Evry L. Schatzman
Françoise Praderie

The Stars

Springer-Verlag

ASTRONOMY AND ASTROPHYSICS LIBRARY

Evry L. Schatzman   Françoise Praderie

The Stars
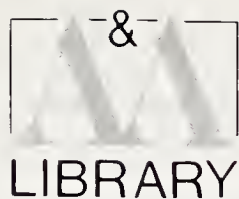
Translated by A. R. King

With 143 Figures

Professor Evry L. Schatzman

Observatoire de Paris, Département d'Astrophysique Stellaire et Galactique, 5, place Jules Janssen, F-92195 Meudon, France

Dr. Françoise Praderie

Observatoire de Paris, Département de Recherche Spatiale, 5, place Jules Janssen, F-92195 Meudon, France

Translator

Professor Andrew R. King

University of Leicester, Astronomy Group, University Road, Leicester LE1 7RH, United Kingdom

Series Editors

Martin Harwit

The National Air and Space Museum
Smithsonian Institution
7th St. and Independence Ave. S.W.
Washington, DC 20560, USA

Virginia Trimble

Astronomy Program
University of Maryland
College Park, MD 20742, USA
and Department of Physics
University of California
Irvine, CA 92717, USA

Rudolf Kippenhahn

Rautenbreite 2
W-3400 Göttingen
Fed. Rep. of Germany

Jean-Paul Zahn

Observatoire Midi-Pyrénées
14, Avenue Edouard Belin
F-31400 Toulouse, France

Cover picture: The Pleiades, an open star cluster consisting of more than a hundred young stars, 5×10^7 years old, lies at a distance of 130 parsec. The brightest stars emit strongly in the ultraviolet (UV), here in the wavelength range 1150–2600 ångström. This false-colour image was taken as part of the UV calibration of the *EXOSAT* satellite telescopes. The field is about 2° by 2°; north is to the right of the figure. (Reproduced with the kind permission of the European Space Agency.)

Preface

The year 1996 will see the seventieth anniversary of the publication of *The Internal Constitution of the Stars* by Sir Arthur Eddington. This book remains a masterpiece, a classic in the astrophysical literature. In essence it proved that it was possible to construct a theory that made it possible to look inside the stars. However, theoretical astrophysics remained the province of a very small number of people for many years. Quite a few remarkable textbooks have been published since, but there always remained the need to have a book, however incomplete, that took into account the fact that the study of the stars raises global problems. Whichever part of a star is studied, that part is of course connected to the rest of the star. Furthermore, all fields of physics are involved. The incredible increase in precision of astrophysical data and the wide variety of new types of observation (from neutrinos to asteroseismology) obliged researchers to take into account new physical processes and to improve their knowledge of basic physical processes (the equation of state, opacities, nuclear cross-sections, plasma physics and magnetohydrodynamics, to name just the most important).

Remember that astrophysics is first and foremost a science of observation. But theoretical developments are essential if we are to understand the observed phenomena. Furthermore, that understanding is necessary if we are to move forward, to unravel the discrepancies between the observational data and the theoretical models, and if we intend to build new instruments that will help us progress in our collection of critical data.

The spirit of this book is reflected in what we have just written. Let us repeat, though, that in astrophysics everything is linked together, and that at present our knowledge of stellar properties is absolutely fundamental to our undertanding of the universe.

We have taken the opportunity of this English translation to thoroughly revise and update the text.

Meudon, October 1992

Evry Schatzman
Françoise Praderie

V

Preface to the French Edition

In 1959 there appeared the book *Astrophysique Générale* by Pecker and Schatzman. It is a little difficult today to realise the stages that knowledge of the stars have passed through in thirty years, whether in respect of observational data or development of theoretical results. In the domain of stellar astrophysics alone, major observational discoveries have been made: neutron stars, through the pulsars (1967), the solar-neutrino deficit following the theoretical prediction (the 1970s), data from helioseismology (after 1980), and Supernova 1987A in the Large Magellanic Cloud and its neutrinos. The existence of neutron stars had been predicted by Landau (1932), but observation of them was nevertheless a considerable scientific event!

In thirty years, the volume of publications in astrophysics has increased more than ten-fold. To make a choice and write a book designed for students which will accurately reflect our knowledge of the stars presents considerable difficulties. The sensitivities of the authors and their perception of the importance of this or that topic and its future have played a role at least as great as acquaintance with scientific currents through the meetings, colloquia, conferences, and symposia which animate the subject. Only the coming years will allow one to say if the desired aim has been reached.

Acknowledgments

Contents

Introduction

In the splendour of a moonless night, far from the pollution of the sky by artificial lighting, the first revelation is that of the stars. The Milky Way, crossing the whole sky, is easily resolved into stars, even with a very small instrument. In the 18th century Herschel (1738–1822) discovered objects then called *nebulae* whose exact nature as gigantic star systems was not definitively established until 1923, when it became possible to find their distances. During this time the Milky Way had been recognised as a vast flattened system, containing billions of stars. The position of the Sun, far from the centre of this system, in all respects analogous to the *extragalactic nebulae*, was established by Shapley (1915). Today this term is obsolete; we speak of galaxies (with a small g) and we belong to the Galaxy (with a large G).

Within the galaxies, as in our Galaxy, we find *stars* and *interstellar matter*. The stars make up 90% of the mass of the galaxies and are the essence of the visible Universe, even if invisible *dark matter* (whose existence is suspected but not yet truly established) represents in its turn perhaps 90% of the mass of the Universe. Whatever the case, the study of the stars, their structure, and evolution is the key to a large number of questions which we may ask ourselves about the Universe. Without studying the stars it is impossible to understand the chemical composition of our Galaxy and of the Universe and the constitution of the galaxies, and thus to enter the immense domain of cosmology.

The basic data are clearly those one obtains from observations of individual stars: (1) for each star under study, we determine the spectral properties of its radiation in all possible wavelength domains; (2) once we have found the distance of a star from the Earth we know the power it radiates; (3) by a simple application of the law of universal gravitation to the Keplerian motion of the two component stars of a double star we find their masses and we can then associate masses and radiation properties; we discover very quickly that there is no one-to-one relation between these properties because of the role played by a further parameter which we identify as the age of the star; (4) the measurement of the radii of the stars is a more delicate process. The orbital plane of certain double stars contains the Earth, so we see two eclipses per period of one component star by the other. Called eclipsing variables, these double stars directly give values of the radii whose

interpretation is often difficult. For several tens of nearby stars the radius can be obtained by interferometric measurements. But the determination of the majority of stellar radii is indirect and consequently model-dependent.

The essence of the study of the stars is immediately seen to be the theoretical analysis of the observed properties. The principle of this analysis is after all no more than the rigorous application of almost all the laws of physics. Simple models can be developed, leading to elementary statements describing properties for which evidence appears only after the fact. It is quite remarkable that these models also give a realistic picture of the main physical processes at work in the stars. In contrast, quantitative agreement is only possible by means of elaborate calculations, using the best and most complete physical data. Despite some very brilliant results, there are numerous cases where there remain disagreements whose origin is not understood between the predictions of the models and the values of the observed quantities: the future of the theory of stellar interiors remains open, like that of the theory of the outer layers.

The study of the surface regions allows in the first place a determination of the abundances of chemical elements in the stars. While a first approach to the question reveals the difference essentially between old stars (poor in metals) and young stars (rich in metals), detailed analysis shows that all stars (almost!) are different. The differences may be caused by anomalies in the region of formation or by the evolutionary history of the star. The classification into different types and varieties allows one to bring out the phenomena believed to be the cause of these differences: transport processes under the effect of forces applied to atoms (gravity, radiative forces), diffusion, and nuclear reactions.

Two phenomena play an important role in the structure and evolution of the stars. On the one hand, the broadening of spectral lines by the Doppler effect demonstrates to us that the stars *rotate*. We cannot list here all the numerous effects of rotation, one of the most important being perhaps the origin of a whole series of hydrodynamical instabilities capable of modifying the star's structure in an appreciable fashion. On the other hand, the profile of spectral lines reveals the presence of an astonishing *stellar wind*, corresponding to a remarkable rate of mass loss, itself capable of having an enormous repercussion on the star's evolution.

A star is an object out of equilibrium, which evolves in an irreversible fashion since photons escape it permanently, as does a flux of matter (the stellar wind). We can single out a certain number of characteristic timescales, which we evaluate here for the Sun: the time of sound propagation radially from one side of the star to the other (of the order of 30 minutes), the time for radiative transport of heat from the centre to the surface (of the order of one hundred thousand years), the time to radiate the gravitational energy (or Kelvin-Helmholtz time, of the order of ten million years), the nuclear timescale (consumption of a certain fraction q of the nuclear energy reserves; for $q = 0.1$, the nuclear timescale of the Sun is

about ten billion years). Depending on the timescale of a phenomenon, one or the other of these processes plays a dominant role.

The Sun, which we have just taken as an example, is only one of many similar stars. A single chapter is devoted to it in this book, where we stress the solar phenomena that are better known because of this star's being close to the Earth. All the properties of the Sun which are common to other stars are treated in the course of the book.

Significant magnetic properties are observable in the outer layers of the stars, without doubt leading to complex hydrodynamical and magnetohydrodynamical phenomena. We summarise these properties by saying that they are caused by the famous dynamo effect, associating rotation and turbulence. Here too, the incontestable successes of mean field electrodynamics are accompanied by enormous unresolved difficulties. This does not prevent us recognising the role of the magnetic field in stellar activity, in the loss of angular momentum and braking of rotation, and in mass loss by a stellar wind.

The nuclear timescale plays a primary role. Stellar evolution allows us to date star systems and to establish a timescale which permits us to estimate the ages of the globular clusters, and to evaluate the age of the Universe. The irreversible evolution of the chemical composition of the stars is accompanied, through the various mass loss processes, by a progressive change of the chemical composition of the Galaxy, which in turn modifies the conditions of star formation, in a remarkable reaction effect of stars–Galaxy–stars.

Studies of stability, quite apart from their intrinsic interest for the understanding of variable stars, link up with current work on nonlinear systems. In sketching this relation briefly we wished to show that the study of variable stars can provide original data on these systems. Conversely, we can hope in the near future that we will be able to deduce unrivalled information about stellar interiors from this analysis.

Binary stars show properties which raise many questions: tidal phenomena, exchange of angular momentum, periastron advance. More spectacularly, the components of close binaries exchange mass between themselves, an inevitable effect of stellar evolution. A great number of unusual objects are thus explicable, of which the most remarkable are the cataclysmic variables, the novae, and probably one class of supernovae. We should perhaps add to this list the neutron stars, whose distribution in the Galaxy can apparently only be explained by the disruption of binary systems at the moment of formation.

The last-born of stellar studies, neutrino astrophysics, with the first flux measurements of solar neutrinos and Supernova 1987A, which are a credit to what is called underground astronomy, and asteroseismology (with helioseismology to the fore) open new paths and pose new questions. Helioseismology gives us direct access to the Sun's core, subjecting the theory of the internal structure to fierce testing. If the solar neutrinos which are detected at the Homestake Gold Mine and Kamiokande are proof of the

thermonuclear reactions which unfold in the centre of the Sun, the deficit of neutrinos (compared with the theoretical prediction) poses the most acute of questions.

The study of the stars is today confronted with progress in the acquisition of data (high signal-to-noise ratio in ground-based and space observation), with new sources of information (stellar seismology, neutrinos) and with results to come from large astronomical projects of all types. A considerable task awaits astrophysicists. However voluminous the present book, it is still only a sketch of the physical basis of our knowledge and of our understanding of the stars; it may nevertheless – let us hope – provide a means of access to an immense world literature and help to realise some progress in stellar astrophysics.

1. Basic Stellar Data

In this book we shall mainly study the individual physical characteristics of the stars, rather than their appearance in vast aggregates as galaxies. We begin by examining the observational data for the stars (Chap. 1) and the Sun (Chap. 2), before considering the structure and evolution of the stars in the subsequent chapters.

The observable properties of a star are its distance d, radius R, mass M, emitted energy per unit time (intrinsic luminosity) L, and the spectral distribution of this energy. We assume for the most part stars in a steady state. The basic source of information for all stars is their radiation; all methods of studying stellar properties such as distance, radius, mass, and energy rely on the measurement and analysis of radiation. We assume that the reader is familiar with the basic definitions of photometric quantities (see for example Mihalas 1978, Chap. 1; Léna 1986, Chap. 3).

1.1 Stellar Distances. Magnitudes

There are two geometric methods giving stellar distances; these are the primary methods. Other (secondary) methods extend these results to many more stars.

1.1.1 Primary Distance-Determination Methods

Trigonometric Parallaxes. Because of the motion of the Earth E around the Sun $\odot$, a star S is seen at intervals of six months at two positions S_1 and S_2 an angular distance 2π apart on the sky (see Fig. 1.1). To an observer on the moving Earth, the star appears to oscillate through a very small angle ($< 1''$ even for the nearest stars) against the background of the "fixed" stars. The angle π subtended by the semi-major axis of the Earth's near-circular orbit at the star S is the star's *trigonometric parallax*.

The parallax π is expressed in seconds of arc. Using the relation

$$1'' = 1/206\,265\,\text{rad} \tag{1.1}$$

and the fact that a is much smaller than $d = \odot S$, we have

$$d = 206\,265\,a/\pi'' . \tag{1.2}$$

Parallactic eclipse

"Sky"

Orbit of the earth
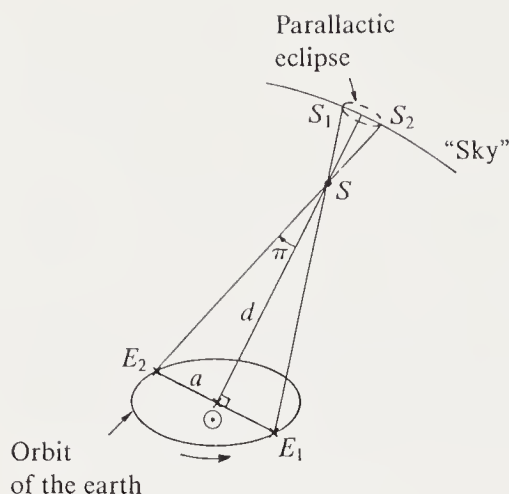
Fig. 1.1. Trigonometric parallax of star S, at distance d from the Sun. The Earth is shown at two positions E_1, E_2 in its orbit, separated by six months

The average Earth–Sun distance, or astronomical unit (AU; $a = 1.495979 \times 10^{13}$ cm), is not appropriate for measuring stellar distances. One uses instead the *parsec*, which is the distance at which the semi-major axis of the Earth's orbit subtends $1''$. We thus have

$$d(\mathrm{pc}) = 1/\pi'' \qquad \text{or} \tag{1.3}$$

$$1\,\mathrm{pc} = 206\,265 \ \mathrm{AU} = 3.086 \times 10^{13} \ \mathrm{km} = 3.26 \ \text{light-years}.$$

Distances in the Galaxy are of the order of kiloparsecs (kpc); for distances between galaxies one uses megaparsecs (Mpc).

The definition of π assumes that the star S is located within a reference system consisting of stars or very distant objects which are not affected by parallax effects. When we measure the parallactic motion of a star from the surface of the Earth we obtain *relative parallaxes*, i.e. the star is related to faint objects in its field.

We can obtain an *absolute parallax* by subtracting from a relative parallax the mean parallax of the distant stars in the fields covered. It is very difficult to determine absolute parallaxes using observations made from the ground; this requires the observation of a large number of standard stars and the fixing of the system's zero point. But the field of photographic plates is always very small. Accumulating several tens of plates with a dedicated telescope over four or five years leads to an accuracy of $0.005''$ in trigonometric parallaxes obtained from the ground. But the resulting parallaxes are affected by systematic errors of the same order of magnitude. About 250 stars have a parallax known to better than 20 %, all of them closer than 10 pc. Table 1.1 gives the parallaxes and distances of the stars closest to the Sun. Column 2 gives their numbers in the catalogue of Gliese (1969). Columns 5 and 6 give the spectral types and visual magnitudes of the stars. These parallaxes are taken from Van de Kamp (1981), who gives a list of the 47 known stars closer than 5 pc.

Table 1.1. The nearest stars

Name		Gliese No.	π''	d (pc)	Spectral type	V
α Cen A		559 A	0.753	1.33	G2V	− 0.01
α Cen B		559 B	0.753	1.33	K5V	1.33
α Cen C (Proxima)		551	0.753	1.33	M5e	11.05
Barnard's star		699	0.544	1.84	M5V	9.54
Wolf 359		406	0.432	2.31	M8	13.53
BD + 36°2147		411	0.400	2.50	M2V	7.50
UV Cet	A	65 A	0.385	2.60	M5e	12.45
	B	65 B	0.385	2.60	M6e	12.95
Sirius (α CMa)	A	244 A	0.377	2.65	A1V	− 1.46
	B	244 B	0.377	2.65	DA	8.98

The launch of the *Hipparcos* satellite (1989) should change the situation appreciably, since it will obtain the parallaxes of 118 000 stars in three and a half years, to an accuracy of 0.002″ for stars brighter than 9th magnitude. We will thus get direct parallaxes for stars with $d \leq 500$ pc. Space observations allow one to overcome the main sources of systematic error affecting parallaxes: the random motion of the Earth's atmosphere, atmospheric refraction and absorption, and flexure of the instruments. Moreover, a major advantage of *Hipparcos* 's optical system is that it gives the parallaxes of stars in both hemispheres in the same system; the fact that we can simultaneously measure stars separated by large angles allows us to get *absolute parallaxes*. The reference frame constituted by the set of *positions* measured by *Hipparcos* will be related to an inertial system for our local volume of space. To do this the *Hipparcos* frame is related to a frame defined by radio sources by simultaneous observations with *Hipparcos* and radio VLBI (very long baseline interferometry). It is also intended to relate the *Hipparcos* frame to that of quasars, by simultaneous observations with *Hipparcos* and the Hubble Space Telescope (in its astrometric mode) of stars at small angular distances from quasars; the position of these stars relative to the quasars is itself measured by very long baseline radio interferometry. Quasars are very distant point sources and define the best absolute reference frame.

The Convergent-Point Method. This method is used for star clusters. It is necessary to measure the two components of the space velocity V (radial velocity V_R and proper motion μ) with respect to the Sun. Figure 1.2 defines V_R and μ for the star S. The heliocentric radial velocity V_R is found by using the Doppler effect for the stellar spectrum, after correcting for the Earth's orbital motion (29.8 km s^{-1}) and its rotation (0.5 km s^{-1} at the equator); the proper motion μ is the angle subtended by the tangential component of V; it measures the secular change of the stars's position on the sky. Like parallaxes, proper motions are relative if they refer to distant stars in the
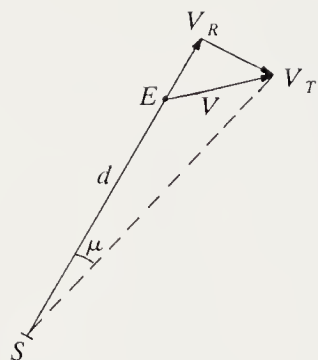
Fig. 1.2. Space velocity V, radial velocity V_R and proper motion μ for star S. The motions are relative to the Sun

field observed or absolute if they have been corrected for the mean proper motion of the reference stars.

For certain clusters the space velocities V are all directed at the same point in the sky, called the *convergent point*. Let λ be the angle between the direction of this point and the direction $\odot S$. Assuming that the cluster is not contracting, expanding, or rotating and the stars in it have no residual random motions with respect to the convergent point, the distance d of any of the stars from the Sun is given by

$$\mu = \frac{V \sin \lambda}{d} \quad \text{and} \quad V_R = V \cos \lambda, \tag{1.4}$$

giving

$$d = \frac{V_R \tan \lambda}{\mu}. \tag{1.5}$$

(In (1.4) and (1.5) μ is expressed in rad s^{-1}.)

We then assume that the distance d is the same for all the stars of the cluster (which is not correct for the nearest clusters) and obtain d as the mean distance of all the stars in the cluster for which μ and V_R are known to good accuracy ($0.005''$ per year for μ in the northern hemisphere, 1 km s^{-1} for V_R).

This method has been used for the clusters given in Table 1.2.

Table 1.2. Distances of three clusters found by the convergent-point method

Cluster	Number of stars	d (pc)
Hyades	200	42
UMa	60	21
Sco Cen	100	170

1.1.2 Secondary Distance-Determination Methods

Statistical and Secular Parallaxes. As a preliminary to describing this method, consider a group of stars in the solar neighbourhood whose heliocentric velocities V have been measured. In reality one measures the radial velocity V_R and proper motion μ for each star as we have seen above. The tangential velocity V_T cannot be obtained unless d is known; let us assume that this is so. We note the basic fact that heliocentric space velocities V may equally well be regarded as describing the motion of the Sun with respect to a group of stars as the motion of the stars with respect to the Sun. We thus seek *the motion of the Sun in the Galaxy with respect to the velocity centroid of the group of stars*. The principle of this determination is given here only schematically.

(a) We define a system of *galactic* coordinates at a point S of the galactic plane at a distance r from the centre: one axis (Π) is directed towards the galactic anticentre, one (θ) in the direction of the galactic rotation, and the other perpendicular to the galactic plane (Z).

(b) A star situated at S and following a circular orbit in the galactic plane has coordinates $\Pi = 0, \theta = \theta_c, Z = 0$.

(c) We define a second system of coordinates called the *local standard of rest* (LSR), with axes parallel to the first and origin S moving as described in (b). With respect to this system (LSR), a star has coordinates $u = \Pi, v = \theta - \theta_c, w = Z$. u, v, w, are the components of what is called the *peculiar velocity* of the star, $\mathbf{S}$.

(d) The observed heliocentric velocity of a star $\mathbf{V}(U_*, V_*, W_*)$ can be written as

$$\mathbf{V} = \mathbf{S} - \mathbf{S}_\odot.$$

(e) A group of N stars is studied to find the motion of the Sun relative to them. If N is large, $\langle u_* \rangle = 0, \langle w_* \rangle = 0$, where $\langle\ \rangle$ denotes the mean over N. $\langle v_* \rangle$ does not vanish, because of the differential motions in the Galaxy, but we may assume it to at this point. Thus the components of $\mathbf{S}_\odot$ are

$$u_\odot = -\langle U_* \rangle, \qquad v_\odot = -\langle V_* \rangle, \qquad w_\odot = -\langle W_* \rangle.$$

The Sun appears to move towards a point in the sky called the *apex* with a velocity $\mathbf{S}_\odot$.

The coordinates of the apex and $S_\odot$ differ from one group of stars to another; they must be known when using the method of statistical and secular parallaxes. (As an example, the *standard* value of $S_\odot$, defined with respect to the most numerous types of stars in the catalogues of V_R and μ, is 19.5 km s^{-1}.) Similarly the use of this method assumes that there are measured proper motions and radial velocities V_R for the stars whose distance one seeks.
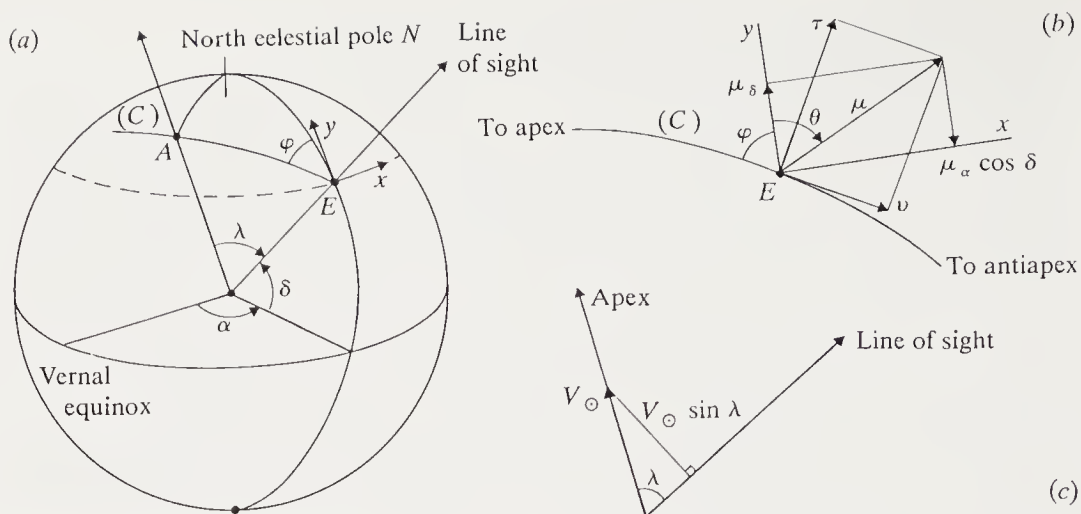
Fig. 1.3. Definition of the quantities used in the method of statistical parallaxes

Figure 1.3 makes these notions precise: Fig. 1.3a shows the star S and apex A on the celestial sphere; Fig. 1.3b shows the great circle (C) passing through the star S and the apex, and the proper motion μ of S. Star S has equatorial coordinates α and δ. The direction of μ is given by the position angle θ. The components of μ in an equatorial system are $\mu_a \cos \delta$ and μ_δ; we denote by τ and v the components of μ along and perpendicular to the great circle (C). In Fig. 1.3c λ is the angular distance between the star and the apex of the Sun, and $V_\odot \sin \lambda$ is the component of the Sun's motion towards the apex normal to the line of sight.

