.L.E. a Corning Glass Co. trade name for a fused quartz glass with a very small expansion coefficient (the letters mean ultra low expansion).

umbra 1) the central part of a shadow where, geometrically, the light from the source is cut off.
2) the dark central area of a sunspot.

universal time the local mean time of the prime meridian.

variable star any star that varies noticeably in brightness.

variation of latitude changes in latitude of places, on the earth due to the shift of the earth's axis within the earth.

velocity of escape the velocity required to escape from the gravitational field of a body; $v = [2G(M + m)/r]^{\frac{1}{2}}$ where M is the mass of the body, m is the mass of the object escaping and r is the distance from the center of the body.

vertical circle a great circle perpendicular to the horizon passing through the zenith.

visual binary a physical double star resolved by the eye at the telescope.

visual photometer an instrument attached to the telescope for measuring brightness visually.

VLA initials for the very large array of radio telescopes proposed for long baseline interferometry and aperture synthesis.

VLB initials meaning very long baseline radio interferometry.

Vulcan a hypothetical planet nearer the sun than Mercury.

W Virginis star a cepheid of population II.

wandering of the poles the shifting of the earth with respect to its axis.

weight a measure of the attraction of a body on a given mass.

white dwarfs faint stars usually having high surface temperature and small diameters, it lies well below the main sequence in the H-R diagram.

Winter solstice the point on the ecliptic where the sun is farthest south from the equator.

Wolf-Rayet star a class of extremely hot stars that eject shells.

World Calendar a calendar proposed to have the same dates fall on the same day every year.

year the period of the earth's revolution around the sun.

ZAMS zero age main sequence.

zenith the point overhead on the celestial sphere, the extention of a plumb line to where it pierces the celestial sphere.

zenith distance the angular dis-

tance along a great circle from the zenith to another point on the celestial sphere.

zodiac an imaginary band sixteen degrees wide centered on the ecliptic.

zone of avoidance an irregular band along the Milky Way devoid of galaxies due to obscuration by interstellar gas and dust.

zone time standard time used in a band of longitude roughly 15 degrees wide.

NAME INDEX

SUBJECT INDEX

(italicized listings appear in the glossary)

C0518-000-9

Baker and Fredrick

ASTRONOMY

NINTH EDITION

DA BAK

VNR

VAN NOSTRAND REINHOLD