— *Analysis of the components v.* Each measured quantity v is the sum of a stellar term v_* and a solar contribution of modulus $V_\odot \sin \lambda / d$, which is collinear with v. Expressing the proper motions in seconds of arc as is customary, we have for each star

$$v_{\text{mes}} = v_* + \frac{\pi'' V_\odot \sin \lambda}{4.74}. \tag{1.6}$$

The coefficient 4.74 results from the conversion of the parallax π to seconds of arc and assumes that $V_\odot$ is expressed in km s^{-1}. We thus seek the *mean parallax* $\overline{\pi}''$ of the N stars in the group. Noting that we can assume that the motions of the stars are random about their velocity centroid, we get

$$\overline{\pi}'' = \frac{4.74 \langle v \sin \lambda \rangle}{V_\odot \langle \sin^2 \lambda \rangle}. \tag{1.7}$$

— *Analysis of the components τ.* If we also know the heliocentric radial velocity $V_{R,i}$ of the stars, we get

$$\overline{\pi}'' = \frac{4.74 \langle |\tau| \rangle}{\langle V_R + V_\odot \cos \lambda \rangle}. \tag{1.8}$$

10

A detailed demonstration of relations (1.7) and (1.8) is given by Mihalas and Binney (1981).

This method can be used for stars having common properties (for example distant stars of the same spectral type, or a given group of variable stars) provided that their absolute proper motions are accurately known and the group of stars is really homogeneous with respect to the chosen property. It is applicable once genuine trigonometric parallaxes become unreliable ($d \gtrsim 50$ pc) and allows extension of the distance scale out to about 500 pc.

Dynamic Parallaxes. This method applies to stars belonging to a double system (cf. Sect. 1.5). If the orbital elements of the system are known (in particular the semi-major axis a of the orbit and the binary period P), we use Kepler's third law to obtain the distance d of the system through the relation

$$d(\mathrm{pc}) = 1/\pi'' = \frac{(M_1 + M_2)^{1/3} P^{2/3}}{a''} , \tag{1.9}$$

where M_1 and M_2 are the masses of the two stars in solar masses and P is the orbital period in years. This assumes some knowledge of the masses M_1 and M_2. For example we may assume that the spectral type of a star fixes its mass, or use a *mass–luminosity* relation (cf. Sect. 3.4.4). For those few systems where the masses and orbital elements are known this becomes a primary method of distance determination.

Although binary systems are very numerous this method has only been applied to a small number of objects (out to about 200 pc). Conversely the parallaxes supplied by the *Hipparcos* satellite will allow us to deduce masses of binaries from dynamical relations, rather than distances from masses.

1.1.3 The Magnitude Scale

The apparent magnitude of a star measures the amount of energy received from this star per unit surface area and time just outside the Earth's atmosphere. The amount of energy f has the dimensions of an energy flux (erg cm^{-2} s^{-1}). Pogson's law is based on the empirical observation that the eye responds linearly to a logarithmically increasing excitation:

$$m = m_0 - 2.5 \log_{10} f . \tag{1.10}$$

The coefficient 2.5 is chosen so that there are 5 magnitudes between two stars whose received fluxes at Earth differ by a factor 100. If $\Delta m = m_2 - m_1 = 1$, we have $f_1/f_2 = 2.512$. The minus sign has a historical origin, giving a magnitude scale in which smaller numbers correspond to brighter stars; m_0 is a constant. Table 1.1 gives the apparent visual magnitudes of

the stars closest to the Sun. Sirius is the brightest star in the sky, except for the Sun, which has $m_v = -26.74$.

We have just mentioned the apparent *visual* magnitude. m actually depends on the wavelength of the radiation or the passband of the detector which makes the measurement. Star catalogues such as the Henry Draper (HD) or the Bonner Durchmusterung (BD) give photographic (m_{pg}) or visual (m_v) magnitudes, whose reference wavelengths are $\lambda = 4250$ Å and $\lambda = 5480$ Å respectively. In the U, B, V system (see Léna 1988, Chap. 3) the V magnitude corresponds to $\lambda = 5500$ Å.

We define the *apparent bolometric magnitude* m_{bol} with respect to the full range of electromagnetic radiation emitted by a star. This is related to the luminosity L of the star, since if $f(\lambda)$ is the monochromatic flux of energy received at the Earth, then

$$f = \int_0^\infty f(\lambda) d\lambda = \frac{L}{4\pi d^2}.$$

We get

$$m_{bol} = -2.5 \log_{10} L + 5 \log_{10} d + \text{const}, \tag{1.11}$$

a relation which shows that the apparent magnitude of a star depends both on its luminosity and on its distance.

To compare the luminosities of stars it is useful to define an *absolute magnitude* for any star as the apparent magnitude it would have at a distance of 10 pc from the Sun. For the bolometric magnitude, (1.11) gives

$$M_{bol} = -2.5 \log_{10} L + 5 + \text{const}, \tag{1.12}$$

where d is measured in pc, which gives

$$m_{bol} - M_{bol} = 5 \log_{10} d - 5. \tag{1.13}$$

Relation (1.13) holds for all apparent and absolute magnitudes, whatever the wavelength for which the magnitude is defined.

A difficulty in using (1.13) is that one measures the radiation of the stars through differing amounts of absorbing medium between the source and the detector. A general treatment of ground-based photometry is given by Young (1974). We will only point out here that interstellar absorption reduces the amount of energy before the radiation reaches the top of the atmosphere. Let A_λ measure in magnitudes the reduction in the radiation at wavelength λ. Then we always have

$$m_\lambda - M_\lambda = 5 \log_{10} d - 5 + A_\lambda. \tag{1.14}$$

The determination of A_λ is an important part of the observational study of the Galaxy.

12

The quantity $m - M$ is called the *distance modulus* of the object.

We also define the *bolometric correction BC* as the difference between visual and bolometric magnitudes:

$$BC = M_{\mathrm{bol}} - M_v = m_{\mathrm{bol}} - m_v. \tag{1.15}$$

This quantity gives M_{bol} from an observed M_v. Bolometric corrections are known to an accuracy of about 0.1 magnitudes, and vanish for spectral types close to F5V.

In practice one uses the absolute bolometric magnitudes of the stars by referring them to that of the Sun, whose distance and intrinsic luminosity are well known. For the Sun $M_{\mathrm{bol}}(\odot) = +4.75, m_{\mathrm{bol}}(\odot) = -26.82, BC(\odot) = -0.08$. Relation (1.12) becomes

$$M_{\mathrm{bol}}(*) - M_{\mathrm{bol}}(\odot) = -2.5 \log_{10}(L_*/L_\odot). \tag{1.16}$$

The determination of absolute visual magnitudes or intrinsic visual luminosities for the stars depends on a good knowledge of their distances. Conversely, if it is possible a priori to assign an absolute magnitude to a certain type of star this gives an indirect method of finding distances.

1.1.4 Spectroscopic or Photometric Parallaxes

This method proceeds in two stages:

(a) From the direct distance determination for the Hyades cluster, we calibrate the absolute magnitudes M_v of the stars in the cluster.

(b) We make the hypothesis that the main-sequence stars of other clusters have the same properties as those of the Hyades, so that a spectral type corresponds to a luminosity. This is the technique of *main-sequence fitting*. Thus the spectral classification of the stars in these clusters gives their luminosities and hence their distances.

This method has been widely used for the determination of cluster distances in the Galaxy. It is a typical secondary method. It is affected by various types of errors, some photometric, some connected with the fact that the main sequence has non-zero width on the Hertzsprung–Russell diagram (cf. Sect. 1.4). The absolute magnitudes M_v found in this way are accurate to ± 0.2 mag, implying a distance error $\Delta d/d$ which reaches or sometimes exceeds 20 %.

A good account of methods of finding absolute magnitudes of stars is given by Heck (1978).

1.1.5 The Importance of Stellar Distance Determinations

The determination of distances (in the Euclidean sense) in the whole Universe is fundamentally based on the distance of stars from the Sun. For this

reason the paucity of direct distance determinations (via trigonometric parallaxes) is a major stumbling block. The neighbourhood of the Sun where distances are well calibrated ($\Delta d/d \leq 20\,\%$) is at present a volume of 10 pc radius. The *Hipparcos* satellite will extend this volume to 100 pc with the same accuracy, at least for stars brighter than magnitude $B = 9$, with lower accuracy for stars up to $B = 13$.

For the first time, because of *Hipparcos*, we can expect that some intrinsically luminous stars (supergiants, Cepheids) will have distances which are precise, or fairly precise (the nearest Cepheid is at 100 pc), for these are the stars that we see most easily in other galaxies. Distance estimates for galaxies are based on the hypothesis that their most luminous stars have the same intrinsic luminosities as similar stars in our Galaxy. It is therefore vital that the first steps in this method of stepwise calibration are well calibrated. Similarly the distance to the standard candle which the Hyades constitutes for the method of main-sequence fitting should be known to $1\,\%$ from *Hipparcos* (rather than $10\,\%$ as today).

The *Hipparcos* satellite gives parallaxes, proper motions, and positions all for the first time subject to homogeneous errors. Of course the observed sample will remain small (120 000 stars, while the Galaxy contains 100 billion), and even in the solar neighbourhood the faintest stars will not be accessible. The importance of this first astrometric satellite is nevertheless very considerable.

1.2 Stellar Spectra

When dispersed by a spectrograph, the visible light from a star shows a *continuum* spectrum generally cut by dark *lines* (absorption lines), but occasionally with lines brighter than the continuum (emission lines). The same is true for radiation in the X-ray, UV, IR, submillimetre, millimetre or longer-wavelength radio.

1.2.1 Continuum Spectrum of a Star

Comparison with a Black Body. The continuum spectrum of a star may be compared with that of a black body of intensity $B_\lambda(T)$, in three different ways.

— If the observed spectrum is relatively calibrated, that is, we may compare the received energy $f(\lambda)$ at two wavelengths λ_1 and λ_2, we can define the *colour temperature* of the star between these two wavelengths as that of the black body having the same gradient $\Phi(\lambda_1, \lambda_2)$ as $f(\lambda)$:

$$\Phi(\lambda_1, \lambda_2) = -\frac{\Delta \ln}{\Delta(1/\lambda)} \left\{ \frac{c_1}{\lambda^5} \left[\exp\left(\frac{c_2}{\lambda T}\right) - 1 \right]^{-1} \right\}$$

$$= -\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) \ln \left(\frac{\lambda_1^5}{\lambda_2^5} \frac{F(\lambda_1)}{F(\lambda_2)}\right), \tag{1.17}$$

where $f(\lambda)$ is proportional to the flux $F(\lambda)$ of the black body,

$$F(\lambda) = \pi B_\lambda(T) = \frac{c_1}{\lambda^5} \left[\exp\left(\frac{c_2}{\lambda T}\right) - 1 \right]^{-1}.$$

The constants c_1 and c_2 are $c_1 = 3.741\,85 \times 10^{-5}$ (erg cm^{-2} s^{-1}) if λ is in cm, and $c_2 = 1.438\,83$ cm K.

— If the observed spectrum is absolutely calibrated, and if we know the star's distance, then

$$f(\lambda) = \frac{L(\lambda)}{4\pi d^2} = \frac{4\pi R^2}{4\pi d^2} F(\lambda), \tag{1.18}$$

where $F(\lambda)$ is the flux emitted at the star's surface at wavelength λ, and R is the star's radius. We define the *radiation* or *brightness* temperature T_b as that of a black body of the same brightness at this wavelength:

$$F(\lambda) = \pi B_\lambda(T_b). \tag{1.19}$$

— Finally, the intrinsic luminosity of the star $L = \int_0^\infty L(\lambda)d\lambda$ is equal to that of a black body at the *effective temperature* of the star:

$$L = 4\pi R^2 \sigma T_{\text{eff}}^4, \tag{1.20}$$

where σ is Stefan's constant ($\sigma = 5.669\,56 \times 10^{-5}$ erg cm^{-2} K^{-4} s^{-1}). For a given stellar spectrum the effective temperature is unique if the radius is specified (cf. Sect. 1.3). The colour temperature varies from spectral region to another. The brightness temperature is particularly useful for the Sun (Sect. 2.2); for other stars it is affected by distance errors. Calibrated stellar spectra are found not by comparison with the Sun, which is much too bright, but with Vega (α Lyr), which is itself calibrated by means of a laboratory radiation source.

Colour Indices. Let us assume we are studying a continuous spectrum using a photometer rather than a spectrograph. We may use a photometric system with wide bands (U, B, V, or the Geneva system), with intermediate bands (u, v, b, y of Strömgren), or even with narrow bands (see Golay 1974 for a description of photometric systems). Each filter defines the radiation received in a spectral interval $\Delta\lambda$, and is characterised by its effective wavelength

$$\lambda_{\text{eff}} = \frac{\int_{\Delta\lambda} S(\lambda) T(\lambda) \lambda \, d\lambda}{\int_{\Delta\lambda} S(\lambda) T(\lambda) \, d\lambda}, \tag{1.21}$$

where $S(\lambda)$ is the response function of the photometer and $T(\lambda)$ the atmospheric transmission.

We define as the *colour index* between two wavelengths denoted 1 and 2 the quantity

$$CI = m(\lambda_{\text{eff},1}) - m(\lambda_{\text{eff},2}) + \text{const}, \tag{1.22}$$

the constant being defined so that for a well-defined spectral type, CI is zero for any choice of the λ_{eff} (the convention is type A0). The notion of a colour index is formally equivalent to that of the colour temperature defined in (1.17), but it has the great advantage of being easily measurable.

In the UBV system for example one uses the index $U - B$ to characterise the strength of the near ultraviolet in the continuum spectrum of the star, while $B - V$ measures the slope of the Paschen continuum and, with very good accuracy, the star's effective temperature.

Calibration of the colour indices is an essential step in getting to the characteristic parameters of a star, which are L and T_{eff} (we should also add the chemical composition, as we shall see later). It relies on model-atmosphere calculations which predict the radiation spectra of atmospheres with specified T_{eff} and surface gravity (Hayes et al. 1985; Böhm-Vitense 1981).

We note finally that the luminosity L in solar units is given by M_{bol} (1.16), which itself relies on a knowledge of the bolometric corrections for different types of star (1.15). This in turn requires a knowledge of the stellar energy distribution throughout the electromagnetic spectrum and a determination of M_v.

1.2.2 Line Spectra: Spectral Classification

Stellar spectra differ not merely in the colour of their visible continua, but also in their absorption line spectra, which vary widely from star to star. It is natural to classify stars on the basis of uniformly measured spectra (same spectrograph dispersion). This ordering of spectral types was carried out mainly by the Harvard astronomers of the early 20th century. Thus stars are classified by their *spectral types*, which are always denoted by the sequence

O B A F G K M

to which can be added W (at left) and types C and S at right; the former is level with O, and the latter two level with M. Originally the order was supposed to be alphabetical, but soon had to be modified, eliminating letters

to arrive finally at the present sequence O, B, A, ... The types are divided into ten; for example type B contains the subtypes B0, B1, ... , B9.

This sequence is based on the change of the absorption-line spectrum as one moves from blue stars (O, B) to red (K, M). Type W is different since its stars have *emission* lines (Wolf–Rayet stars).

The main characteristics of the various spectral types and the criteria used in the Harvard classification are as follows.

- Type O. The hottest stars. Lines of[1] He II, O II, C III, N III, Si IV. Example: ζ Pup.
- Type B. Spectra are dominated by lines of He I. Lines of He II disappear. Lines of hydrogen strengthen and those of singly ionised elements (C II, Si II, Mg II) appear. Example: α Leo (Regulus)
- Type A. Hydrogen lines dominate and are a maximum at subtype A0. Lines of Mg II and Si II are strong, those of Ca II become progressively stronger. Example: α Lyr (Vega), α CMa (Sirius).
- Type F. Metal lines grow in strength (neutral atoms and ions such as Ca II). Hydrogen lines become weaker. Example: α CMi (Procyon).
- Type G. Stars of solar type, in the strict sense. The resonance line of Ca I is very strong, as are certain lines of Fe I. Ca II is always strong. The G band (spectrum of the CH radical) appears. Example: α Cen A (Rigil Kent).
- Type K. The resonance lines of Ca II reach their maximum. Lines of H I become weak. Other molecular bands appear (CN, ...). Example: α Boo (Arcturus).
- Type M. The spectrum is dominated by molecular bands, in particular TiO. Lines of neutral elements (Ca I) remain very strong.

Besides this standard sequence there are parallel to type M

- Carbon stars (type C, sometimes subdivided into types R and N) where the molecular bands C_2, CN, CO, CH are very strong, while TiO is absent. Example: o Ceti (Mira).
- Stars of type S, where the dominant molecular bands are ZrO, LaO, YO. Example: R Gem.

Representative spectra of this classification are given in Abt et al. (1969) or Jaschek and Jaschek (1987). The spectral types of the stars in Table 1.1 are given in column 5 of that table.

We note an unusual designation in this table: the white–dwarf companion of Sirius, Sirius B, has type DA; D denotes white dwarfs; DO, DB, DA, DF are different subtypes (Sect. 8.3). Other designations are encountered for groups of stars which, like white dwarfs, represent a particular stage of stellar evolution. For example the hot Wolf–Rayet stars (type W) have

---

[1] Note: the spectroscopic designation of elements is He I for neutral helium, He II for singly ionised helium, etc.

subtypes WC or WN according as the lines of carbon (C III and C IV) or nitrogen (N III, N IV, N V) dominate. We shall return later to departures from the spectral classification.

The spectral classification always uses the *visible* spectrum. There are *standard* stars for each type and subtype. Other stars are classified by interpolation by eye; the classification is thus entirely empirical.

However, it soon became clear that the spectral classification by lines followed that by colours, and that both arrange the stars by *temperature*. O stars are the bluest and hottest, and their emission regions contain elements as ionised as He II. M stars are the reddest and coolest, so lines from neutral elements are found in their spectra, as well as lines from molecules which are not dissociated at the low temperatures of the emission region.

It also became clear, but not so easily, that the strength of the lines does not depend solely on the temperature of the star. The *luminosity* of the star influences its spectrum, for example by narrowing the lines. This was noticed by Hertzsprung around 1905 when comparing stars of the same spectral type, for which he knew approximate distances and hence L, but whose lines had different widths. The most luminous stars have the narrowest lines for a given spectral type. This suggests that their *surface densities* are lower than those of less luminous stars, since the lines are broadened by collisions of the radiating atoms with electrons or hydrogen atoms, and the effect is proportional to the density. A more extended atmosphere, or a larger radius R, will mean a lower density. This explains the terminology dwarfs, giants, supergiants, etc. to denote more and more luminous stars, and thus by (1.20), for the same effective temperature, larger and larger radii.

The spectral classification thus becomes a two-parameter classification, of spectral type and *luminosity class*, the latter giving the size (or intrinsic luminosity) of the star. In the MK (Morgan-Keenan) system, luminosity classes are denoted by Roman numerals I to V, the supergiants (I) being further classified into Ia0, Ia, Ib, by decreasing luminosity. Giants are denoted II or III, by decreasing luminosity; class IV are the subgiants and class V the dwarfs. For example Vega (α Lyr) is classified A0V in the MK system.

The two-dimensional spectral classification is simultaneously a basic achievement of stellar astrophysics and a reference grid against which to measure departures from normality. In fact many stars are not easily accommodated in the MK classification; they are *peculiar*. Examples are Be stars — B stars with emission lines, a prototype being γ Cas; stars of types A and B with strong lines of rare elements such as gadolinium, mercury, or the rare earths, or abnormally strong lines of certain elements (Mn, Si, Fe): these are called Ap and Bp stars; K and M-type stars which flare or have emission lines, denoted Ke or Me (e.g. α Cen C and UV Cet, Table 1.1); barium stars, and so on. Departures from the spectral classification are always the sign of important physical phenomena occurring either in the atmo-

sphere or the subphotospheric layers of the star. Before we discuss these phenomena, we must describe the basis of the quantitative study of stellar spectra.

1.2.3 Spectral-Line Analysis

The *profile* of a line is its intensity distribution $I(\lambda, \theta)$ or the flux $F(\lambda)$ as a function of wavelength; θ denotes the angle between the normal to the emitting layers and the line of sight. In the case of the Sun, where we can resolve the surface in angle, we can measure $I(\lambda, \theta)$. For stars, where the disc cannot be resolved, the energy $f(\lambda)$ received at the Earth's surface is proportional (1.18) to the flux from the star's surface $F(\lambda)$. One can easily show (Mihalas 1978, Chap. 1) that $F(\lambda)$ is equal to the mean intensity emitted by the disc (the mean is over angle). Here we consider stellar spectral lines and their profiles expressed through the flux.

The amount of information contained in a spectral line depends on the spectral resolution $\lambda/\Delta\lambda$ of the spectrograph and the signal-to-noise ratio S/N (see Léna 1988, Chap. 7). Today a good spectrum has resolution $\geq 50\,000$ and $S/N \geq 100$.

Stellar spectral lines are formed in an inhomogeneous region of gas, so that their profiles cannot in general be taken to have a simple form (Gaussian or Lorentzian, corresponding to a single density and temperature, as holds for lines observed in the laboratory). The emergent flux from the star at each wavelength results from the balance of photon absorptions and emissions along the ray as it passes through the star to the surface. The *radiative-transfer equation* describing this out of equilibrium process will be discussed in Sect. 2.2; here it is sufficient to note that the observed profile results from two properties of the medium: its emissivity and its ability to absorb photons, both functions of the density and temperature.

Equivalent Width.   This is the integral quantity

$$W = \int_{\text{line}} \frac{F_c(\lambda) - F_l(\lambda)}{F_c(\lambda)} \mathrm{d}\lambda = \int_{\text{line}} (1 - r(\lambda)) \mathrm{d}\lambda \,, \tag{1.23}$$

where F_c is the continuum flux and F_l the line flux (Fig. 1.4). It represents the emitted (or absorbed) energy in the line, compared with the continuum. It has dimensions of a wavelength (Å in general), hence the name. The ratio $r(\lambda) = F_l(\lambda)/F_c(\lambda)$ represents the flux removed from the continuum, and $1 - r(\lambda)$ is the *depth* or *depression* of the line (if it is in absorption).

The notion of equivalent width is important in the *abundance* analysis of the absorbing elements. Methods for studying abundances will be developed for the case of the Sun (Sect. 2.3), but they apply to other stars and the interstellar medium, etc.

Fig. 1.4. Spectral line in absorption

$r = \dfrac{F_r}{F_c}$

Depression $= 1 - r(\lambda)$

Full width at half maximum (FWHM)

λ

(Heliocentric scale)

$\lambda_{\text{lab}} \quad V_R \dfrac{\lambda_0}{c}$

λ_{lab}
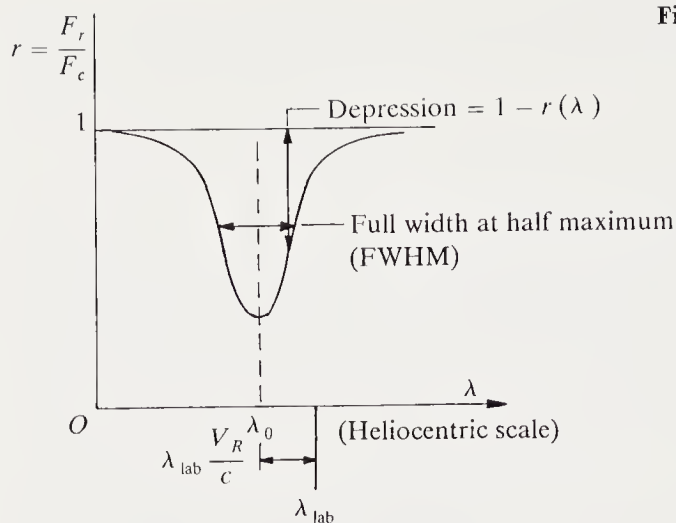
Line Profiles: Symmetrical Lines. We consider symmetrical absorption lines.

— The position of minimum observed flux $F(\lambda_0)$ may be compared to the wavelength of the same line observed in the laboratory (it is assumed that the stellar wavelength scale has been corrected for the Earth's velocity, cf. Sect. 1.1). The deviation $\lambda_0 - \lambda_{\text{lab}}$ measures the *heliocentric radial velocity* V_R of the star.

$$\frac{\lambda_0 - \lambda_{\text{lab}}}{\lambda_{\text{lab}}} = \frac{V_R}{c}, \tag{1.24}$$

where c is the speed of light.

— V_R is taken as positive if the star is moving away from the Sun, negative if it approaches. The wavelength scale must be corrected by V_R if we wish to consider a reference frame fixed in the star. If moreover λ_0 varies periodically in time, the star may belong to a *binary or multiple* system, two or more stars orbiting their common centre of gravity, or be a *pulsating* star.

— The observed profile, which we denote $O(\lambda)$, is the convolution of the true stellar profile $F(\lambda)$ with all modifying geometrical effects $B(\lambda)$ and the spectrograph's instrumental profile $S(\lambda)$. We thus have

$$O(\lambda) = F(\lambda) * B(\lambda) * S(\lambda). \tag{1.25}$$

The instrumental profile must be determined for each spectrograph used. The profile $B(\lambda)$ is a *broadening function* which may be due to

— rotation of the star's surface
— the fact that the line is formed in a turbulent medium whose cells are large compared with the mean free path of the photons involved in the line formation.
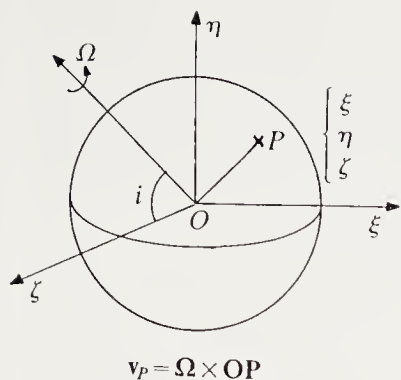
$$v_P = \Omega \times \mathbf{OP}$$

We obtain $B(\lambda)$ as follows:

(a) *Rotational broadening.* Figure 1.5 shows the star, assumed spherical and rigidly rotating with angular velocity Ω. The axis of rotation makes an angle i with the line of sight. We can divide the stellar surface into isovelocity regions, i.e. zones contributing to the same radial velocity $v_\zeta = -\xi \Omega \sin i$ along the line of sight.

The maximum Doppler shift of a line of wavelength λ is

$$b = \lambda \Omega R \sin i / c = \lambda V \sin i / c, \tag{1.26}$$

where $V \sin i$ is the projected equatorial velocity of the star.

The observed profile $O(\lambda)$ is obtained from the emitted profile $F(\xi, \zeta, \lambda)$ at position (ξ, ζ) on the disc. For an infinitely narrow instrumental profile we have

$$O(\lambda) = \int_{-1}^{+1} F(\lambda - bx)\, B(x)\, \mathrm{d}x\,, \tag{1.27}$$

where $x = \xi/R$, $F(\xi, \zeta, \lambda) = \text{const}(1+\beta)\cos\theta$, where β is the limb-darkening coefficient in the continuum, assumed constant for all λ in the profile, and $\cos\theta = (1 - \xi^2 - \zeta^2)^{1/2}$. The broadening profile $B(\lambda)$ (see Unsöld 1968) is given by

$$B(x) = \left[\frac{2}{\pi}(1 - x^2)^{1/2} + \frac{\beta}{2}(1 - x^2)\right] \Big/ (1 + 2/3\beta), \tag{1.28}$$

with $B(\lambda) = 0$ outside $[-1, +1]$, and $\int_{-\infty}^{+\infty} B(x)\mathrm{d}x = 1$.

Equation (1.27) is not strictly a convolution. To invert it and get $B(x)$ we may rewrite it as

$$\begin{aligned}
O(\lambda) &= \int_{-\infty}^{+\infty} F(\lambda - bx)B(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} F(\lambda - t)B\left(\frac{t}{b}\right)\frac{\mathrm{d}t}{b} \\
&= \int_{-\infty}^{+\infty} F(\lambda - t)A(t)\,\mathrm{d}t \quad \text{with} \quad A(t) = \frac{1}{b}B\left(\frac{t}{b}\right),
\end{aligned} \tag{1.29}$$

so that $A(t)$ gives the broadening function for rotation.

(b) *Broadening by macroturbulence.* Because the size of the turbulent elements exceeds the mean free path of the photons, the line is formed entirely within one cell. The line profile is broadened but its equivalent width is unaltered. The broadening function $B(\lambda)$ is generally assumed to be Gaussian; it follows the distribution function $N(U)dU$ of random velocities U of turbulent elements

$$B(\lambda)d\lambda \sim N(U)dU = \text{const} \times \exp(-U^2/V_M^2)\,dU$$
$$= \text{const} \times \exp(-\Delta\lambda^2/\Delta\lambda_D^2)\,d\lambda/\Delta\lambda_D\,, \qquad (1.30)$$

where $\Delta\lambda/\lambda = U/c, \Delta\lambda_D/\lambda = V_M/c$. V_M is called the *macroturbulent velocity*. V_M^2 is a quadratic mean velocity. There is no requirement that the macroturbulent velocity field should be isotropic.

The general relation (1.25) can be considered in Fourier space. $B(\lambda)$ is obtained from its Fourier transform $\widehat{B}$ if there exists a theory of the intrinsic profile $F(\lambda)$. Stellar-atmosphere theory gives $F(\lambda)$. We thus get

$$\widehat{B} = \widehat{O}/(\widehat{F}\cdot\widehat{S})\,. \qquad (1.31)$$

This method has been successfully applied to find the rotational broadening when the macroturbulent broadening is negligible (dwarf stars). Moreover it is the only method giving a good estimate of the macroturbulent velocity V_M. The book by Gray (1976) gives a good explanation of this. Using extremely high quality profiles it has been possible to measure velocities $V\sin i$ of the order of 2 km s^{-1} to an accuracy of ± 0.5 km s^{-1}.

Before Fourier analysis of the profiles became current, for velocities $V\sin i$ large enough compared to the width of the intrinsic profile $F(\lambda)$ and the instrumental width, it was usual to measure just the *full width at half maximum* (FWHM); this is the width corresponding to a depth $(1 - r(\lambda_0))/2$ (see Fig. 1.4). This gives

$$V\sin i = \tfrac{1}{2}\text{FWHM}\,.$$

The accuracy then obtainable did not allow determination of rotation velocities below 10 km s^{-1}. This method is still used for very faint stars, observed at low resolution.

— The wings of the line are barometers of the medium forming the lines, as suggested above (Sect. 1.2.2) (except for the hydrogen Balmer lines, which are barometers up to spectral type A0, and thermometers after this). If we compare the same line in a dwarf and a supergiant of the same spectral type, the wings are stronger in the dwarf than in the supergiant because of the sensitivity of the mass absorption coefficient $\kappa_l(\lambda)$ to the density ρ. The density is several orders of magnitude higher in the dwarf. $\kappa_l(\lambda)$ can be expressed as

$$\kappa_l(\lambda) = \kappa_0 \, \varphi(\lambda) \,, \tag{1.32}$$

where κ_0 is the absorption coefficient at line centre and $\varphi(\lambda)$ is the *intrinsic line profile* (cf. Sect. 2.2). The general form of $\varphi(\lambda)$ is the convolution of a Gaussian and a Lorentzian, called the *Voigt profile* and written $H(a,v)$:

$$\varphi(\lambda) \equiv H(a,v) = \frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{\exp(-x^2)}{a^2 + (v-x)^2} \, \mathrm{d}x \,. \tag{1.33}$$

A general point of the line is given by

$$v = \frac{\lambda - \lambda_0}{\Delta\lambda_D} = \frac{\nu - \nu_0}{\Delta\nu_D} \,,$$

where $\Delta\nu_D$ and $\Delta\lambda_D$ denote the Doppler width. The dimensionless *damping constant* a is defined by

$$a = \frac{\gamma}{4\pi\Delta\nu_D} = \frac{\gamma_{\mathrm{rad}} + \gamma_{\mathrm{col}}}{4\pi\Delta\nu_D} \,, \tag{1.34}$$

where γ_{rad} is the width (FWHM) of the intrinsic profile caused by the finite lifetime of the levels, and γ_{col} is the width due to collisions of the radiating atom with electrons, ions, or neutral atoms. γ_{col} always has the form

$$\gamma_{\mathrm{col}} = \gamma_1 N_e + \gamma_2 N_H + \gamma_3 N_p + \ldots \,, \tag{1.35}$$

where N_e, N_H, N_p denote the number densities of electrons, neutral hydrogen atoms, and protons. a is thus a function of density.

Study of absorption line wings is a standard method of finding the density of the atmospheric layers in stars (see e.g. Mihalas 1978).

Line Profiles: The Asymmetric Case. In some stars the absorption lines are not symmetrical about λ_0. Observed asymmetry always signifies motion of the medium forming the lines. The asymmetry can take different forms.

— The line bisector does not coincide with λ_0. The bisector is the curve formed by the midpoints of sections of the profile parallel to the λ axis, with the line represented in reduced flux units. In the Sun and some dwarfs of similar spectral type, the bisector has a C-shape (Fig. 1.6), concave towards long wavelengths.

The interpretation of this asymmetry and the associated wavelength shift of the bisector involves the granular structure of a stellar surface through the emergence of blobs of material formed in the convection zone. Figure 1.6a shows a schematic 2D model of the granulation (see Sect. 2.4), consisting of bright ascending convective elements and a dark intergranular structure forming the descending flow. The profiles formed in the various regions are shown as if in very high spatial resolution (Fig. 1.6b). The profile resulting from the integrated disc has a C-shaped bisector and has the line centre shifted to the blue (Fig. 1.6c).
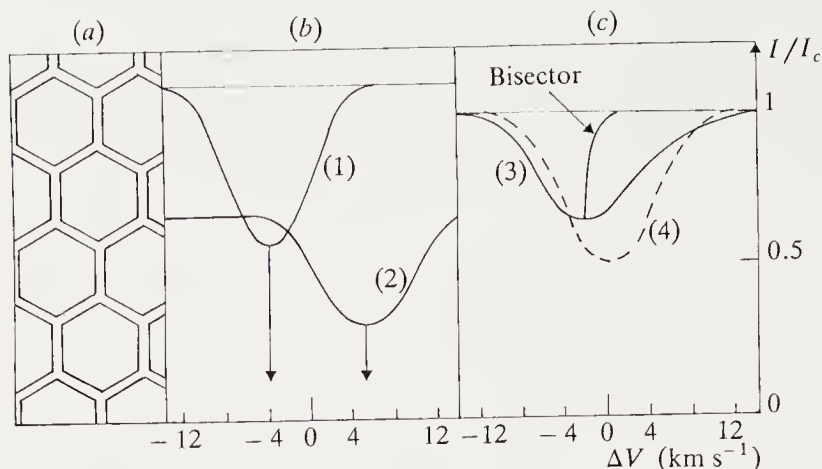
Fig. 1.6. Asymmetric line with C-shaped bisector. Explanatory scheme for the origin of the asymmetry and the wavelength shift due to convective motion. (1) Profile formed in ascending elements; (2) profile formed in descending elements; (3) mean observed profile (averaged over many granules); (4) profile in the absence of the velocity field. ΔV is the shift measured in km s^{-1} in a frame fixed at the star's centre. (From D. Dravins, *Ann. Rev. Astron. Astrophys.* **20**, 61 (1982). Reproduced with the kind permission of *Annual Reviews of Astronomy and Astrophysics*; (© 1982 Annual Reviews Inc.)

— One of the line wings is stronger than the other (for absorption lines, deeper). Assume that the blue wing is stronger (Fig. 1.7); if there are no granules, $\kappa_l(\lambda - \lambda_0)$ is larger for $\lambda - \lambda_0 < 0$ than for $\lambda - \lambda_0 > 0$, for the same $|\lambda - \lambda_0|$. Here we must distinguish the reference frame of the observer, who measures $\lambda - \lambda_0$, from the reference frame of the flow of emitting atoms. The observed asymmetry reflects the bulk motion (of velocity $V(z)$ with respect to the observer) of these atoms. Let ν be the frequency seen by the observer; this corresponds to a local value ν_L in the moving frame,

$$\nu_L = \nu(1 - V(z) \cdot n/c) \simeq \nu - \nu_0 V(z)\mu/c, \tag{1.36}$$

where z is the coordinate normal to the layers, and $\mu = \cos\theta$. The mass absorption coefficient $\kappa_l(\nu)$ defined in (1.32) is now

$$\kappa_l(\nu) = \kappa_0 \varphi(\nu_L - \nu_0), \tag{1.37}$$
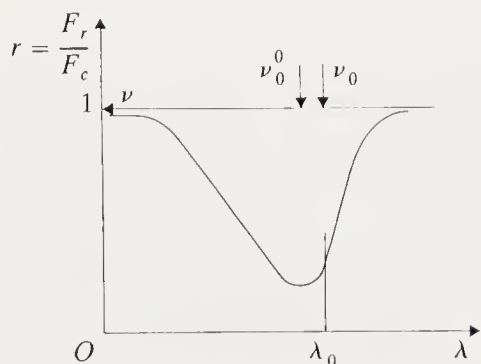
Fig. 1.7. Asymmetric line caused by a monotonic velocity field
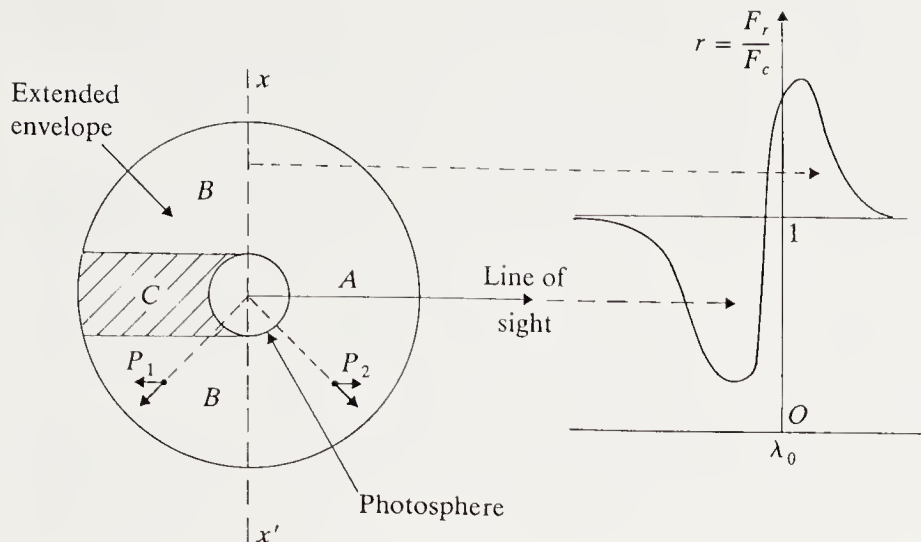
Fig. 1.8. P Cygni line profile and schematic view of the expanding envelope producing it

where $\varphi(\nu_L - \nu_0)$ is the intrinsic line profile in the moving frame. We note that the frequency ν_L seen by the moving atoms varies with z and the direction $\boldsymbol{n}$ of the photons. This means that the central frequency in the observer's frame is no longer ν_0, but is given by the maximum of $\varphi(\nu_L - \nu_0)$ in (1.37), at $\nu_L = \nu_0$. We call this value ν_0^0. We have

$$\nu_0^0 = \nu_0(1 + V(z)\mu/c) = \nu_0(1 + V_r(z)/c) = \nu_0^0(z, \mu). \tag{1.38}$$

If $V(z) = V = \text{const}$, the sign of $\nu_0^0 - \nu_0$ shows whether the medium approaches or recedes from the observer. In the case of Fig. 1.7, $\nu_0^0 - \nu_0$ is positive and the motion is towards the observer. If $V(z)$ is not constant, the gradient $dV_r(z)/dz$ also affects the sign of $\nu_0^0(z, \mu) - \nu_0$. The interpretation of an asymmetric line as in Fig. 1.7 becomes complex.

— P Cygni profiles. This type of profile (Fig. 1.8) is characteristic of very luminous stars (it is named after the hypergiant P Cygni, classified B1Ia$^+$), and has both an absorption and an emission component. The former is shifted to shorter wavelengths and the latter is centred near wavelength λ_0 (in the star's frame). The line is formed in an extended envelope which the observer sees partly in projection against the layers forming the continuum (region A) and partly outside this (regions B); a third region (C) of the envelope is occulted. Region A forms a blueshifted absorption line if the envelope is expanding. Regions B, accessible to lines of sight not crossing the core, are generally very optically thin and produce an emission line which would be centred at λ_0 if the occulted zone was negligible, but which is redshifted in many cases. Indeed, if region C is negligible we can always find in regions B volumes P_1, P_2 symmetrical about the line xx' (and thus with respect to the plane normal to the line of sight passing through xx') such that their radial velocities cancel for the observer. If region C (and hence the similar region in front of it) is not negligible, a part of the emission

line centred at $V = 0$ is missing, resulting in a redshift of the emission component.

This qualitative interpretation, due to Beals (1949), is fully confirmed by detailed calculations (see Mihalas 1978, Chap. XIV).

P Cygni profiles are usually seen in resonance lines of abundant ions (Mg II, C IV, N V, O VI, etc.) but also in subordinate lines such as Hα, Hβ, He I λ 10830, etc. All these lines are strong enough in hot stars that their centres reflect conditions in the outer parts of the atmosphere where the wind has a significant velocity (several hundred km s^{-1}). In cool stars the resonance lines of e.g. Na I have P Cygni profiles. They are a very important tool in the study of *stellar winds* (see Chap. 4).

A general point about asymmetric lines is that *it is vital to establish carefully the wavelength scale of the star's centre; otherwise the velocities estimated from the profiles will be meaningless.*

Absorption Lines, Emission Lines. The spectral classification is based on an ordering of absorption-line strengths, following a temperature scale. However, we have seen that lines can appear in emission, i.e. be stronger than the neighbouring continuum. We should ask what is unusual about such stars, remembering that for e.g. planetary nebulae, all the lines are in emission.

We have just seen that the standard P Cygni profile has an emission component, which we attributed to the formation of the line in an extended envelope. By constrast, all the lines in the visible spectrum used for the spectral classification are formed in the layers producing the visible continuum also. This layer has a depth which is small compared to the star's radius R, and is called the *photosphere* (see also Sect. 2.2). Its temperature decreases outwards. Figure 1.9 shows how an absorption line forms in a plane-parallel photosphere. Let $\kappa_l(\lambda_0)$ be the mass absorption coefficient at line centre, and κ_c the absorption coefficient in the continuum. We assume that the observer sees into the atmosphere to a depth such that $\int^z \kappa\rho\,\mathrm{d}z = 1$. To a first approximation
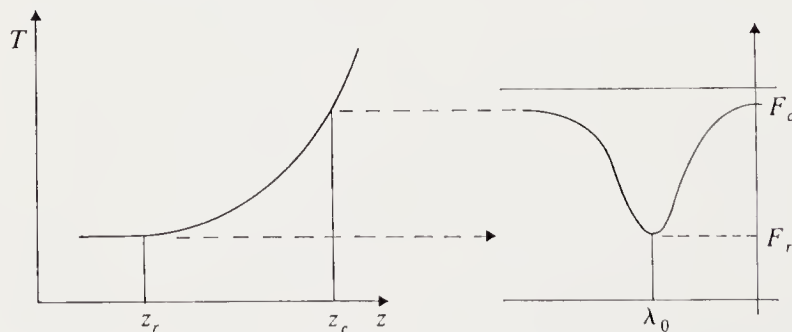
Fig. 1.9. Formation of an absorption line. *Left*, the temperature distribution in the atmosphere; *right*, the emergent line profile

— for the continuum: $\int^{z_c} \kappa_c \rho_c \mathrm{d}z = 1$
— at line centre: $\int^{z_l} (\kappa_l(\lambda_0) + \kappa_c) \rho \mathrm{d}z = 1$
— as $\kappa_l(\lambda_0) \gg \kappa_c$, we see that $z_l \ll z_c$, and thus $T(z_l) < T(z_c)$.

In general this means $F_l < F_c$; the line is darker than the continuum on which it appears. If emission lines appear it means that one or more of the above assumptions does not hold:

— The line is formed in another layer of the atmosphere (chromosphere, corona) where the temperature does not decrease outwards.
— The line is formed in a very extended layer (envelope) whose geometry invalidates the definition of the formation depth assumed above, whatever the temperature. Examples: P Cygni profiles, but also emission lines in Be and Wolf–Rayet stars, etc.

The foregoing will have shown that the analysis of line profiles gives a wealth of information about the thermodynamic and dynamical state of the line-forming region.

The *thermodynamic* state is specified by the temperature T and density ρ, which, as suggested above, vary with depth z. Study of this dependence leads to the construction of *model atmospheres*; the principles will be discussed in Chap. 2. But we have also seen above that different ions of the same element appear as we consider the range of spectral types and luminosity classes. For a given spectral type the lines of Fe II for example strengthen as we go from dwarfs to supergiants. Following the dwarf sequence, we go by decreasing temperatures through lines of He II (O stars) to lines of Ca I or K I (G, K stars), neutral elements which are very easily excited. The interpretation of the density variation at given spectral type (given T) was provided by Saha in 1920, as a law of *ionisation equilibrium*. The Saha law is a form of the law of mass action, which holds for thermodynamic equilibrium, applied to two successive ions of an element, called N^+ and N^0 for simplicity:

$$\frac{N^+ N_e}{N^0} = 2 \frac{U^+}{U^0} \left(\frac{2\pi m_e kT}{h^2} \right)^{3/2} \exp(-\chi_{\mathrm{ion}}/kT), \tag{1.39}$$

where χ_{ion} is the ionisation energy, N_e the electron number density, and U^+ and U^0 the partition functions of the two ions; the factor 2 comes from the statistical weight of the electron, whose mass is m_e; h and k are the Planck and Boltzmann constants. This law gives a simple interpretation of the changes in ionisation balance observed in the spectral classification. A more refined analysis of stellar spectra requires other state variables besides T and ρ: we need the populations of all the atomic energy levels, and the description of the gas becomes one departing from thermodynamic equilibrium (see Thomas 1965; Oxenius 1986).

The *dynamical* state of the medium is given by the velocity vector at each point; this may have a systematic component (for example, a wind) and a random component (for example, turbulence on various scales). Diagnosing the dynamical state of the stellar gas is a difficult problem (see Chap. 5); however, it is the empirical basis for the study of many phenomena: heating of exterior layers, generation of magnetic fields, diffusion, etc.

1.3 Stellar Radii

The notions of dwarf, giant, and supergiant stars arose in the discussion of the spectral classification. Measurement of stellar radii, when it is possible, confirms the differences in the sizes of stars.

We return to relations (1.18) and (1.20). The received flux f (outside the atmosphere) is related to the flux F emitted at the star's surface by

$$f = \frac{R^2}{d^2} F = \frac{\theta^2}{4} F,$$

(1.40)

where θ is the star's angular diameter ($\theta = 2R/d$). To measure a stellar radius usually means measuring θ. This is always a very small angle ($< 0.1''$), and stellar discs are not at present resolved in the visible, in the sense that we have no images of the structure of their discs.

1.3.1 Direct Measurement of the Radius

This uses interferometry: either phase, amplitude or speckle. These methods are described by Léna (1988, Chap. 6). By now the angular diameter of about 100 stars has been measured directly. The largest star in angular size (apart from the Sun) is Betelgeuse (α Ori, M2 Iab), which has $\theta = 0.067'' \pm 0.005''$ at $\lambda = 4880$ Å. Table 1.3 gives the angular diameters θ, parallaxes π, distances d, and deduced radii R (in solar units) of a set of stars of various spectral types. It is clear from the table that giants and supergiants have radii which are respectively of the order of 10 and 100 (or even 1000) times that of the Sun.

The angular diameter of a star depends on λ. Generally "radius" will mean "visible radius" ($\lambda = 5500$ Å), corresponding to the atmospheric layer emitting the visible continuum. Direct measurements of stellar radii will benefit from the improved accuracy of distances supplied by the *Hipparcos* satellite.

1.3.2 Radii of Eclipsing Binary Stars

Variable stars (discussed in Chap. 6) include some binaries whose orbital plane is close to the line of sight. A consequence of this geometrical situation

Table 1.3. Angular diameter, distance, and radius of bright stars (adapted from Allen 1976)

Star	Name	Spectral type	Angular diameter (″)	Parallax (″)	Distance (pc)	Radius (in $R_\odot$)
α Ori	Betelgeuse	M2Iab	0.067	0.005	180	1 297
α Tau	Aldebaran	K5III	0.024	0.048	21	54
α Boo	Arcturus	K2III p	0.022	0.090	11	26
α Aur	Capella	G5III/ G0III	0.0047	0.075	14	7.1
α CMi A	Procyon	F5IV-V	0.0055	0.287	3.5	2.1
α PsA	Fomalhaut	A3V	0.00210	0.149	6.7	1.5
α Cyg	Deneb	A2Ia	0.0027	− 0.013 (ᵃ)	500	145
α CMa A	Sirius	A1V	0.00589	0.377	2.65	1.7
α Lyr	Vega	A0V	0.00324	0.126	7.9	2.76
β Ori	Rigel	B8Ia	0.00255	− 0.003 (ᵃ)	270	74
α Leo	Regulus	B7V	0.00137	0.041	24.4	3.6
α Vir	Spica	B1IV	0.00087	0.021	65	7.8

(ᵃ) A negative parallax implies one at the limit of current accuracy.

is that each of the stars passes in front of the other once per orbital period, an occultation phase following a transit phase. Thus the light received from the system varies periodically, with two successive minima: primary minimum corresponds to the eclipse of the hotter star by the cooler, and secondary minimum to the reverse situation. Algol (β Per) is an example of an eclipsing binary.

Analysis of the light curve of an eclipsing binary can give the radii of the two components, in the following way.

(1) We normalise the measured intensities to the brightness between minima, which we can take as constant. Let the normalised intensity be l.

(2) We measure the phase angle $\theta = \frac{2\pi}{P}(t - t_0)$, where P is the heliocentric orbital period and t_0 the time of primary minimum.

(3) We assume the orbit to be circular. Let i be the inclination of the orbital plane to the plane of the sky (tangent to the celestial sphere), and d the distance between the two stars.

(4) The two stars of radii r_1 for the larger and $r_2 = kr_1$ for the smaller have discs deprived of limb darkening. The following relations are easily proved:

$$l_1 + l_2 = 1, \quad l_{\text{occul}} = l_1 = 1 - l_2, \quad l_{\text{transit}} = 1 - k^2 l_1 = 1 - k^2 l_{\text{occul}}$$

$$k^2 = \frac{1 - l_{\text{transit}}}{l_{\text{occul}}}. \tag{1.41}$$

The quantity k^2 is obtained from the light curve.

At exterior contact, $r_1 + r_2 = r_1(1 + k) = d(\sin^2 \theta' \sin^2 i + \cos^2 i)^{1/2}$.

At interior contact, $r_1 - r_2 = r_1(1 - k) = d(\sin^2 \theta'' \sin^2 i + \cos^2 i)^{1/2}$.

If the phases θ' and θ'' are established with enough precision we can in principle obtain d/r_1 and $\cos i$ from these two equations. If in addition, as is the case for spectroscopic binary systems, we have radial velocity curves for the two components as functions of θ we can use the value of i found previously to find d and $m_1 + m_2$, finally obtaining r_1 and r_2.

Photometric observations have finite accuracy, the stellar orbits are sometimes elliptical, the eclipses may be partial, and the stars can be deformed by tides. These difficulties complicate the practical application of this method of determining stellar radii.

1.3.3 Indirect Radius Determinations: The Infrared-Flux Method

This method (Blackwell and Shallis 1977) obtains θ from (1.40) by using this equation at several wavelengths:

$$f(\lambda) = \frac{\theta^2}{4} F(\lambda).$$

It relies on a measurement of $f(\lambda)$, the flux received at Earth, which assumes that the energy distribution of the star is available in absolute units; this gives $\int_0^\infty f(\lambda) d\lambda$. One also measures $f(\lambda_{IR})$; stellar-atmosphere theory then gives $F(\lambda_{IR})$ and thus a first approximation to θ. For the method to work well, it is necessary that $F(\lambda_{IR})$ can be computed without having to specify the parameters of the star too closely. These parameters are the effective temperature T_{eff} (cf. (1.20)), the surface gravity defined by

$$g = GM_*/R^2 , \tag{1.42}$$

and the star's chemical composition. The surface gravity is almost never found from the stellar mass M_* and radius R, but usually by spectroscopic methods. For the purpose of finding R it suffices to note that dwarfs have gravities which are close to 10^4 cm s^{-2} (from 1 to 3×10^4), and that the IR continuum is very insensitive to g and abundances. λ is chosen in the near infrared ($\lambda \sim 2 \, \mu$m); for normal dwarfs the energy distribution is well represented by the Rayleigh–Jeans approximation to a black body: $F(\lambda) = (c_1/c_2)T_{\text{eff}}\lambda^{-4}$. The linear dependence on T_{eff} is less significant than at shorter wavelengths.

This method works by iteration: one starts with a first approximation for T_{eff} and g and calculates $F(\lambda = 2 \, \mu$m) using a model atmosphere with these parameters. As seen above, one uses (1.40) to give an estimate of θ from the measured $f(\lambda = 2 \, \mu$m). One also calculates $F(\lambda)$ for the entire visible spectrum and adjusts T_{eff} using

$$\int_0^\infty f(\lambda)\,\mathrm{d}\lambda = \frac{\theta^2}{4}\int_0^\infty F(\lambda)\,\mathrm{d}\lambda = \frac{\theta^2}{4}\sigma T_{\mathrm{eff}}^4\,.$$

The procedure is repeated to convergence.

1.3.4 Other Radius-Determination Methods

There are other methods for finding radii which have more limited application. The use of lunar or planetary occultations gives results for stars near the ecliptic plane. The varying radii of pulsating stars can be found by the Baade–Wesselink method (see Sect. 6.2).

In concluding this section on radius determinations we note the importance of measurements of distances and small angles. The development of interferometric techniques is indispensable for better radius determinations for stars, and particularly for a better picture of their surface structure. At present the Sun is the only star whose disc is resolved (see Sects. 2.4 and 2.5).

1.4 The Hertzsprung–Russell Diagram

The MK classification has two parameters: the spectral type and the luminosity class. These parameters give a first idea of the star's properties, but their limits are clear: they are discontinuous and are not easily transformed to give the physical parameters of effective temperature and luminosity (but see de Jager and Nieuwenhuijzen 1987).

We have seen (Sects. 1.1.3 and 1.2.1) that stellar photometry gives magnitudes and colour indices. These quantities are more easily converted to T_{eff} and L than the parameters of the MK classification. Consider a magnitude V and colour index $B - V$. For stars at a given distance, V gives the absolute visual magnitude M_v, which is linked to the luminosity by

$$\begin{aligned}
M_v - M_v(\odot) &= M_{\mathrm{bol}} - M_{\mathrm{bol}}(\odot) - (BC - BC(\odot)) \\
&= -2.5\log_{10}(L/L_\odot) - (BC - BC(\odot))\,.
\end{aligned} \tag{1.43}$$

The index $B - V$ (similarly other indices built in the same way) can be calibrated in T_{eff}, as indicated in Table 1.4 for dwarfs. The index $B - V$, which measures the slope of the Paschen continuum (see Sect. 1.2.1) is sensitive to interstellar absorption, which reddens the observed colour, as the extinction curve $A(\lambda)$ rises towards short wavelengths (see e.g. Allen 1976) in the optical region. In the following we consider the dereddened index $B - V$, often denoted $(B - V)_0$. To achieve the dereddening requires the *colour excess* $E(B - V)$, so that

$$(B - V)_{\mathrm{obs}} = (B - V)_0 + E(B - V)\,. \tag{1.44}$$

Table 1.4. Colour index and effective temperature as a function of spectral type for main-sequence stars (adapted from Böhm-Vitense 1981)

$B - V$	T_{eff} (K)	MK Type	$B - V$	T_{eff} (K)	MK Type
− 0.30	30 500	B0	0.70	5 540	
− 0.20	17 700	B3	0.80	5 330	K0
− 0.10	11 800	B8	0.90	5 090	
0.00	9 480	A0	1.00	4 840	K3
+ 0.10	8 530	A4	1.10	4 600	
0.20	7 910	A7	1.20	4.350	
0.30	7 450		1.30	4 100	K7
0.40	6 800	F2	1.40	3 850	M0
0.50	6 310	F7	1.50	3 500 ± 300	
0.60	5 910	G0	1.60	3 350 ± 300	M5

A plot of M_v as a function of spectral type or $B - V$, for field stars or stars of a cluster, is called a *Hertzsprung–Russell* (abbreviated to HR) or colour–magnitude diagram (C–M). Figure 1.10 shows such a diagram for stars whose distances from the Sun are known.

We note that the stars do not fill this diagram. The majority lie in a narrow band, called the *main sequence*, extending from the highest luminosities (top left) to the lowest (bottom right). The scale of absolute visual magnitudes runs from -10 to $+15$, corresponding to a factor 10^{10} in luminosity. In contrast the effective temperatures run from about $100\,000$ K to 2500 K. The stars of the main sequence are also called *dwarfs*, and have luminosity class V in the MK classification.

The *giant* branch, of luminosity class III, separates from the main sequence in the direction of lower temperatures, at roughly constant luminosity.

The *supergiants* cover an enormous region from blue to red stars, with $M_v \lesssim -5$ ($\sim 10^4$ times brighter than the Sun).

The lower part of the diagram is not empty, but contains *white dwarfs*, stars about 10^4 times fainter than the Sun. These stars have very large apparent magnitudes (cf. Sirius B, Table 1.1).

The ranges of L and T_{eff} we have given above confirm that stellar radii vary widely over the diagram. The lines of constant radius have similar slopes to that of the main sequence; they can easily be drawn using (1.20) if the M_v–$(B - V)$ diagram is calibrated, i.e. converted to one of $L/L_\odot$ vs T_{eff} (see Schwarzschild, 1958, Fig. I.5.) Supergiant radii are the largest at about $1000\ R_\odot$; those of the white dwarfs are about $0.01\ R_\odot$.
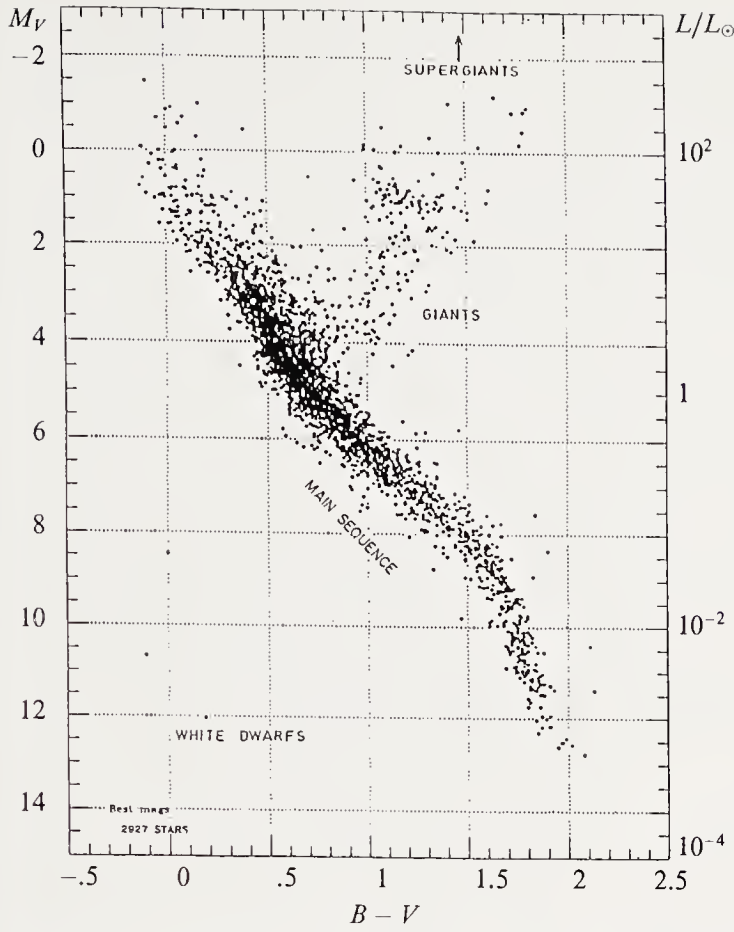
Fig. 1.10. HR diagram for stars at known distances from the Sun. This diagram was constructed for 2927 stars observed with the Hipparcos satellite during its first year of operations. The parallaxes are determined to better than 10 %, corresponding to 0.22 units in absolute magnitude. The colour scale is in the Tycho photometric system (that of the Hipparcos satellite). One has $(B - V)_{\text{Tycho}} \simeq 1.2(B - V)_{\text{Johnson}}$. (After L. Lindegren, ESA SP-349, 1992, in press. Reproduced by kind permission of the European Space Agency)

1.4.1 The HR Diagram of Open Clusters

The diagram of Fig. 1.10 shows nearby stars (field stars), the majority being dwarfs, as is the Sun itself. The errors in M_v come mainly from the distances, and those in $B - V$ from extinction and reddening corrections. The dispersion of the points is thus markedly reduced in an HR diagram of stars at the same distance and having the same extinction. A group of stars like this, all at the same distance except for the depth of the group, is a *cluster*. We have seen (Sect. 1.1.1) how the parallax of a cluster is measured. For stars in nearby clusters we then have both M_v and $B - V$. In practice one plots V as a function of $B - V$, which for a cluster is equivalent (up to the distance modulus) to M_v as a function of $B - V$.

Figure 1.11 shows the HR diagrams of several clusters, called *open* or *galactic*, which are clusters containing a few tens of stars to a few hundred

33
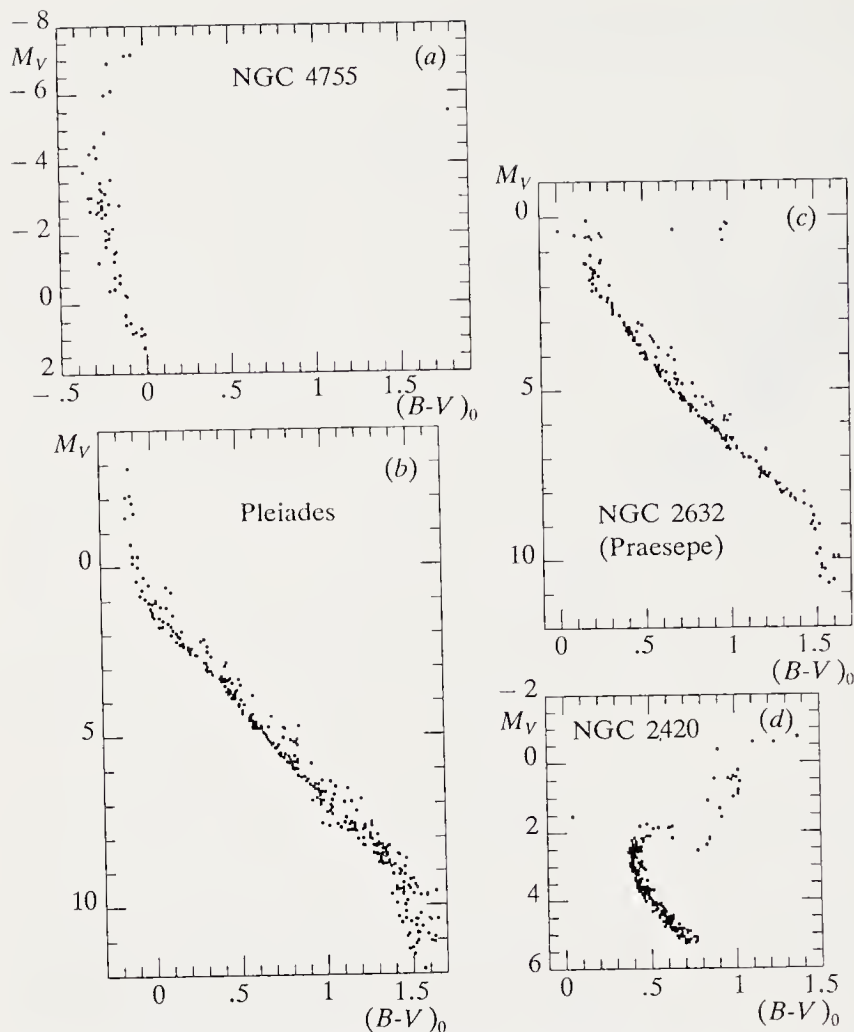
Fig. 1.11. HR (colour–absolute magnitude) diagrams of 4 open clusters. The diagrams were obtained by photoelectric photometry (private communication from J.C. Mermilliod, Observatoire de Genève)

stars in varying degrees of concentration, and having irregular shapes. We see that each of the diagrams shows a well-developed main sequence, and sometimes stars to the right of it, i.e. giants or supergiants. If we move to increasing temperatures on the main sequence (decreasing $B - V$), we note that there are no dwarfs below a certain $B - V$, or equivalently above a certain luminosity; the diagram (see e.g. NGC 2420 [2]) has a knee, and the bright stars of the cluster are to the right.

We shall show in Chap. 3 that this termination of the main sequence and the existence of a knee (*turnoff*) is an effect of *stellar evolution*, more apparent for cluster stars because these can all be assumed to have approximately the same age and initial chemical composition. The HR diagram of a cluster is thus a sign of the evolutionary state of the stars in it, as indicated by the

---

[2] The designation NGC stands for *New General Catalogue*

colour index $B - V$ giving the knee. We shall show in Chap. 3 that in the HR diagram of a cluster, which is an isochronic line of evolution (the stars there are the same age), the turnoff point gives the *age* of the cluster.

For the moment we return to an assumption stated in Sect. 1.1.4: let us imagine that the main sequences of clusters coincide, i.e. the intrinsic luminosity of a dwarf corresponds uniquely to a spectral type or to a colour index $B - V$. Then we can construct (Sandage 1957) a *composite* colour–magnitude diagram, in which open clusters are ordered by age. The main sequences are arranged to coincide, and the turnoff points where the sequences bend to the right follow each other, from $M_v = -5$ for NGC 4755 to $M_v = +3$ for NGC 2420. The main sequence thus produced has a lower envelope called the *zero-age main sequence* (ZAMS); it is the position of the unevolved stars in each of the clusters, i.e. the position of stars in which central hydrogen burning has only just begun (Chap. 3). The higher the luminosity at the turnoff point (the smaller M_v there), the younger the cluster. Figure 1.12 shows a composite colour–magnitude diagram.

The hypothesis of a zero-age main sequence is empirically justified for most galactic clusters; but it may simply be difficult to recognise the main

NGC 6231, 4755, 457, Pleiades, NGC 2516, 2287,
Praesepe, NGC 752, M67
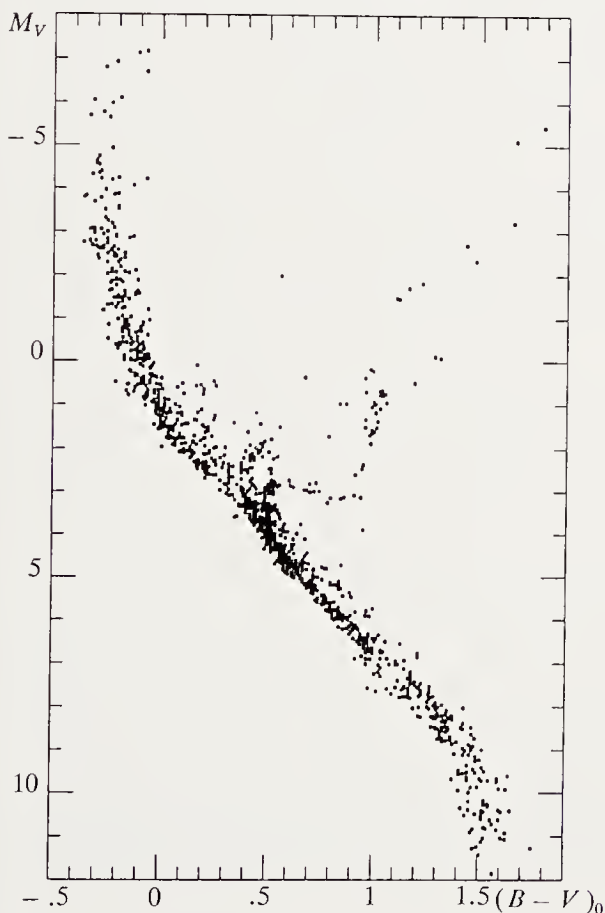
Fig. 1.12. Composite HR diagram. The absolute magnitude M_v is shown as a function of $(B - V)_0$. The diagram is constructed using HR diagrams for NGC 6231, 4755, 457, the Pleiades, NGC 2516, 2287, Praesepe, NGC 752 and M 67 (private communication from J.C. Mermilliod, Observatoire de Genève)

sequence. For example Walker's cluster (NGC 2264) has a very short main sequence, so short that it is difficult to adjust it to the ZAMS in Fig. 1.12 (on the high-luminosity side). The stars in this cluster are much younger than those represented in Fig. 1.12, and most of them have not yet arrived on the main sequence, i.e. they are still in the phase of gravitational contraction, and *a fortiori* none has yet left the main sequence, bifurcating to the right, back towards the giant branch.

1.4.2 The HR Diagram of a Globular Cluster

Figure 1.13 shows the C–M diagram of a globular cluster, M3. Globular clusters have many more stars (10^5 to 10^6) and are gravitationally more stable than open clusters. Their stars have remained bound, with few escapes, throughout the immense ages of the clusters.

The observed HR diagram ($V, B - V$) of a globular cluster is quite different from that of an open cluster. The main sequence is formed of very unevolved dwarf stars. It terminates at a later spectral type than in open clusters. This sequence bends into a branch of the subgiants, never seen in open clusters, where the evolution is too rapid for stars to be found there. It is followed by the red-giant branch which in $M_v, B - V$ coordinates extends to $M_v \sim -2$. The most remarkable part of the HR diagram of a globular cluster is the *horizontal branch*, at luminosity $M_v \simeq +0.5$, which can ex-
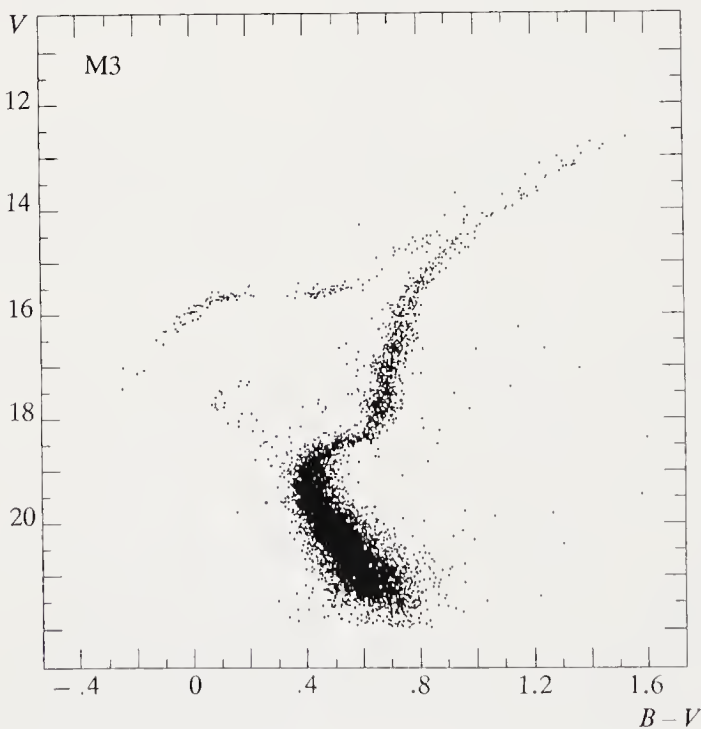
Fig. 1.13. HR diagram (colour–visual magnitude) for the globular cluster M3. The diagram was obtained by photoelectric photometry (private communication from R. Buonanno, after R. Buonanno, A. Buzzoni, C. E. Corsi, F. Fusi Pecci, A. Sandage (IAU Symp. No. 126, 621 (1987))

tend to varying distances towards the blue. In the middle of the horizontal branch there is a zone of instability which contains the RR Lyrae variables. Finally the horizontal branch joins the red-giant branch by an ascending branch which tends asymptotically to the giant branch. (See also Fig. 3.20.)

The HR diagrams of globular clusters are sensitive to the chemical composition of the material from which the cluster formed. As a result, main-sequence fitting to produce a composite HR diagram is more difficult than for open clusters. In fact stellar-evolution calculations show that the position of the ZAMS in the theoretical (i.e $L/L_\odot$ vs T_{eff}) HR diagram of a globular cluster depends on the abundance of helium and global abundance of heavy metals (metallicity). There is a dependence of the ZAMS position on the helium abundance for open clusters too, but it is less severe. The helium abundance is generally difficult to determine, except in O and B stars, which are not found in globular clusters. Moreover globular-cluster stars are metal-deficient (see Sect. 1.6).

1.4.3 Determination of M_v Revisited

The composite colour–magnitude diagram for clusters allows us to complete the discussion above (Sect. 1.1.4) on estimating distances by main-sequence fitting or by the method of spectroscopic or photometric parallaxes.

From the colour–magnitude diagram (Fig. 1.12) the colour index of any cluster star gives its absolute visual magnitude and hence its distance.

Thus for cluster stars we have a relation between $B - V$ (or spectral type) and M_v. Then study of the spectra of these stars shows that the sensitivity of the hydrogen spectrum to density, and hence luminosity, gives spectral-luminosity criteria. Thus for "late-type" stars (A, F, G) a combination of colour indices measures the size of the Balmer discontinuity, which is sensitive to the density. This is the case for Strömgren's index $c_1 = (u - v) - (v - b)$. For O and B stars it is mainly the (weak) hydrogen lines whose wings are sensitive to the density. Strömgren's index β measures the intensity of Hβ.

Table 1.5 gives the indices which are good *luminosity indicators* for stars later than A0 in the Strömgren (intermediate band) and Geneva (wide band) photometric systems, as well as in the Barbier–Chalonge–Divan spectrophotometric classification.

We can therefore use cluster stars to calibrate these luminosity indicators in M_v. The resulting calibration is then used mainly for stars outside clusters, assuming that they can be classified by two parameters only. We refer to Strömgren (1966) for the intermediate-band classification, and to Hauck (1973) and Cramer (1982) for the Geneva classification.

For stars with a chemical composition differing from that of the Sun a two-dimensional classification (T_{eff}, M_v) fails, and we must introduce the *metallicity* as a third parameter. The above references discuss this point (see also Sect. 1.6).

Table 1.5. Spectroscopic luminosity indicators

Spectral classification	Luminosity indicator (stars later than A0)	References
Strömgren Geneva Barbier–Chalonge–Divan	$[c_1] = c_1 - 0.20\,(b - y)$ $[d] = (U - B_1) - 1.430\,(B_1 - B_2)$ $[\Delta] = (U - B_2) - 0.832\,(B_2 - G)$ Balmer discontinuity D and its position λ_1	Strömgren (1966) Golay (1974) Pecker and Schatzman (1959)

1.5 Stellar Masses

Many stars in the Galaxy, and no doubt in other galaxies, belong to multiple systems, i.e. a set (two or more) of stars orbiting a common centre of mass. In general a star can be orbited either by other stars, by planets or planetary-type objects such as rocks or dust grains (as in the star β Pic). All such objects follow Kepler's laws.

Kepler's third law relates the mass of a binary system to the orbital period P and the semi-major axis a of the ellipse which is the relative orbit of the two stars. The application of this law to binary stellar systems with known orbits gives direct estimates for the masses of the stars.

1.5.1 Direct Mass Determination

Let M_1 and M_2 be the masses of two stars forming a binary system. Using Kepler's third law (with $n = 2\pi/P$ the mean angular frequency) we have

$$G(M_1 + M_2) = n^2 a^3 = \frac{4\pi^2}{P^2} a^3 \,. \tag{1.45}$$

To get $M_1 + M_2$, we need a as a length, and thus a known distance d for the system. The orbital period P is easily measured. The main error in $M_1 + M_2$ is three times the error in d. To obtain M_1 and M_2 separately we have to be able to find the absolute orbit of at least one of the stars around the common centre of mass, in a reference frame fixed to distant stars. Let a be the semi-major axis of this absolute orbit. Then

$$\frac{M_2}{M_1 + M_2} = \frac{a_1}{a} \,, \tag{1.46}$$

which gives M_2 and thus M_1.

This method of determining masses, the only direct one, assumes various conditions:

— it must be possible to resolve the angular separation θ of the two stars;
— the system's distance d must be known so as to give $a = d\theta$;
— it must be possible to determine the absolute orbit of one of the components.

These conditions restrict direct mass determinations to *visual doubles*, i.e. binaries with angular separations $0.5''$ for ground-based observations, or $0.1''$ for an instrument such as the Hubble Space Telescope. The advent of *Hipparcos* will improve the accuracy of mass determinations: it is hoped that masses accurate to 15 % will be obtained for about 100 visual binaries.

At present only about fifty stars have directly determined masses (Popper 1980); for only a dozen of these is the mass known to better than 20 %.

1.5.2 The Mass–Luminosity Relation

For some main-sequence binaries the masses M and luminosities L are known well enough that they can be plotted against each other (Fig. 3.1). It is clear from the figure that L and M increase together, with

$$L \sim M^{\alpha}, \tag{1.47}$$

where $\alpha = 4$ for $L > L_{\odot}$ and $\alpha = 2.8$ for $L < L_{\odot}$.

Assuming that this result holds for all main-sequence stars, we see that stellar masses increase along the sequence, from M type ($M = 0.1\,M_{\odot}$ for α Cen C) to O type ($M \sim 30 - 50\,M_{\odot}$). The observed luminosities have a well-defined upper limit ($M_v > -11.6$ for stars in the Galaxy), so this implies an upper mass limit of the order of $60 - 100\,M_{\odot}$. For main-sequence stars the mass–luminosity relation shows that the star's energy production is directly related to its mass, i.e. the available nuclear fuel reserve (see Chap. 3). These observational facts are of very great significance for the theory of stellar structure.

The mass–luminosity relation does not apply to giants or supergiants: some supergiants (e.g. red) have low mass. White dwarfs do not obey the relation either.

1.6 Stellar Populations

1.6.1 Introduction

We discuss here properties shared by stars having different spatial distributions in the Galaxy (and other galaxies).

The notion of stellar populations was introduced by Baade in 1944. Resolving the centre of the Andromeda galaxy (M31) and its two companions M32 and NGC 205 into stars, he was able to construct HR diagrams for

the brightest stars in these galaxies. He noticed that there were two types of HR diagram: that for the spiral arms of M31 resembled the HR diagram of the solar neighbourhood, while those for the central bulge of M31 and the stars of its two companions resembled HR diagrams of globular clusters. Baade called the stars of the solar neighbourhood and the spiral arms of M31 *population I*, and designated as *population II* the stars of the elliptical galaxies M32 and NGC 205, the bulge of M31, and the globular clusters, as well as the RR Lyrae and short-period Cepheid variables.

The idea of a stellar population is now refined to mean stars having not only the same spatial distribution in the Galaxy, but also the same age, kinematic properties, and chemical composition. This set of common properties of the population is based on a model of the dynamical and chemical evolution of the Galaxy, which we shall not discuss here. The position of a star in the Galaxy and its related properties bear the imprint of the history of the Galaxy from its formation, since each star is assumed to have the same chemical composition as the interstellar medium where it was born.

1.6.2 Defining Criteria for Stellar Populations

Here we restrict ourselves to the Galaxy, but the concepts we shall introduce are clearly relevant for other galaxies.

Stellar Spatial Distribution. In a spiral galaxy such as our own the spatial distribution of the stars allows us to distinguish a spheroidal part consisting of a *bulge* and a *halo*, together with a flattened *disc* extending into spiral arms. Table 1.6 summarizes the types of stars mainly or exclusively present in each of these components. This tendency for certain types of stars to occur in different parts of the Galaxy is revealed by systematic number counts in different volume elements.

Spheroidal and Flattened Distribution. The distinction between these two distributions largely coincides with that between the two main kinematic groups. High-velocity stars are largely members of Population II or of the halo, while low-velocity stars are mainly in the solar neighbourhood and the disc. The measured velocity is the space velocity; we often consider the more easily measured radial velocity. High-velocity stars are those with space velocities exceeding 65 km s^{-1} (with respect to the Sun). For comparison, the escape velocity from the Galaxy at the Sun's position is 360 km s^{-1}.

Velocity distributions for spectral types and groups of stars give the *velocity dispersion* as the second moment of the distribution for each group of stars. It is well established that the velocity dispersion is low (1 km s^{-1}) for stars in young associations and open clusters, and grows with age of the stars.

Table 1.6. Stellar population of the galactic halo and disc (after Mould 1982)

Spheroidal population (halo)	Flattened population (disc)

Globular clusters
Metal-poor RR Lyrae variables
 	
Field halo stars
– Subdwarfs
– M and carbon-rich giants
– Some planetary nebulae
  and white dwarfs
– High-velocity A stars
 	
Bulge
– Giants
– OH/IR stars
– Long-period variables
– X-ray burst sources
 	
Outer halo
– Distant globular clusters
– Dwarf spheroidal galaxies
– Magellanic clouds
	Old disc population
	– Dwarf F and M stars
	– Giants
	– Metal-rich RR Lyrae variables
	– Long-period variables
	– C and S stars
	– Barium stars
	– The majority of planetary nebulae
	  and white dwarfs
	
	Young disc population
	– Bright O and B stars
	– A and B dwarfs
	– Pulsars (neutron stars)
	– Mc dwarfs
	– Open clusters younger than the
	  Hyades
	– T Tauri and Herbig stars

We should however note that there exist high-velocity stars in the young disc population: for example O stars of mass above 40 $M_\odot$ almost all have high velocities.

Stellar Chemical Composition. The chemical composition of stars varies according to the part of the Galaxy where they were born. The composition is characterised by the *metallicity index*. Defined spectroscopically, it is written [Fe/H], standing for

$$\log(\mathrm{Fe/H})_* - \log(\mathrm{Fe/H})_\odot \,,$$

[Fe/H] denoting the relative abundance by number of iron relative to hydrogen in the star compared with the solar value; iron is regarded as representative of the abundances of all the metals. Defined from a photometric system (e.g, UBV) the metallicity index measures the effect of metal lines on the energy distribution in certain passbands, for a given effective temperature. One of these indices is $\delta(U - B) = (U - B)_{\mathrm{st}} - (U - B)_{\mathrm{obs}}$, where $(U - B)_{\mathrm{st}}$ denotes the standard colour of normal stars and $(U - B)_{\mathrm{obs}}$ is the observed index; we measure $\delta(U - B)$ corresponding to the observed value of a good indicator of T_{eff}, such as $B - V$ here. Similar indices are defined in the Strömgren system (Δm) or the Geneva system.

It is clearly established that the UV excess $\delta(U - B)$ is very well correlated with the star's metallicity [Fe/H], its distance from the galactic plane, and its space velocity. The excess grows with distance from the galactic plane, with space velocity, and with decreasing metallicity. Metal-deficient stars are defined as those with [Fe/H] < -1; metallicities as low as [Fe/H] $= -3$ have been found in halo and globular-cluster stars.

Of course there are exceptions to these correlations. The chemical composition is not solely a function of distance from the galactic plane: there is a clear gradient of chemical composition with the distance from the galactic centre within the disc. Stars closer to the centre than the Sun are metal richer, and those at the disc edges metal poorer. The metallicity gradient near the Sun is ~ -0.05 kpc^{-1} in units of [Fe/H]. By contrast, the metallicity distribution in globular clusters is bimodal, so that naive correlation of [Fe/H] with cluster age is not possible.

This raises the question of the possible existence of population III: stars without metals, representing the oldest population, formed from a primordial "interstellar" medium consisting only of hydrogen and helium. This population need not exist, in that Cayrel (1986) shows that the oldest globular clusters (those of the lowest metallicity) could have formed from a medium already enriched by the explosions of (still more primordial) supermassive stars, which no longer exist.

Studies of distances, space velocities, and metallicities for stars, including the faintest in the Galaxy, are being extended to other galaxies, particularly by means of the Hubble Space Telescope. Existing instruments have already enabled the study of stellar populations and chemical evolution for galaxies in the local group, such as the Magellanic clouds and M31.

2. The Sun: The Nearest Star

2.1 Introduction

The unusual feature of the Sun as a star is its closeness to the Earth where from we observe it. The average Earth–Sun distance is $1.495\,979 \times 10^{13}$ cm (by definition one astronomical unit); this distance varies between 1.4710×10^{13} cm (perihelion) and 1.5210×10^{13} cm (aphelion) through the orbit.

There are several remarkable consequences of this proximity. First, the Sun has an angular diameter of 32′, allowing detailed study of its surface. The best spatial resolution in ground-based observation is about 0.25″; from space better than 0.1″ is obtainable. As the Sun is also the star whose radius is known most accurately (see Table 2.1) this means that we can distinguish parts of the Sun's surface down to about 180 km from the ground and 70 km from space. Until very recently (supernova 1987A) the Sun was the only star from which we could detect particles: neutrinos, high-energy ions and electrons ejected in flares, and solar-wind particles. The Sun is still the only star from which particles emitted in transient events (flares) or steadily blown away (solar wind) can be counted, rather than being detected by

Table 2.1. General solar properties

Radius (R)	6.95997×10^{10} cm
Mass (M)	1.9892×10^{33} g
Photospheric gravity (g)	2.7398×10^{4} cm s^{-2}
Effective temperature (T_{eff})	5 770 K
Synodic rotation period of the sunspot zone	$13.45° - 3.0 \sin^2 \varphi$ per day (φ = heliographic latitude): $P = 26.76$ days at the equator
Rotation rate (Ω) and synodic period (P) (photosphere, at the equator)	$\Omega = 2.67 \times 10^{-6}$ rad s^{-1}, $P = 26.24$ days
Rotation speed (photosphere, at the equator) (V_{eq})	2 km s^{-1}

Table 2.2. Energetics of the Sun

Total luminosity (L)	3.83×10^{33} erg s^{-1}
Solar constant (at Earth)	1.360×10^6 erg cm^{-2} s^{-1} or 1.360 kW m^{-2}
Mass loss rate (dM/dt)	0.6–$1.0 \times 10^{-14}\ M_\odot$/y or 10^{12} g s^{-1} or 10^{-36} particles s^{-1}
Binding energy ($M^2 G/R$)	3.8×10^{48} erg
Rotational energy $\left(\dfrac{1}{5} M\Omega^2 R^2\right)$	1.4×10^{43} erg
Angular momentum (from surface rotation)	1.63×10^{48} g cm^2 s^{-1}

Notes: (1) the binding energy, rotational energy, and angular momentum in fact depend on the mass distribution and the behaviour of Ω with depth inside the Sun. (2) The solar constant varies in the course of the solar/activity cycle, by about 0.1% from maximum to minimum. It has its maximum at the maximum of solar activity (results obtained between 1980 and 1989 by the *Solar Maximum Mission* satellite (*SMM*)).

their radiative effects. Neutrinos have now been detected from one other object (supernova 1987A) besides the Sun.

In its other properties the Sun is a very ordinary star. Its spectral type is G2V: it is a dwarf. Its age is 4.6×10^9 y: it is quite old. There is nothing special about its position in the Galaxy: it is located about 10 kpc from the Galactic centre. But every property of the Sun is known to far higher precision than for any other star. Table 2.1 summarizes general properties of the Sun and Table 2.2 its characteristics as an energy source.

Because it is the star whose mass, luminosity, and radius are best known, the Sun is the touchstone for all theories of stellar properties: evolution, heating of the exterior layers, generation of magnetic fields, acceleration of the wind, or large-scale motions in the interior. The Sun is a universal stellar prototype, which is the reason for devoting a chapter to it.

The Sun is also a laboratory for plasma physics. Since the end of the 19th century, laboratory spectroscopy and the identification of lines in the solar spectrum have developed in close symbiosis (discovery of helium in 1868, laboratory identification by Ramsey in 1895; identification of coronal lines by Grotrian and Edlèn between 1939 and 1942). Later, non-LTE (local thermodynamic equilibrium) diagnostic analysis was largely established through quantitative study of the solar spectrum. Today the spectroscopy of the solar corona is closely linked with that of fusion plasmas.

We remark finally that the Sun provides units of measurement for all stellar quantities: masses, radii, luminosities, and mass-loss rates are always given in solar units (see Chap. 1).

In the following we give only a brief introductory picture of the Sun and study some aspects of its physics. We treat the Sun here as static and in a steady state, even though its time variability is very important; we shall consider it together with that of other stars (Chaps. 6 and 7). The Sun's structure and evolution are studied in Chap. 3, the expansion of its atmosphere (wind) and the possibility of collecting solar particles in Chap. 4. The relation of the Sun to its system of planets is considered in the book by Encrenaz and Bibring (1990). The influence of the Sun specifically on the Earth is an essential part of the climatology of our planet, not discussed here.

Observation of the Sun from X-rays to radio waves at the highest spatial resolution reveals an object of fascinating surface complexity. Initially we shall ignore these surface inhomogeneities ("fine structure") and idealize the Sun as a sphere of gas in which every property is a function of radius only. This allows us to apply *stellar-atmosphere theory* to the Sun, as if it were a star, i.e. seen without spatial resolution.

2.2 Models of the Solar Atmosphere

Although the Sun has non-radial structure in the surface layers forming its atmosphere, physical quantities (electron temperature T_e, pressure P, density ρ, electron density N_e, etc) in these layers were first studied using *average* homogeneous models. As a next step, models adjusted to each of the observed surface structures can be computed by the same methods.

2.2.1 The Different Regions of an Atmosphere

We introduce these regions in a general manner, so that the same terminology can be used for stars other than the Sun. First, by the *atmosphere* we mean the layers through which the energy produced at the centre of the Sun, here propagating mainly as radiation, interacts with matter to produce the photons we observe. Here the idea of opacity is important. We define a linear absorption coefficient χ_ν at each frequency, and a mass absorption coefficient κ_ν, related by $\chi_\nu = \kappa_\nu \rho$. The first has dimensions $[\mathrm{cm}^{-1}]$, the second $[\mathrm{cm}^2 \mathrm{g}^{-1}]$.

The observed photons come from different layers depending on their frequency. In the spectrum of every star we find frequencies where the material is very transparent, i.e. the absorption coefficient $\chi_\nu = \kappa_\nu \rho$ is small, and at such frequencies we see deep into the atmosphere. Similarly there are frequencies where the material is very opaque; at these frequencies we can see only the upper layers of the atmosphere. The (radial) *optical depth* τ_ν,

defined by $d\tau_\nu = -\kappa_\nu \rho dr$, gives the size of the region of the star producing the observed photons. At a given ν, the received photons are such that $10^{-3} \lesssim \tau_\nu \lesssim 100$, $\tau_\nu \simeq 1$ being a representative value. The atmosphere is thus the zone between the geometrical radial depths r_1 and r_2 such that

$$\tau_\nu = -\int_{r_1}^{\infty} \kappa_\nu \rho \, dr \simeq 100$$

at the most transparent frequency, and

$$\tau_\nu = -\int_{r_2}^{\infty} \kappa_\nu \rho \, dr \simeq 10^{-3}$$

at the most opaque frequency.

In the Sun, the frequency where the matter is most transparent is in the near infrared ($\lambda = 1.6\ \mu$m), and the most opaque frequencies are in the core of the Lyα line and in the centimetre radio region. These frequencies differ for stars of other spectral types.

The reference temperature in the atmosphere is the effective temperature T_{eff} already defined as that of a black body radiating the same amount of energy as the star,

$$L = 4\pi R_*^2 \sigma T_{\text{eff}}^4 \,, \tag{2.1}$$

where L is the luminosity, R_* the radius, and σ Stefan's constant. The classical radius of the star is that of the layer where $T_e = T_{\text{eff}}$, which for all spectral types is characteristic of the formation of the visible continuum.

The notion of the star's radius must be made precise. An observer viewing the Sun at frequency ν (Fig. 2.1) "sees" on average down to $\tau_{l,\nu} \simeq 1$ at each point of the disc ($\tau_{l,\nu}$ here means an optical depth along the line of sight). He thus sees down to the dashed curve in the figure. Its shape results from seeing in more deeply at the centre of the disc, where the line of sight is vertical, than at the edge, where the rays received by the observer are highly inclined to the normals to the layers and are thus more absorbed,
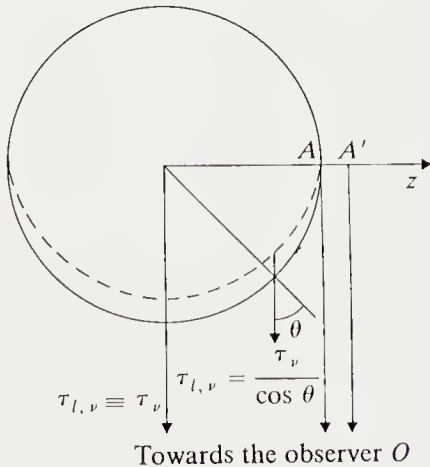
Fig. 2.1. The solar limb. θ is the angle between the normals to the layers and the line of sight, $\tau_{l,\nu}$ is the optical depth along the line of sight, τ_ν the radial optical depth, and z the depth variable in the plane-parallel approximation. The *dashed* curve is the locus $\tau_{l,\nu} = 1$
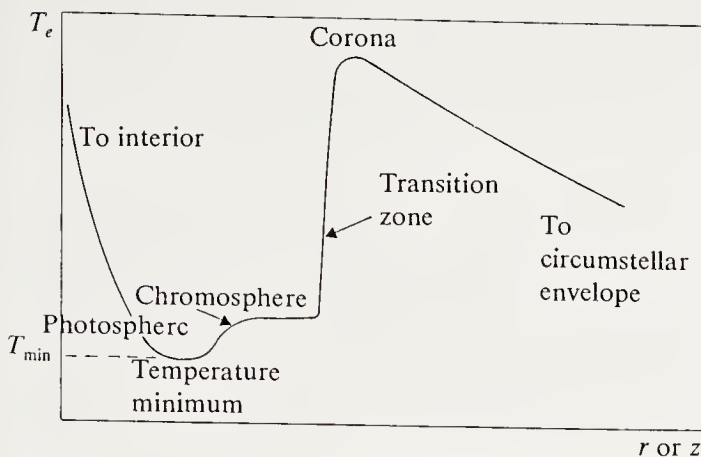
Fig. 2.2. The various regions of a stellar atmosphere, classified by their variation of electron temperature with geometrical depth z

since $\tau_{l,\nu} = \tau_\nu/\cos\theta$. We define the edge (or *limb*) of the disc as the point where $\tau_{l,\nu}$ goes from 1 to a near-zero value, for a displacement Δz of the tangent ray OA small compared to R_*, because of the variation of the density ρ with z: to a very good approximation we have $\rho = \rho_0 e^{-z/H}$, with a scaleheight H of the order of 100 km in the Sun. We note that $H/R_\odot = 10^{-4}$. For $\Delta z = AA' \simeq 3H$ there is too little material along the line of sight OA' for $\tau_{l,\nu}$ to be $\simeq 1$. This means that we can see a sharp edge to the Sun (other stars also have sharpe edges, but we cannot resolve their discs) and that the Sun does not have the same radius at different frequencies. For example the solar radius in the decametric radio region is about 1.8 times its visible radius.

Figure 2.2 shows schematically the various regions of a stellar atmosphere, as in the Sun. Each region corresponds to a distribution $T_e(r)$, or $T_e(z)$ if the atmosphere is thin enough compared to R_* to be regarded as plane-parallel rather than spherically symmetrical. We consider this approximation more precisely, for a vector V whose radial component V_r is non-zero. If the scaleheight H defined by $1/H = \partial \ln V_r/\partial r$ obeys $H \ll R_*$, we also have $z \simeq r$ and div $V = \partial V_r/\partial r + 2V_r/r \simeq \partial V_r/\partial r$. This is easily satisfied if V is the radiation flux integrated over all frequencies, F_R, in a stellar atmosphere.

The names for the various regions of an atmosphere have empirical origins. However, it is more illuminating first to distinguish these regions by their dominant modes of energy transport, which fix $T_e(r)$; we shall do this below. We refer to stellar-atmosphere theory for the radiation-dominated regions and hydrodynamics and magnetohydrodynamics for the outermost layers.

In the *photosphere*, energy is transported essentially by radiation. Then in the stationary approximation the local energy equation at depth z is the condition of *radiative equilibrium*:

$$\text{div } \boldsymbol{F}_R(z) = 0. \tag{2.2}$$

47

We have

$$F_R(z) = \int_0^\infty F_\nu(z)\mathrm{d}\nu = \sigma T_{\mathrm{eff}}^4\,, \quad \text{with} \quad \boldsymbol{F}_\nu(z) = \int_\Omega I_\nu(\boldsymbol{n},z)\boldsymbol{n}\,\mathrm{d}\omega\,, \quad (2.3)$$

where $\boldsymbol{n}$ is the propagation direction of the radiation, $I_\nu(\boldsymbol{n},z)$ its specific intensity, and $\mathrm{d}\omega$ the element of solid angle around $\boldsymbol{n}$.[1] From the base of the photosphere to its outermost parts, different frequency domains become successively transparent, i.e. conditions in the stellar material cause τ_ν to range from ~ 100 to about 10^{-3} at each ν. These frequency ranges define sub-regions; in each the intensity is found as the solution of the transfer equation.

In the stellar interior, matter is by contrast opaque at all frequencies and characteristic scales; the flux can be found from a local approximation (quasi-isotropic, or diffusion) and has the form of a heat flux:

$$F_R = -K_R \frac{\mathrm{d}T}{\mathrm{d}z}\,, \quad \text{with} \quad K_R = +\frac{16\sigma}{3}T^3/\kappa\rho\,, \quad (2.4)$$

where κ is a mean absorption coefficient and K_R the radiative conductivity.

For stars of some spectral types there is convective transport in the photosphere. Then the local energy-transport equation becomes

$$\mathrm{div}(\boldsymbol{F}_R(z) + \boldsymbol{F}_c(z)) = 0\,, \quad (2.5)$$

where F_c is the convective flux. The theory of convection is discussed in Chap. 5, together with the conditions for convection to occur.

In the *chromosphere*, radiation carries off a non-radiative energy contribution which may be due to wave dissipation (acoustic, hydrodynamic, or magnetohydrodynamic). The local energy equation is then

$$\mathrm{div}\,\boldsymbol{F}_R(z) = \dot{Q}_{\mathrm{chrom}}(z)\,. \quad (2.6)$$

The non-radiative, or mechanical, heating term $\dot{Q}_{\mathrm{chrom}}$ raises the local electron temperature above the value it would have had in radiative equilibrium. At present we have a crude theory of convective transport and a very elaborate theory of the radiative flux; but the physics of the heating term $\dot{Q}_{\mathrm{chrom}}$ remains poorly known. The typical temperature range in a chromosphere is 6000–10 000 K.

In the *transition zone*, i.e the region between chromosphere and corona, matter is optically thin. The electron temperature varies from 3×10^4 to 3×10^5 K over a few hundred kilometres. The energy equation reduces to

$$\mathrm{div}(\boldsymbol{F}_R(z) + \boldsymbol{F}_{\mathrm{cond}}(z)) = 0\,, \quad (2.7)$$

---

[1] We shall always use the flux defined in (2.3) in the following. In the literature there exists an "astrophysical" flux defined as $F_\nu(z)/\pi$.

assuming no matter motions; it is unclear whether this is justified for the Sun. $\boldsymbol{F}_{\text{cond}}$ is the conductive flux, which is significant because the temperature gradient is large:

$$F_{\text{cond}} = -K \, \mathrm{d}T_e/\mathrm{d}z \,. \tag{2.8}$$

K is the thermal conductivity of the matter, with $K = 1.1 \times 10^{-6} T_e^{5/2}$ in cgs units, for a fully ionized gas.

In the *corona*, the temperature reaches 10^6 to 10^7 K. Matter is optically thin. The energy equation contains a wind term, taking account of the expansion of the medium at velocity v, and a non-radiative-heating term $\dot{Q}_{\text{cor}}$, so that

$$\mathrm{div}(\boldsymbol{F}_R + \boldsymbol{F}_{\text{cond}} + \boldsymbol{F}_{\text{conv}}) = \dot{Q}_{\text{cor}} \,, \tag{2.9}$$

with

$$\boldsymbol{F}_{\text{conv}} = 4\pi\rho v r^2 \left(\frac{1}{2}v^2 + h - \frac{GM_\odot}{r} \right) \,,$$

where the first term is the kinetic energy flux of the wind, the second its thermal flux (h is the specific enthalpy), and the third the gravitational energy flux. The term $\dot{Q}_{\text{cor}}$ is still the subject of active research: Joule heating, magnetic reconnection, and wave dissipation are some of the phenomena invoked to heat the solar corona (see Chap. 4). The plane-parallel approximation is no longer valid in the corona; r replaces z for average models.

A stellar wind is not always optically thin as in the Sun. In many types of star (blue and red giants and supergiants, pre-main-sequence stars, Wolf–Rayet stars, etc.) it is dense and optically thick, and its spectroscopic effects are seen first in the upper photosphere and chromosphere (Chap. 4).

In some stars no corona is detected, but only a thick chromosphere. The geometrical depth of this layer is about $1R_*$. All such chromospheres are observed to be expanding (red giants and supergiants, pre-main-sequence stars).

Above the corona (or thick chromosphere) there may be an envelope involving more than gas: a circumstellar dust envelope. Envelopes of this kind surround stars of many spectral types: pre-main-sequence stars, the Sun (the F corona), red giants and supergiants, Mira variables, etc.

2.2.2 Models of the Mean Solar Atmosphere

Theoretical Photosphere Models. The photosphere is the only region of an atmosphere for which one can compute a theoretical model. In the absence of convection, the energy equation $\mathrm{div}\,\boldsymbol{F}_R = 0$ can be written explicitly in the plane-parallel approximation at each depth z as

$$\frac{\mathrm{d}\boldsymbol{F}_R}{\mathrm{d}z} = 0 = -4\pi \sum_i \int_{\nu_{0,i}}^{\infty} \chi_{\nu_i}(J_\nu - S_\nu)\,\mathrm{d}\nu$$

$$+ \sum_{i,j>i} h\nu_{ij} A_{ji} n_j \left(1 - \frac{\int_0^\infty J_{\nu_{ij}}\varphi_\nu \mathrm{d}\nu}{S_{ij}}\right). \tag{2.10}$$

Each of the terms on the right represents a balance of emissions minus absorptions of radiation; the first refers to bound–free atomic transitions, the second to bound–bound transitions, i.e. lines formed at frequencies ν_{ij} between levels i, j. χ_{ij} is the continuous absorption coefficient, of dimensions $[\mathrm{cm}^{-1}]$,

$$\chi_{\nu i} = n_i \alpha_{\nu i} = \kappa_\nu \rho. \tag{2.11}$$

A_{ij} is the Einstein spontaneous-emission coefficient, $\alpha_{\nu i}$ the photoionisation cross-section, and n_i and n_j are the level populations. J_ν is the mean intensity, or zeroth moment of the specific intensity,

$$J_\nu = \frac{1}{2} \int_{-1}^{+1} I_\nu(\mu)\mathrm{d}\mu \quad \text{with} \quad \mu = (\boldsymbol{n} \cdot \boldsymbol{S}) = \cos\theta. \tag{2.12}$$

$\boldsymbol{S}$ is the unit vector normal to the surface element $\boldsymbol{S}$ that the radiation crosses. The physical flux F_ν is the first moment multiplied by 4π, i.e.

$$F_\nu = 4\pi \frac{1}{2} \int_{-1}^{+1} I_\nu(\mu)\mu\,\mathrm{d}\mu \tag{2.13}$$

(see also (2.3)), S_ν and S_{ij} are the source functions in the continuum and the transition ij, and ϕ_ν is the absorption profile of the line ij.

Equation (2.10) immediately shows that at each z we need consider only frequencies such that:

(a) $\chi_{\nu i}$ (for the continuum) or $A_{ij} n_j$ (lines) are not very small: if the medium is transparent it has no effect on the radiative balance;

(b) J_ν differs from S_ν; equality holds in the quasi-isotropic case referred to above, i.e. when the medium is opaque at the frequencies considered.

To solve (2.10) we need J_ν, i.e. a solution of the transfer equation at frequencies which are neither transparent nor opaque. In plane-parallel geometry this equation is

$$\mu \frac{\mathrm{d}I_\nu}{\mathrm{d}z} = -\chi_\nu I_\nu(\mu, z) + S_\nu. \tag{2.14}$$

The source function S_ν is in general a function of T_e and of the radiation fields in the continuum or the lines.

The other equations required to find T_e from (2.10) are (in the static case) as follows.

50

— Hydrostatic equilibrium:

$$\frac{\mathrm{d}P}{\mathrm{d}z} = -\rho g \, , \tag{2.15}$$

where g is the gravity and P the total pressure,

$$P = P_g + P_R \, , \tag{2.16}$$

where P_g is the gas pressure and P_R the radiation pressure,

$$P_R = \frac{4\pi}{c} \int_0^\infty K_\nu \mathrm{d}\nu \, , \tag{2.17}$$

where K_ν is the second moment of the specific intensity,

$$K_\nu = \frac{1}{2} \int_{-1}^{+1} I_\nu(\mu)\mu^2 \mathrm{d}\mu \, . \tag{2.18}$$

— The equation of state of a perfect gas:

$$\begin{aligned} P_g &= k(N_a T_a + N_{\mathrm{ion}} T_{\mathrm{ion}} + N_e T_e) \\ &= NkT_e = (N_a + N_{\mathrm{ion}} + N_e)kT_e \, , \end{aligned} \tag{2.19}$$

assuming equality of the kinetic temperatures of all constituents of the gas (justified if collisions are frequent), N_a, N_{ion}, N_e being the number densities of atoms, ions, and electrons respectively; k is Boltzmann's constant.

The density ρ is given by

$$\rho = N_H m_H \left(1 + 4\frac{N_{He}}{N_H} + \overline{m}\frac{N_{\mathrm{met}}}{N_H} \right) \, , \tag{2.20}$$

where m_H is the proton mass, N_H the number density of hydrogen nuclei; $\overline{m}$ is the mean mass of elements heavier than helium, collectively termed "metals", N_{met}/N_H their abundance relative to hydrogen by number; N_{He}/N_H is the abundance by number of helium.

— The equation of charge conservation (of a plasma) or electrical neutrality:

$$N_e = \sum_l \sum_k k N_k \, , \tag{2.21}$$

where N_k denotes the number of ions in state k belonging to the element l of the medium; k is the charge of this ion. We thus have to solve an ionisation equation to get the various N_k corresponding to the element l, such that $\sum_k N_k = N_l$. Moreover $\sum_l N_l = N_{\mathrm{ion}}$.

— Similarly, to find the occupation numbers n_i and n_j of the levels of a given ion we have in general to solve a set of conservation equations, the equations of *statistical equilibrium*:

$$\frac{\partial n_i}{\partial t} = \mathcal{P}_i - \mathcal{R}_i \,, \tag{2.22}$$

where $\mathcal{P}_i$ stands for the processes populating level i, and $\mathcal{R}_i$ for those de-populating it. The form of the terms $\mathcal{P}_i, \mathcal{R}_i$, the radiation source functions, and methods of solving the transfer equation are given by Heyvaerts et al. (1992) or by Mihalas (1978).

To construct a theoretical model of a stellar photosphere we solve (2.10–22) simultaneously, with the boundary conditions:

— no radiation incident from outside the atmosphere;

$\forall \nu \quad I_\nu(\mu) = 0 \quad \text{for} \quad \mu < 0 \text{ at the surface} \,,$

— radiation becomes quasi-isotropic at all frequencies for $\tau_\nu \gg 1$: in this case

$$I_\nu(\tau_\nu, \mu) = B_\nu(\tau_\nu) + \mu \frac{\mathrm{d}B_\nu}{\mathrm{d}\tau_\nu} \quad \text{and} \quad J_\nu \cong S_\nu = B_\nu(T_e(\tau_\nu)) \,,$$

where B_ν is the Planck function;
— gas pressure tends to zero at the surface.

The radiation flux integrated over all ν must be constant with z and the value of this constant gives the effective temperature of the model (cf. (2.3)):

$$F_R(z) = \int_0^\infty F_\nu(z)\mathrm{d}\nu = \sigma T_{\mathrm{eff}}^4 \,. \tag{2.23}$$

A theoretical model of the photosphere in radiative equilibrium is then characterised:

— by two parameters: the effective temperature T_{eff} and gravity g, as well as the abundance of the elements of the medium $(N_{He}/N_H, N_{\mathrm{met}}/N_H)$,
— by LTE or not (NLTE): in LTE the ionisation equations and statis-tical equilibrium equations giving the occupation numbers n_i and n_j are replaced by the Saha and Boltzmann equations respectively. In this case the n_i, n_j, N_k depend only on temperature T_e for a given density. The LTE approximation has to be checked whenever it is used. It holds only if col-lisions alone determine the atomic-level populations. In the general case where the level populations are determined by both collisional and radia-tive processes, the ionisation and statistical-equilibrium equations must be solved in detail. The n_i, n_j, N_k then depend not only on T_e but on the radia-tion field at the relevant transitions, and there is complete coupling between the transfer equation, giving the radiation distribution functions, and the statistical balance equations, giving the material distribution functions.

If part of the energy is transported by convection at the photosphere we must calculate the convective flux at each depth (see Chap. 3), and (2.23) is replaced by

$$F_R(z) + F_c(z) = \sigma T_{\text{eff}}^4 \,. \tag{2.24}$$

LTE photosphere models for the Sun are given by Kurucz (1974). For other stars similar models are in many cases the only ones available. Thus the LTE models of Kurucz (1979) cover the range $T_{\text{eff}} = 50\,000\text{--}5500$ K. The first NLTE models are by Mihalas (1972) (range $T_{\text{eff}} = 50\,000\text{--}15\,000$ K), while Borsenberger and Gros (1978) extend the grid of NLTE models to $10\,000$ K.

For atmospheric layers outside the photosphere the formal energy equations given above (2.6, 7, 9) are not usable, since we do not know the terms containing dynamical and magnetic effects well enough. Thus *semi-empirical models* have been developed for these layers of the Sun, to find the distribution of T_e with z. As we shall see, this approach has been applied to the photosphere also.

Semi-empirical Solar Photosphere.   (a) *Deep photosphere*. We take advantage of the fact that an observer sees solar radiation at different emergent angles θ_i ($\mu_i = \cos\theta_i$). The method proceeds by *inversion of data from centre to limb of the disc*. Absolute photometric calibration gives the emergent intensity $I_\nu(0, \mu_i)$ in energy units (erg cm^{-2} s^{-1} Hz^{-1} str^{-1}). The formal solution of the transfer equation is

$$I_\nu(0, \mu_i) = \int_0^\infty S(t_\nu) e^{-t_\nu/\mu_i} \, dt_\nu / \mu_i \,, \tag{2.25}$$

which in the photosphere we may always write as

$$I_\nu(0, \mu)_i = S(t_\nu = \tau_*) \,, \tag{2.26}$$

where τ_* is by definition the "formation depth" of the radiation.

The *Eddington–Barbier* approximation assumes that the source function varies monotonically with t_ν, e.g. $S = a + bt_\nu$. Then it is easy to show from (2.25) and (2.26) that $\tau_* = \mu_i$.

If we measure a *continuum* formed in LTE we have

$$S(t_\nu) = B_\nu(T_e(t_\nu)) \,. \tag{2.27}$$

We can always define a brightness temperature T_b for the observed radiation:

$$I_\nu(0, \mu_i) = B_\nu(T_b(\mu_i, \nu)) \tag{2.28}$$

(as an example, for $\mu = 1$ and $\lambda = 1.6$ μm, $T_b = 6800$ K).

Then from the above

$$T_e(t_\nu = \mu_i) = T_b(\mu_i, \nu) \,. \tag{2.29}$$

To model the deep solar photosphere we use several frequency ranges where, if possible, LTE provides a good approximation for the source function. In practice the range 1–2.5 μm is convenient for studying the deep photosphere. Once the temperature $T_e(t_\nu)$ has been found using various frequencies ν and angles μ_i, we have to order all the t_ν on the same optical-depth scale, chosen for convenience at 5000 Å, τ_{5000}, using

$$\mathrm{d}t_\nu = -\kappa_\nu \rho \, \mathrm{d}z \,,$$
$$\mathrm{d}\tau_{5000} = -\kappa_{5000} \rho \, \mathrm{d}z \,, \tag{2.30}$$

and calculate the other quantities $(P, N_e, \rho,...)$ from (2.15–22). The designation τ_{5000} for the reference optical depth is standard even though we are working in frequency units.

Figure 2.3a shows a compendium of semi-empirical models of the deep solar photosphere, which we may compare with the theoretical model of Kurucz (1974).

(b) *Upper photosphere.* The same inversion method as above is applicable down to the temperature minimum, which is probed in the far IR (100–200 μm) or the UV near 1600 Å.

However, we may also find a model of the upper photosphere by a trial-and-error method. Instead of getting the electron temperature from T_b we assume a priori a function $T_e(z)$. We retain all the equations describing the theoretical photosphere except for the energy equation, which we replace by this assumption. This model allows one to compute the emergent intensity at all frequencies where the radiation is formed in the layers considered. We can compare these intensities with those measured in as large a ν domain as possible. Then we adjust $T_e(z)$ and iterate. Figure 2.3b summarizes the state of upper-solar-photosphere models in 1977. Maltby et al. (1986) give an improved version.

We note that the theoretical model of Kurucz has a smaller gradient $\mathrm{d}T_e/\mathrm{d}\tau_{5000}$ than the semi-empirical models in the deep photosphere (Fig. 2.3a). We see also that the temperature in this model continues to decrease above the temperature minimum (Fig. 2.3b). The surface cooling of this theoretical model comes from the fact that it is a pure photosphere, while in the Sun the chromosphere is heated by non-radiative energy dissipation. A theoretical photosphere in radiative equilibrium has a surface temperature T_0, which in the plane-parallel case is reached and remains constant once all the constituents of the medium have become transparent. From Fig. 2.3b we see that there are still opaque layers in the Kurucz model at depths $\tau_{5000} = 10^{-5}$, since the temperature has not stabilised; this results from the fact that Kurucz's calculation contains many lines (almost
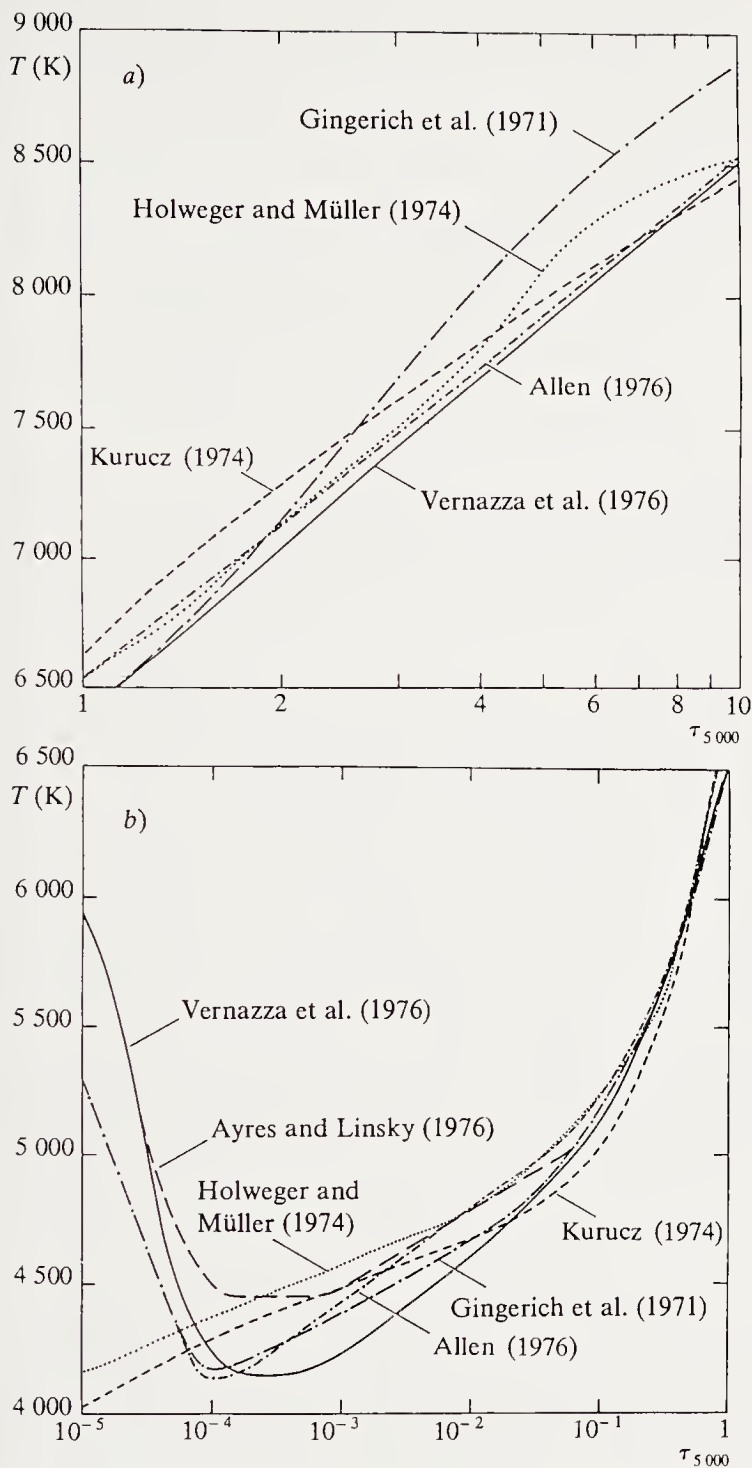
Fig. 2.3. Models of the solar photosphere. (a) deep photosphere, (b) upper photosphere. (After E.H. Avrett in *The Solar Output and its Variations*, ed. by O.R. White, Colorado Associated University Press, 1977)

10^6) whose radiation still interacts with the matter at reference depths as low as 10^{-5}.

What is the value of the minimum temperature? If we use continuum observations to adjust T_{min}, we do not find the same value as when T_{min} is

Table 2.3. Determination of the temperature minimum of the homogenous solar atmosphere

Diagnostic used	Source function	Frequency redistribution v	T_{min} (K)	Authors
IR continuum (100–200 µm)	LTE	—	$4\,200 \pm 300$	Avrett (1977)
UV continuum (1 400–2 000) Å	LTE	—	$4\,330 \pm\ \ 70$	Samain (1980)
	NLTE	—	$4\,150 \pm 160$	—
H and K lines	NLTE	complete	$4\,250 \pm\ \ 50$	Dumont (1967)
of Ca II	NLTE	partial	$4\,450 \pm 130$	Ayres and
(λ 3 933–3 968 Å)				Linsky (1976)

N.B. The problem of redistribution of radiation within spectral lines arises once the levels are broadened by radiation and collisions, both elastic and inelastic. The source function then contains a variable proportion of coherent and incoherent emission (see Omont et al. 1972).

derived from resonance lines such as the H and K lines of Ca II. In reality T_{min} depends on line-formation theory in the second case, and calculations of the continuum source function in a suitable approximation in the first case. Table 2.3 shows the influence of these approximations.

Is the solar minimum temperature necessarily unique? The above analysis applies to *homogeneous* models, and Table 2.3 shows the difficulty of identifying a good theory. But the Sun has a surface roughness because of the horizontal fine structure, which can lead to different values of T_{min} being obtained when different frequencies and inclination angles are used.

Solar Chromosphere. As we shall see in Sect. 2.4, it is only possible to approximate the solar chromosphere as spherically symmetric over a thickness of about 2000 km, since structures (spicules) emerging from it reach heights of about 10 000 km while retaining a relatively low temperature (10^4 K). We provisionally retain the assumption of plane-parallel geometry for chromosphere models.

Semi-empirical chromosphere models are calculated by two methods.

(a) *Trial and error.* We proceed as described above for the upper photosphere. But for the chromosphere we have to give a priori not only $T_e(z)$ but also the field of unresolved velocities $V(z)$. This function is needed (1) to calculate the Doppler width of the lines, which are observed at high accuracy; (2) to calculate the turbulent pressure, which affects the scaleheight of the chromosphere, measurable in eclipses (see below). These models use *disc* observations of optically thick lines and continua, all formed out of LTE: the calculated lines and continua have to agree with observation over all available frequencies and positions on the disc. Figure 2.4 shows the temperature distribution in such a model, giving the formation depths of spectral characteristics formed in the chromosphere. Figure 2.5 gives the
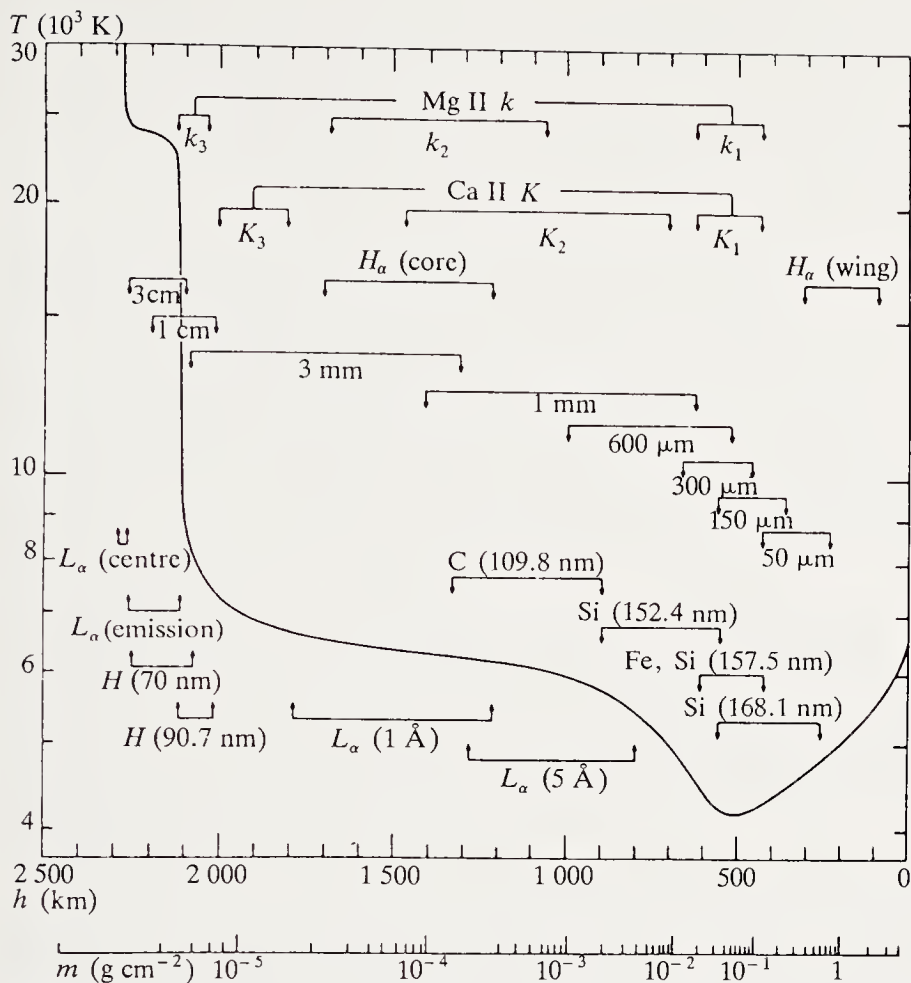
Fig. 2.4. Model of the solar chromosphere: temperature distribution. h is the height above the point defined by $\tau_{5000} = 1$, m is the column density of matter. (After J.E. Vernazza et al., *Astrophys. J.* **45**, 635 (1981). Reproduced with the kind permission of *The Astrophysical Journal* published by The University of Chicago Press; © 1981 The American Astronomical Society)

density distribution as a function of height h in the atmosphere for the same model.

We note that the typical electron temperature in the chromosphere is about 6000 K, forming a first plateau of around 1000 km near this temperature. But the core of the Lyα line is formed at 2.4×10^4 K, where there is a second narrow temperature plateau. The Lyman continuum comes from the region above the mean chromosphere, where T_e reaches 10^4 K. The matter density ρ varies as N_H, and decreases outwards by a factor 10^4 to 10^5 over the chromosphere, but N_e is remarkably constant over a region coincident with the first temperature plateau because of the increase in the number of electrons through the ionisation of hydrogen as the density drops.

(b) *Eclipse-observation method.* This gave the first solar-chromosphere models in the 1950s.
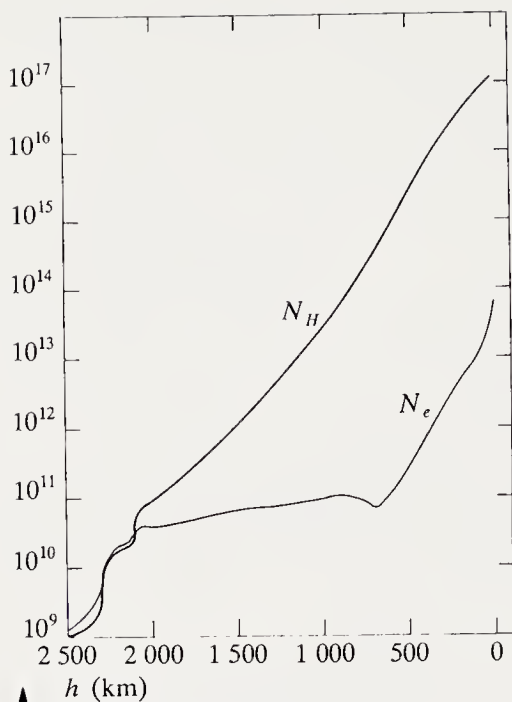
Fig. 2.5. Solar-chromosphere model: hydrogen and electron number densities N_H, N_e (cm^{-3}) as functions of height h. (After Vernazza et al. 1981)
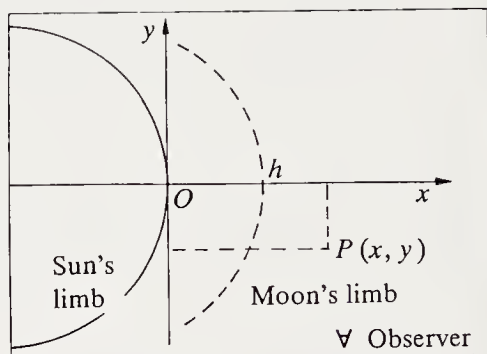
Fig. 2.6. Eclipse observations. The Moon's limb (*dashed*) moves across the Sun and blocks out its bright disc. The chromosphere is then directly observable at field points $P(x, y)$

The method is almost entirely empirical and specific to the chromospheric layers. We observe the light emitted by that volume of the chromosphere at a height larger than h at the limb, defined by $\tau_l(\lambda = 5000 \text{ Å}) = 1$. The corresponding energy integrated over all wavelengths is (see Fig. 2.6)

$$E(h) = \int_h^\infty I(x)\mathrm{d}x \quad \text{erg s}^{-1} \text{ str}^{-1} \text{ cm}^{-1} \tag{2.31}$$

with

$$I(x) = \int_0^{+\infty} \int_0^\infty \varepsilon_\lambda(x, y)\mathrm{e}^{-\tau_{l,\lambda}}\mathrm{d}y\,\mathrm{d}\lambda, \tag{2.32}$$

where $\varepsilon_\lambda(x, y)$ is the monochromatic emissivity per unit volume at the point P, and $\tau_{l,\lambda}$ is the tangential optical depth defined by

$$\tau_{l,\lambda} = \int_{-\infty}^{+\infty} n_i \alpha_\lambda \mathrm{d}y, \tag{2.33}$$

where n_i and α_λ have the same meanings as in (2.11).

In spherical symmetry we have $\varepsilon_{\lambda(x,y)} = \varepsilon_\lambda(h)$. In the following we shall use λ rather than ν.

The method has the following advantages: we observe at different λ (say p wavelengths) and at different values of h. We write analytic expressions

for ε_λ as a function of the unknowns (T_e, densities, say k unknowns). In the continuum, the emissivity per wavelength interval is the sum of all the radiative recombination emissivities

$$\varepsilon_\lambda = 4\pi B_\lambda(T_e) \sum_i n_i^* \alpha_{\lambda,i} [1 - \exp(-h\nu/kT_e)], \tag{2.34}$$

where the symbol $*$ denotes an LTE quantity. A similar expression gives the net line emissivity.

At each height h we thus have p equations between k unknowns. By judicious choice of wavelengths we can obtain all the unknowns.

The difficulties with this method arise in extracting $\varepsilon_\lambda(h)$ from $E(h)$. This requires (1) a profile of the solar limb, to fix h; (2) good observational resolution in altitude; (3) absolute photometric calibration of E; (4) an accurate numerical method for extracting the second derivative of the data $E(\lambda)$.

The method was applied to observations of the eclipse of 1952 at two wavelengths in the visible continuum ($\lambda = 3646, 4700$ Å) from 500 to 2400 km above the limb. This gave the first realistic model of the solar chromosphere, supplying the electron density and temperature as well as the ionisation degree of hydrogen as a function of h. We refer to Thomas and Athay (1961) for details.

Transition Zone.   Semi-empirical models of the transition zone use diagnostic methods valid when the lines are effectively but not optically thin. A line is optically thin at frequency ν if $\tau_\nu = -\int_0^\infty \kappa_\nu \rho dz$ is less than 1. The line is *effectively* thin if at frequency ν the mean free path, $mfp = 1/\kappa_\lambda\rho$, is shorter than the thermalisation length of the photons, defined as the depth beyond which the source function equals $B_\nu(T_e)$. Pottasch (1964) gave the first analysis of the transition region and lower coronal emission lines.

(a) *Emission measure as a function of* T_e. For an optically thin line the emergent intensity on a line of sight defined by $\mu = \cos\theta$ is

$$I_\nu(\tau = 0, \mu) = \int_0^\infty S_\nu e^{-t_\nu} dt_\nu \cong \int_0^\infty S_\nu dt_\nu, \tag{2.35}$$

where t_ν is the optical depth along the line of sight ($= \tau_\nu/\mu$).

We consider an atom with just two levels i, j, with i the lower. Then adopting the hypothesis of complete redistribution of radiation (not discussed here: see Omont et al. 1972), S_ν is

$$S_\nu = \frac{n_j A_{ji}}{n_i B_{ij} - n_j B_{ji}}, \tag{2.36}$$

where A_{ij}, B_{ij}, B_{ji} are the Einstein coefficients. Further,

$$\mathrm{d}t_\nu = \frac{h\nu_{ij}}{4\pi}(n_i B_{ij} - n_j B_{ji})\varphi_\nu\,\mathrm{d}s\,, \tag{2.37}$$

where φ_ν is the absorption-line profile (with $\int_0^\infty \varphi_\nu \mathrm{d}\nu = 1$) and $\mathrm{d}s$ is the path element along the line of sight. Hence the emergent line intensity integrated over ν is

$$I(0) = \frac{h\nu_{ij}}{4\pi}\int_0^\infty n_j A_{ji}\mathrm{d}s\,. \tag{2.38}$$

In the transition zone the excitation of the upper level j is mainly collisional. For a two-level atom we thus have $n_j A_{ji} = n_i C_{ij}$, C_{ij} being the collisional-excitation-rate coeffcient for level j from level i.

It is convenient to write n_i in the form

$$n_i = \frac{n_i}{N_k}\frac{N_k}{N_A}\frac{N_A}{N_H}\frac{N_H}{N_e}\,N_e\,, \tag{2.39}$$

where N_k is the number density of line-producing ions, N_A the number density of nuclei of the element:

$$N_A = \sum_k N_k\,, \tag{2.40}$$

and N_H the hydrogen number density.

The rate C_{ij} $(\mathrm{cm}^{-3}\ \mathrm{s}^{-1})$ is a function of excitation energy E_{ij} and T_e

$$C_{ij} = 8.63\times 10^{-6}\frac{\Omega_{ij}}{g_i}T_e^{-1/2}N_e\exp(-E_{ij}/kT)\,; \tag{2.41}$$

Ω_{ij} is an atomic quantity called the *collision strength*, and g_i is the statistical weight of level i.

A completely ionised plasma with $N_{He}/N_H = 0.10$ has $N_H/N_e \simeq 0.8$. Transition-region lines are often resonance lines, $i = 1$ and $n_i/N_k \simeq 1$. N_A/N_H is the abundance A by number of the element forming the line.

The integrated line-of-sight intensity $(\mathrm{erg}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{str}^{-1})$ of an emission line is

$$I(0) = h\nu_{ij}\frac{\Omega_{ij}}{g_i}A\int_{\Delta s} N_e^2 G(T_e)\mathrm{d}s\,, \tag{2.42}$$

where

$$G(T_e) = T_e^{-1/2}N_k/N_A\exp(-E_{12}/kT_e) \tag{2.43}$$

depends only on T_e and is calculated from the ionisation equations. This function is strongly peaked at a particular value T_m of T_e; if $\langle G(T_e)\rangle$ is an average of $G(T_e)$ we have

60

$$\frac{N(\mathrm{Si})}{N(\mathrm{H})}\int N_e^2\,\mathrm{d}h$$

10^{23}

├── S II ──┤

├── SiX II ──┤
TeXV
├─ AlXI ─┤
├── MgX ──┤ TeXV ─┤

├── C II ──┤
├─ NII ─┤
├── Si III ──┤

├── SVII ─┤ AlX ─┤
├── NeVII ──┤ AlIX ─┤
├── Si X ──┤

├── O II ──┤
├── SIII ──┤

├── MgIX ──┤

├── N III ──┤
├── Si IV ──┤
├── C II ── C IV ──┤

├── Ne VII ──┤

10^{22}

├── MgVIII ──┤
├── Si VIII ──┤

├── N IV ──┤
├── SVI ── O VI ──┤
├── S V ──┤ ├── MgVI ──┤
├── S IV ──┤ ├── SiVI MgVII ──┤
├── O V ──┤
├── NeIV ─┤ ├── N V ──┤
├── O IV ── Ne V ──┤
├── O III ──┤ ├── Ne VI ──┤

10^{21}

10 000   30 000   100 000   300 000   1 000 000 T (K)
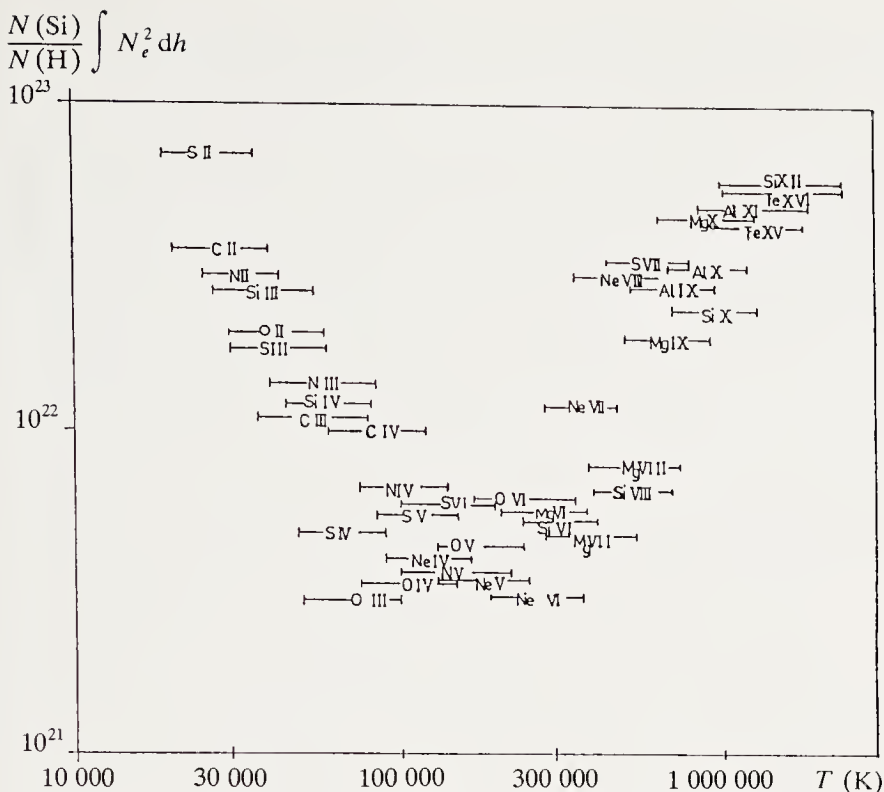
Fig. 2.7. Emission measure for the quiet Sun. The ordinate is the emission measure multiplied by the abundance of silicon, and the abscissa the maximum temperature at which the ions are formed. (From S.R. Pottasch, *Space Science Reviews* **3**, 816 (1964). Reproduced with the kind permission of Kluwer Academic Publishers)

$$I(0) = h\nu_{ij}\frac{\Omega_{ij}}{g_i}A\langle G(T_e)\rangle \int_{\Delta s} N_e^2\,\mathrm{d}s. \tag{2.44}$$

$I(0)$ is measured, and the other quantities $(\nu_{ij}, \Omega_{ij}, g_i, A)$ are known; we thus get the *emission measure*, often denoted EM:

$$EM = \int_{\Delta s} N_e^2\,\mathrm{d}s. \tag{2.45}$$

If we choose $\langle G(T_e)\rangle = G(T_m \pm 0.15\ \mathrm{dex})$, which is a very good approximation, we get $EM = \psi(T_m)$. (dex = "exponent of 10").

Studying lines formed at different temperatures gives the emission measure as a function of T_m. Figure 2.7 gives the curve $EM(T_m)$ found by Pottasch for the quiet Sun.

The emission measure is an integrated quantity for which we do not require an a priori assumption about the geometry of the medium. In general Δs is chosen as the geometrical thickness over which $\Delta \log T_e = 0.30$, in agreement with the choice of $\langle G(T_e)\rangle$ made above. In the solar case, Δs is small (about 100 km) so the plane-parallel assumption allows one to construct semi-empirical models of the transition region. In reality the integral

(2.42) should be over a volume rather than a length since the observations have limited spatial resolution. In practice all models assume infinite resolution.

(b) *Transition-region models.* The full thickness of this region is less than the pressure scaleheight $H_P = kT_e/\mu g$, where μ is the mean molecular mass; the gas pressure thus varies little here. If we assume no motion, the equation of hydrostatic equilibrium applies. Actually short-lived jets of matter with velocities up to 200 to 400 km s^{-1} have been seen in the CIV lines (Brueckner 1981). Their Doppler shifts are generally to shorter wavelengths, so this is matter moving away from the Sun. Flows directed towards the solar surface (*downflows*) are common elsewhere in the lower transition zone and chromosphere.

In the following we consider observations made at the centre of the disc: we thus replace s by z. We also neglect matter motions to a first approximation.

If we assume that the temperature gradient dT_e/dz is meaningful and constant over the emission region Δz, we can write

$$
EM = \int_{\Delta z} N_e^2 \, dz = \int_{2\Delta T} N_e^2 \frac{1}{dT_e/dz} \, dT_e
$$

$$
= \left(\frac{dT_e}{dz}\right)^{-1} \int_{2\Delta T} (T_e N_e^2) \, d \ln T_e \tag{2.46}
$$

with $\Delta T = 0.15$ dex. We assume also, as justified above, that over the width Δz of the emission region of each line we have $P_e = N_e k T_e = \overline{P}_e \simeq$ const. Then

$$
EM = 0.3 \left(\frac{dT_e}{dz}\right)^{-1} 2.3 \frac{\overline{P}_e^2}{k^2 T_e}, \tag{2.47}
$$

giving

$$
\frac{dT_e}{dz} = 3.6 \times 10^{31} \overline{P}_e^2 / (T_e \times EM). \tag{2.48}
$$

The hydrostatic-equilibrium equation then gives the dependence of P_e on z over the transition zone. With

$$
P_g = NkT_e = (N_H + N_{He} + N_e)kT_e = (1.1 N_H + N_e)kT_e
$$

and $N_H/N_e = 0.8$ we get

$$
\frac{dP_e}{dz} = -7.1 \times 10^{-9} \overline{P}_e g / T. \tag{2.49}
$$

We therefore have two equations, (2.48) and (2.49), which we can integrate, given the function $EM(T_m)$, which is a single-valued function of T_e. To find a solution we need the density or pressure at a reference height z_0.

We can choose $P_e(z_0)$ at the top of the chromosphere, or take the value given by various diagnostics: there are pairs of lines whose ratio is sensitive to N_e, e.g. the 1176 and 977 Å lines of C III. These two lines respectively involve a metastable level[2] ($2p^2\ ^3P$) and the ground level ($2s^2\ ^1S_0$) of C III. The intensity ratio of the triplet and resonance transitions depends on collisional and radiative rates between levels. Radiative excitations, ionisations, and recombinations are negligible in the transition zone. For a given T_e the intensity ratio is sensitive to N_e when the collisional and spontaneous deexcitation rates are of the same order.

Values $dT_e/dz = 1000$ K km^{-1}, $P_g = 0.1$ dyne cm^{-2} are typical of the transition region of the quiet Sun.

Transition-region emission-line analysis can be applied to other stars. We have to replace the emergent intensity $I(0)$ by $F(0)$, the integrated emergent flux over the line ($F(0) = 2\pi I(0)$). Averaging gives a stronger smoothing of the inhomogeneities in the lateral structure than for the Sun.

The Corona. The idea of a mean corona is even less meaningful than that of a mean chromosphere or transition region, since observations from the space platform Skylab (1973) showed pronounced structure of the corona to be a general phenomenon. There are also marked daily and solar-cycle variations (see Chap. 7).

To fix orders of magnitude Fig. 2.8 gives T_e and N_e and the wind velocity as functions of distance from the photosphere for the quiet corona, in a polar and an equatorial coronal hole. Coronal holes are described in Sect. 2.5.3.

The main diagnostics for the corona are (1) for the lower corona ($r < 0.5R_\odot$ above the visible limb) the same as those used in the transition region; (2) for the outer corona, the emission-line profiles (T_e diagnostic) and the Thomson-scattered photospheric continuum (N_e diagnostic).

In the outer corona we distinguish from the point of view of radiation the K corona (continuum radiation produced by electron scattering) and the F corona (continuum radiation from scattering by interplanetary grains). The F corona becomes brighter than the K corona, which nevertheless exists, above $r = 1.2R_\odot$ from the limb.

(a) *Electron-temperature diagnostics.* The observed emission lines in the corona result either from collisional excitation, as in the transition zone, or from radiative excitation involving *resonant scattering*. We observe on the one hand lines from highly ionised species such as Mg X, Ne VII, Ne VIII, and on the other hand the Ly α lines of H I and He II, and the resonance lines of N V, O VI, etc. The former are excited by electrons from the lower corona, the latter by photons from the chromosphere or the transition zone. In the latter case the intensity of the scattered radiation depends on the temperature of the scattering ions or atoms through their velocity distribution function.

---

[2] A metastable atomic level is one not linked to the ground level by a permitted transition.
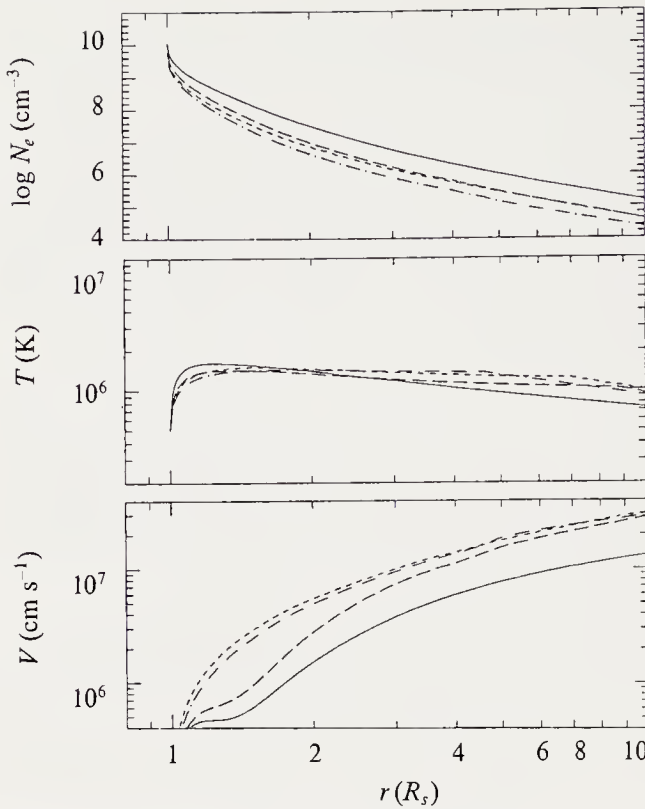
Fig. 2.8. Distribution of electron density and temperature and wind velocity as functions of distance from the coronal base defined as $r = R_s$. *Full curves* refer to a quiet region, *long dashes* to a polar coronal hole at solar maximum, and *short dashes* to an equatorial hole; *long–short dashes* show a polar hole at solar minimum. (From G.L. Withbroe, *Astrophys. J.* **325**, 442 (1988), Fig. 13, p. 457. Reproduced with the kind permission of the *Astrophysical Journal*, published by The University of Chicago Press; © 1988 The American Astronomical Society)

Consider a coronal line $(i - j)$, of rest frequency ν_0. Level j is excited by a beam of radiation of frequency $\nu', \nu' + d\nu'$, direction $\boldsymbol{n}'$, in solid angle $d\omega$. The number k of atoms or ions excited in the velocity interval $\boldsymbol{v}, \boldsymbol{v} + d\boldsymbol{v}$ is then

$$dn_j(\boldsymbol{v}) = n_i(\boldsymbol{v}) \frac{B_{ij} h \nu'}{4\pi} I(\nu', \omega) \delta \left(\nu' - \nu_0 - \frac{\nu_0}{c} \boldsymbol{v} \cdot \boldsymbol{n}' \right) d\omega \, d\nu' \, d\boldsymbol{v} , \quad (2.50)$$

where B_{ij} is the Einstein coefficient for radiative excitation, and h Planck's constant. The Dirac δ function appears as a factor, since an atom of velocity v can only scatter frequencies $\nu' = \nu_0 + (\nu_0/c)\boldsymbol{v} \cdot \boldsymbol{n}'$. The natural width of the line is ignored in (2.50). The number of photons scattered into the observer's direction $\boldsymbol{n}$ is then

$$dn \left(\nu = \nu_0 + \frac{\nu_0}{c} \boldsymbol{v} \cdot \boldsymbol{n} \right) = dn_j(\boldsymbol{v})(a + b(\boldsymbol{n} \cdot \boldsymbol{n}')^2) , \quad (2.51)$$

where $a + b(\boldsymbol{n} \cdot \boldsymbol{n}')^2$ is the angular dependence of the scattering process. For Lyα scattering, $a = 11/12, b = 3/12$.

From (2.39) we can write the fraction of atoms in level i and the velocity interval $\boldsymbol{v}, \boldsymbol{v} + \mathrm{d}\boldsymbol{v}$ as

$$n_i(\boldsymbol{v})\mathrm{d}\boldsymbol{v} \simeq 0.8 A N_e (N_k/N_A) f(\boldsymbol{v})\mathrm{d}\boldsymbol{v}\,, \tag{2.52}$$

where A is the elemental abundance by number ($A = 1$ for hydrogen), $N_k/N_A = F(T_i)$ is the degree of ionisation of the ion k, and T_i is the ionisation temperature, assumed equal to T_e, and we recall that $N_H/N_e = 0.8$ in the corona if $N_{He}/N_H = 0.1$. We have assumed that i is the ground level and $n_i/N_k \simeq 1$. The intensity scattered into the line of sight, taken as the x-axis is

$$
\begin{aligned}
I(\nu) = \frac{0.8 h B_{ij}}{4\pi} &\int_0^{+\infty} N_e F(T_e) f\left(v_x = (\nu - \nu_0)\frac{c}{\nu_0}\right)\mathrm{d}x \\
&\times \int_\omega (a + b(\boldsymbol{n}\cdot\boldsymbol{n}')^2)\,\mathrm{d}\omega \int_0^{+\infty} I(\nu',\omega)\mathrm{d}\nu' \\
&\times \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(v_y, v_z)\,\delta\left(\nu' - \nu_0 - \frac{\nu_0}{c}\boldsymbol{v}\cdot\boldsymbol{n}'\right)\mathrm{d}v_y\,\mathrm{d}v_z\,. \quad (2.53)
\end{aligned}
$$

If $f(v)$ is assumed Maxwellian (at temperature T), part of expression (2.53) may be integrated analytically. Analysis of the Lyα line gives T as the temperature characterising the thermal motions of neutral hydrogen, which at 125 km s^{-1} exceed the non-thermal motions (wind $\leq$ 20 km s^{-1}, turbulence $\leq$ 30 km s^{-1}) in the corona. The temperature T is quite close to the proton temperature T_p; T_p is equal to T_e until about $2R_\odot$. Above this height T_e and T_p differ, and the corona must be treated as two fluids, electrons and protons.

(b) *Electron-density diagnostics.* The white light of the corona (K corona) is caused by coronal electron scattering of the photospheric continuum into the line of sight. The number of incident photons scattered by coronal electrons with velocities in $\boldsymbol{v}_e, \boldsymbol{v}_e + \mathrm{d}\boldsymbol{v}_e$ and number $N_e(\boldsymbol{v}_e)\mathrm{d}\boldsymbol{v}_e$ is

$$\mathrm{d}n = N_e(\boldsymbol{v}_e)\sigma I(\nu',\omega)\,\mathrm{d}\nu'\,\mathrm{d}\omega\,\mathrm{d}\boldsymbol{v}_e\,, \tag{2.54}$$

where $I(\nu',\omega)$ is the intensity of the incident beam of direction $\boldsymbol{n}'$ in solid angle ω. Of these incident photons, $\mathrm{d}n(\nu)$ are scattered towards the observer (direction $\boldsymbol{n}$),

$$\mathrm{d}n(\nu) = \frac{3}{16\pi}\mathrm{d}n(1 + (\boldsymbol{n}\cdot\boldsymbol{n}')^2)\,\delta\left[\left(\nu' - \frac{\nu'}{c}\boldsymbol{v}_e\cdot\boldsymbol{n}'\right) - \left(\nu - \frac{\nu}{c}\boldsymbol{v}_e\cdot\boldsymbol{n}\right)\right]\,, \tag{2.55}$$

where σ is the Thomson scattering cross-section and $\frac{3}{16}(1 + (\boldsymbol{n}\cdot\boldsymbol{n}')^2)$ its angular dependence. The observed intensity is

$$I_e(\nu) = \frac{3\sigma}{16\pi} \int_0^\infty N_e \mathrm{d}x \int_\omega \mathrm{d}\omega (1 + (\boldsymbol{n} \cdot \boldsymbol{n}')^2) \int_0^\infty I(\nu', \omega) \mathrm{d}\nu'$$

$$\times \int_v f(\boldsymbol{v}_e) \delta\left[\left(\nu' - \frac{\nu'}{c} \boldsymbol{v}_e \cdot \boldsymbol{n}'\right) - \left(\nu - \frac{\nu}{c} \boldsymbol{v}_e \cdot \boldsymbol{n}\right)\right] \mathrm{d}\boldsymbol{v}_e . \qquad (2.56)$$

If we again assume that $f(v_e)$ is Maxwellian, the $\boldsymbol{v}_e$ integral can be calculated analytically. The observed intensity then gives $\int_0^\infty N_e \mathrm{d}x$.

(c) *Limits of the diagnostics.* Coronal densities are low (Fig. 2.8). This means that the velocity distributions will deviate from Maxwellian. Also there are mass motions in the corona (solar wind and turbulence associated with waves); these too modify the distribution functions. Finally estimates of $\beta = P_g/P_{\mathrm{mag}}$, where P_g is the gas pressure and P_{mag} the magnetic pressure, show that β is much smaller than 1 in the corona, i.e. the magnetic field structures the corona. At best we can apply the above diagnostics to individual structures, showing the importance of high spatial resolution in coronal lines and continuum so that we can take account of the size and geometry in each case.

Analysis of radio emission also gives a diagnostic of the mean corona. We refer for example to McLean and Labrum (1985) for this.

2.2.3 Radiative Losses

One of the aims of semi-empirical modelling of the outer layers of the Sun and stars is the estimate of radiative losses. This is the term in the energy equation equal to the divergence of the radiative flux. Its local expression is given by (2.10). We often describe the integrated quantity $\int \frac{\mathrm{d}F_R}{\mathrm{d}z} \mathrm{d}z$, (incorrectly) as the radiative losses, whereas it has the dimensions of a flux. Clearly we need the local expression, i.e. the net radiative cooling rate, to characterise the various terms in the energy equation.

In (2.10), $\mathrm{d}F_R/\mathrm{d}z$ is expressed as a balance between emissions and absorptions. It is positive for net radiative cooling. A continuum tends to cool the medium if $J_\nu < B_\nu$.

The line contribution to $\mathrm{d}F_R/\mathrm{d}z$ introduces the factor, often called NRB (*net radiative bracket*),

$$(\mathrm{NRB}) = 1 - \frac{\int_0^\infty J_\nu \varphi_\nu \mathrm{d}\nu}{S_{ij}} . \qquad (2.57)$$

This term contains the effects of radiative transfer in the lines. In the case of strict LTE, $J_\nu = S_\nu = B_\nu$ and NRB = 0. In the case of restricted LTE, where the source function S_ν is not equal to the mean radiation intensity J_ν, but $S_\nu = B_\nu$, a line generally cools the medium if $J_\nu < B_\nu$ at a given depth. Out of LTE, for a two-level atom, with complete redistribution of radiation,

$$S_{ij} = \frac{\int_0^\infty J_\nu \varphi_\nu \mathrm{d}\nu + \varepsilon B_\nu(T)}{1 + \varepsilon} \,. \tag{2.58}$$

This expression comes from (2.36) when we use the equation of statistical equilibrium. Here ε is the collisional-deexcitation probability per scattering:

$$\varepsilon = (C_{ji}/A_{ji}) \left[1 - \exp(-h\nu_{ij}/kT_e) \right] \,. \tag{2.59}$$

In this case the NRB term reduces to

$$(\mathrm{NRB}) = \varepsilon (B_\nu(T_e)/S_{ij} - 1) \,, \tag{2.60}$$

which shows that, all things being equal, a line cools more for larger ε. Athay (1976) shows that in the optically thick case, radiative losses in the lines measure the number of photons created by collisional excitation in one thermalisation length. In the optically thin case the line term in $\mathrm{d}F_R/\mathrm{d}z$ reduces to

$$\sum_{i,j>i} h\nu_{ij} A_{ji} n_j \,. \tag{2.61}$$

In the solar chromosphere the lines are optically thick and we thus have to solve the transfer equation for the NRB terms to calculate the radiative losses. In contrast in the transition region and corona the lines are optically thin, considerably simplifying matters.

Table 2.4. Integrated radiative losses (erg cm^{-2} s^{-1})

1. Radiative losses	
– Photosphere	6.4×10^{10}
– Chromosphere	$2\text{–}6 \times 10^6$
Balmer series	5×10^5
H$^-$	4×10^5
Ly α	3×10^5
metal lines (Mg II, Ca II)	34×10^5
– Transition region	$4\text{–}6 \times 10^5$
– Corona: quiet Sun	$\sim 6 \times 10^5$
– Corona: coronal hole	$\sim 10^4$
2. For comparison	
– Conductive flux:	
Quiet Sun	$\sim 2 \times 10^5$
Coronal hole	6×10^4
– Solar wind:	
Quiet Sun	$0?$
Coronal hole	$\sim 6 \times 10^5$

P_{rad} (erg cm^3 s^{-1})
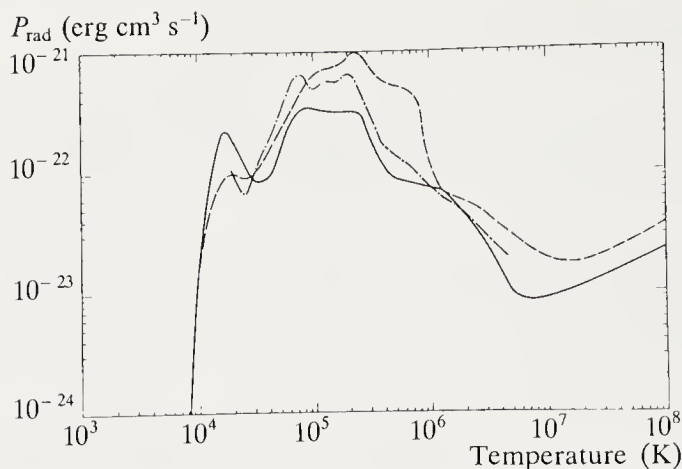
Fig. 2.9. Radiative losses in the solar corona. The ordinate is P_{rad}, defined by $dF_{rad}/dz = N_e^2 P_{rad} = N_e^2 \Lambda(T)$. (From McWhirter et al. 1975)

Table 2.4 gives the integrated radiative losses in the various outer layers of the Sun. It shows the contributions of individual atoms or ions to the radiative energy losses.

When the plasma is optically thin, we have seen that it is not useful to calculate terms of the NRB type to find the radiative losses. In fact a good approximation for dF_R/dz is a single-valued function of temperature, i.e. $dF_R/dz = N_e^2 \Lambda(T)$. Figure 2.9 shows $\Lambda(T)$, valid for the solar corona at heights above that of the formation of Ly α.

2.3 The Chemical Composition of the Solar Atmosphere

The Chemical Composition of the Solar Atmosphere The first paper giving the chemical composition of the solar atmosphere was by H.N. Russell in 1929. The predominance of hydrogen in the solar gas was a major result. Abundances of other elements were naturally given with respect to hydrogen. These abundances were not very different from what is found today, using the photospheric (Fraunhofer) spectrum. This "Russell mixture" is still used to specify the chemical composition of solar and stellar interiors.

The study of the Sun's chemical composition is an archetype for stellar-abundance determinations in general, but also for other objects in the Universe. The solar gas is a sample of cosmic material, and, more accurately, its composition should be that of the interstellar medium 4.6×10^9 y ago. Moreover it was formed from the same protosolar nebula as the contents of the Solar System, and a comparison of abundances in the Sun and objects such as meteorites or minor bodies is rich in information about the history of each of these objects.

More generally, finding abundances is of major importance in astrophysics, for models of the solar and stellar atmospheres and interiors, for checking the chemical homogeneity or inhomogeneity of the stars, for Solar System evolution, for testing theories of nucleosynthesis, and for the chemical evolution of galaxies and the abundance gradients found between their different regions.

2.3.1 Methods of Finding Abundances

There are two main methods of determining the abundance of an element from a photospheric spectrum composed of absorption lines.

The Curve of Growth Method. The curve of growth compares the line equivalent width W to the number of atoms or ions producing the lines. We have

$$W = \int_{-\infty}^{+\infty} \frac{I_c - I_\lambda}{I_c} \mathrm{d}(\Delta\lambda) = \int_{-\infty}^{+\infty} r_\lambda \mathrm{d}(\Delta\lambda), \qquad (2.62)$$

where I_λ is the emergent intensity in the line at the point λ of the profile a distance $\Delta\lambda$ from line centre, and I_c the continuum intensity; r_λ is the depression of the profile. If we consider a star whose disc is unresolved, the specific intensity I is replaced by the mean intensity over the disc, which is numerically equal to the emergent flux.

In the following it is more convenient to use the variable λ rather than ν since UV and visible spectrographs provide I_λ.

We can measure W for each line from the observed spectrum. Formal solution of the transfer equation for the emergent intensities I_c and I_λ in a plane-parallel atmosphere gives an analytic expression for W. We have

$$I_c - I_\lambda = \int_0^\infty S_c(\tau)\mathrm{e}^{-\tau/\mu}\mathrm{d}\tau/\mu - \int_0^\infty S_l(\tau)\mathrm{e}^{-(\tau+\tau_\lambda)/\mu}\mathrm{d}(\tau+\tau_\lambda)/\mu, (2.63)$$

where τ, τ_λ are the optical depths, and S_c, S_l the source functions, in the continuum and line respectively. I_c and I_λ correspond to an angle θ with respect to the normal to the layers ($\mu = \cos\theta$).

In practice one calculates the curve of growth only for lines formed in *pure absorption*, or LTE (i.e. when the source function is $B_\lambda(T_e)$). This is the only case where the source function is the same for all the lines. Generally, the source function S_l has a part due to incoherent scattering (see the expression for the source function of a two-level atom (2.58)); each line is then characterised by its coefficient ε, and it makes sense to speak of the curve of growth only for lines with the same ε (Athay and Skumanich 1968).

In the following we assume $S_c(\tau) = S_l(\tau) = B_\lambda(T_e(\tau))$, i.e. that LTE holds. Then $I_c - I_\lambda$ is easily transformed using the relation $\mathrm{d}\tau_\lambda = \chi_\lambda/\chi\mathrm{d}\tau$ and integrating by parts; for each angle μ

$$I_c - I_\lambda = \int_0^\infty \left(\int_{\tau'}^\infty B_\lambda(\tau) e^{\tau/\mu} d\tau/\mu - B_\lambda e^{-\tau/\mu} \right) e^{-\tau_\lambda/\mu} d\tau_\lambda/\mu \qquad (2.64)$$

and

$$\frac{I_c - I_\lambda}{I_c} = r_\lambda = \int_0^\infty g(\tau, \mu) e^{\tau_\lambda/\mu} d\tau_\lambda , \qquad (2.65)$$

where $g(\tau, \mu)$ is the *weight function*, which depends only on the *model*:

$$g(\tau, \mu) = \frac{1}{\mu} \frac{\int_\tau^\infty B_\lambda(\tau') e^{\tau'/\mu} d\tau'/\mu - B_\lambda e^{-\tau'/\mu}}{\int_0^\infty B_\lambda(\tau) e^{-\tau/\mu} d\tau/\mu} . \qquad (2.66)$$

The term $e^{\tau_\lambda/\mu}$ (2.65) depends only on the *line*. The equivalent width W becomes

$$W = \int_0^\infty g(\tau, \mu) \left(\int_{-\infty}^{+\infty} e^{-\tau_\lambda/\mu} \frac{\chi_\lambda}{\chi} d(\Delta\lambda) \right) d\tau . \qquad (2.67)$$

We have seen from (1.32) that the linear absorption coefficient in the line, χ_λ, can at each depth τ be written

$$\chi_\lambda = \chi_0 \varphi_\lambda , \qquad (2.68)$$

where φ_λ is the broadening profile of the line. We assume that the profile is given by the Voigt function $H(a, v)$, in which a is the damping constant $a = \gamma/4\pi\Delta\nu_D$; v is the general point of the profile, expressed as $v = \Delta\nu/\Delta\nu_D = \Delta\lambda/\Delta\lambda_D$; γ is the full width at half-maximum of the line profile; $\Delta\lambda_D$ and $\Delta\nu_D$ denote the Doppler width in wavelength and frequency; $\Delta\lambda = \lambda - \lambda_0$, where λ_0 is the central wavelength of the line. We set

$$x = \frac{1}{\mu} \int_0^\tau \frac{\chi_0}{\chi} d\tau ,$$

so that

$$\int_{-\infty}^{+\infty} e^{-\tau_\lambda/\mu} \frac{\chi_\lambda}{\chi} d(\Delta\lambda) = \int_{-\infty}^{+\infty} e^{-\tau_\lambda/\mu} \frac{\chi_0}{\chi} H(a, v) d(\Delta\lambda)$$

$$= \frac{\chi_0}{\chi} \frac{d(\Delta\lambda)}{dv} \int_{-\infty}^{+\infty} H(a, v) e^{-x H(a,v)} dv . \qquad (2.69)$$

Pecker defines a *saturation function*

$$\phi(a, x) = \frac{2}{\pi^{1/2}} \int_0^\infty H e^{-xH} dv ,$$

where H stands for $H(a, v)$. We get

$$W = \int_0^\infty g(\tau, \mu) \frac{\chi_0}{\chi} \sqrt{\pi} \, \Delta\lambda_D \phi(a, x) d\tau . \qquad (2.70)$$

The absorption coefficient at line centre is

$$\chi_0 = \frac{\sqrt{\pi}\, e^2}{mc} \frac{\lambda_0^2}{\Delta\lambda_D} n_i(\tau) f \,, \tag{2.71}$$

where n_i is the population of the lower level of the line ij, and f its oscillator strength. Let A be the abundance of the element forming the line, so that $A = N_A/N_H$. Then we have finally

$$W = \frac{\pi e^2}{mc} \lambda_0^2 f A \int_0^\infty g(\tau,\mu) \frac{n_i(\tau)}{\chi(\tau)N_A(\tau)} N_H(\tau)\phi(a,x)\mathrm{d}\tau \,. \tag{2.72}$$

The quantities characterising the line are $\lambda_0, f, n_i/N_A$ and a; the abundance A by number depends on the element, and the other terms in (2.72) are functions of the model only.

— For a weak line, $\chi_0/\chi \ll 1$, a is small, so

$$\phi(a,x) \simeq \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v^2}\mathrm{d}v = 1 \,.$$

In LTE

$$\frac{n_i}{N_k} = \frac{g_i}{U(T)} \exp(-h\nu_{ij}/kT_e)$$

from the Boltzmann law. $U(T)$ is the partition function and N_k the population of the ion producing the line; then if $F(\tau)$ stands for the ionisation degree, $F(\tau) = N_k(\tau)/N_A(\tau)$, and

$$\frac{n_i}{N_A} = \frac{n_i}{N_k} \frac{N_k}{N_A} = g_i \exp(-h\nu_{ij}/kT_e)F(\tau)/U(T) \,. \tag{2.73}$$

The equivalent width of a *weak* line ij is thus

$$W_{ij} = \frac{\pi e^2}{mc} \lambda_0^2 f g_i A \int_0^\infty g(\tau,\mu)\exp(-h\nu_{ij}/kT_e)F(\tau)N_H(\tau)\mathrm{d}\tau \,.$$

In this case W/λ_0 is a *linear* function of $A\lambda_0 f g_i$, and the integral can be calculated using a model atmosphere.

— For a strong line,

$$H(a,v) = \frac{1}{\pi^{1/2}} \frac{a}{v^2} \quad \text{and} \quad \phi(a,x) = \frac{1}{2}(a\pi^{1/2}/x)^{1/2} \,.$$

From (2.71) and the definition of x, x varies as $A\lambda_0^2 f g_i$, so that $W/\lambda_0 \propto (A\lambda_0 g_i f a)^{1/2}$ for a strong line.

The theoretical curve of growth gives $\log W/\lambda$ as a function of $\log(A\lambda_0 g_i f) + \Gamma$, where Γ depends only on the model for *lines of the same multiplet*. It has a linear part of slope 1 for "weak lines", independent of the damping constant a, while the "strong line" part has slope 1/2 depending

on the damping constant as $a^{1/2}$. Between these parts there is a *plateau* describing the abundance interval over which the line begins to saturate, i.e. have profile wings $\varphi_\lambda = H(a,v)$. The position of this plateau depends on a. The damping constant a is not constant with depth τ as both the width γ and the Doppler width $\Delta\nu_D$ depend on τ. We note further that $\Delta\nu_D$ describes the Doppler broadening by all micromotions (thermal and turbulent). Thus in practice we can use the position of the plateau in the curve of growth to find the rms microturbulent velocity in the atmosphere, ξ_t. Through the weight function $g(\tau, \mu)$, the position of the plateau also depends on the gradient of the source function with depth (we can show this for a dependence $B = B_0(1 + \beta\tau)$). All things being equal the plateau position also depends on ε, for non-LTE lines.

To use a curve of growth we plot the empirical curve $\log(W/\lambda_0)$ as a function of $\log(\lambda_0 g_i f)$ using the measured equivalent widths and the known quantities λ_0, g_i, f for each line. We superimpose the theoretical curve calculated from a model atmosphere containing a microturbulence parameter ξ_t, which accounts for the observed line widths without being theoretically justified. Comparing empirical and theoretical curves of growth, a horizontal translation gives A and a vertical one a and hence ξ_t if we know the width γ. Figure 2.10 shows an empirical solar curve of growth for Fe II.

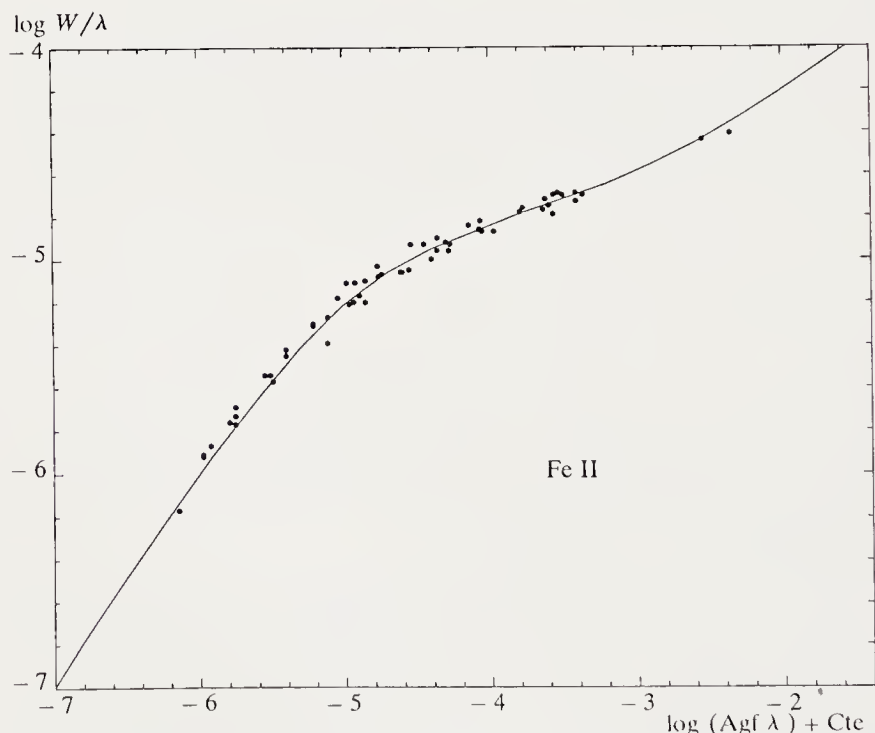**Fig. 2.10.** Solar curve of growth for photospheric lines of Fe II measured at the centre of the disc. The abscissa is $\log(Agf\lambda)+$ const. The full curve is the theoretical curve of growth calculated from the model of Holweger and Muller (1974); the microturbulence parameter is $\xi_t = 0.5$ km s^{-1}. The iron abundance is 7.5 on the scale of Table 2.5. (M. Spite, private communication)

A few remarks are in order. (1) LTE curves of growth give LTE abundances, and this assumption must always be justified; (2) it is not legitimate to plot a curve of growth using lines from too large a spectral interval, since the above procedure uses *one* theoretical curve to adjust one empirical curve; the former is calculated for *one* wavelength, the continuous opacity varying with λ; (3) it is not legitimate to plot a single curve of growth for lines belonging to different ions; (4) in practice we no longer calculate a saturation or a weight function, since computers give W as an integral over λ rather than over τ using the emergent intensities (cf. (2.63)); (5) the main difference between the solar or stellar case and a similar analysis for the interstellar medium is obviously the existence of intrinsic emission in the solar or stellar absorption lines.

The curve of growth has the advantage of giving a value for A *consistent* with all the lines used. If very good spectra are available we can confine ourselves to using the "weak lines" part of the curve to get A. Nowadays it is possible to measure equivalent widths less than 1 mÅ quite accurately.

Direct Use of Line Profiles.   Since high-resolution spectra of good signal-to-noise ($S/N > 100$) are now available and atmosphere and line-formation models are now physically refined enough, it is more satisfactory to provide abundance estimates by adjustment of calculated and observed line profiles. Use of this method and of NLTE calculations of the source functions requires all the atomic data influencing line formation, not simply the oscillator strength and natural width γ. The problem of velocity broadening becomes more severe: in particular the deduced abundance depends not only on ξ_t but also the assumed projected rotation velocity of the star. It also depends on assumptions about the redistribution of the radiation, i.e. the frequency dependence of the line source function over the profile. This effect is ignored in LTE and in NLTE calculations using *complete redistribution.* An advantage of using line profiles directly is that abundance information can be extracted even from line blends.

2.3.2 Results for the Solar Photosphere

Derived solar abundances generally use the curve-of-growth method if the element spectrum has enough lines. If not, individual lines are calculated. There have been efforts to get NLTE abundances for some elements (e.g. OI). Forbidden lines require NLTE analysis.

Table 2.5 gives the photospheric abundances. For comparison we give also abundances obtained from analysis of meteorites.

2.3.3 Further Remarks

Abundances obviously depend on the quality of the atmospheric models used. They further depend on basic atomic data (f and γ), and opacities,

Table 2.5. Abundance of elements in the solar photosphere and meteorites (from N. Grevesse, *Physica Scripta* **8,** 49, 1984)

Z	Element	Solar abundance[a]	Meteoritic abundance[b]
1	H	12.00	12.00
2	He	(11.00)	—
3	Li	1.00	3.33 ± 0.01
4	Be	1.15	1.45 ± 0.03
5	B	(2.6)	2.93 ± 0.11
6	C	8.69	—
7	N	7.99	—
8	O	8.91 ± 0.02	—
9	F	(4.56)	4.48 ± 0.06
10	Ne	(8.0)	—
11	Na	6.33 ± 0.03	6.31 ± 0.03
12	Mg	7.58 ± 0.05	7.59 ± 0.02
13	Al	6.47	6.48 ± 0.02
14	Si	7.55 ± 0.05	7.55 ± 0.02
15	P	5.45	5.57 ± 0.04
16	S	7.21 ± 0.06	7.27 ± 0.05
17	Cl	(5.5)	5.27 ± 0.06
18	Ar	(6.58)	—
19	K	5.12 ± 0.13	5.13 ± 0.03
20	Ca	6.36 ± 0.02	6.34 ± 0.03
21	Sc	3.1	3.08 ± 0.04
22	Ti	5.02	4.93 ± 0.02
23	V	4.0	4.02 ± 0.02
24	Cr	5.67 ± 0.03	5.68 ± 0.03
25	Mn	5.45	5.53 ± 0.04
26	Fe	7.67 ± 0.03	7.51 ± 0.01
27	Co	4.92 ± 0.04	4.91 ± 0.03
28	Ni	6.25 ± 0.04	6.25 ± 0.02
29	Cu	4.21 ± 0.04	4.26 ± 0.05
30	Zn	4.60 ± 0.08	4.65 ± 0.02
31	Ga	2.88	3.13 ± 0.03
32	Ge	(3.63)	3.63 ± 0.04
33	As		2.39 ± 0.04
34	Se	—	3.35 ± 0.03
35	Br	—	2.63 ± 0.07
36	Kr	—	3.21 ± 0.05
37	Rb	2.60	2.40 ± 0.03
38	Sr	2.9	2.93 ± 0.03
39	Y	2.24 ± 0.03	2.22 ± 0.03
40	Zr	2.56 ± 0.05	2.58 ± 0.05
41	Nb	(2.10 ± 0.10)	1.41 ± 0.06
42	Mo	1.92 ± 0.05	1.96 ± 0.02
44	Ru	1.84 ± 0.07	1.82 + 0.02
45	Rh	1.12 ± 0.12	1.09 ± 0.03

Table 2.5 (continued)

Z	Element	Solar abundance[a]	Meteoritic abundance[b]
46	Pd	*1.69 ± 0.04*	1.70 ± 0.03
47	Ag	(0.94)	1.28 ± 0.03
48	Cd	1.86	1.76 ± 0.03
49	In	1.66	0.82 ± 0.03
50	Sn	2.0	2.14 ± 0.04
51	Sb	1.0	1.10 ± 0.07
52	Te	—	2.25 ± 0.05
53	I	—	1.51 ± 0.05
54	Xe	—	(2.19) ± 0.05
55	Cs	—	1.12 ± 0.03
56	Ba	*2.13 ± 0.05*	2.19 ± 0.02
57	La	1.22 ± 0.09	1.21 ± 0.02
58	Ce	1.55	1.62 ± 0.02
59	Pr	0.71 ± 0.08	0.79 ± 0.03
60	Nd	1.34 ± 0.11	1.48 ± 0.03
62	Sm	0.80 ± 0.11	0.97 ± 0.03
63	Eu	*0.51 ± 0.08*	0.54 ± 0.03
64	Gd	1.12	1.07 ± 0.02
65	Tb	0.2	0.32 ± 0.05
66	Dy	1.1	1.15 ± 0.02
67	Ho	(0.26)	0.50 ± 0.02
68	Er	0.93	0.96 ± 0.02
69	Tm	0.00 ± 0.15	0.14 ± 0.03
70	Yb	1.08	0.94 ± 0.02
71	Lu	(0.76)	0.12 ± 0.03
72	Hf	0.88 ± 0.08	0.80 ± 0.03
73	Ta	—	− 0.09 ± 0.04
74	W	*1.11 ± 0.15*	0.69 ± 0.03
75	Re	—	0.26 ± 0.05
76	Os	1.45 ± 0.10	1.41 ± 0.04
77	Ir	1.35	1.37 ± 0.03
78	Pt	1.8 ± 0.03	1.69 ± 0.04
79	Au	(1.13)	0.82 ± 0.06
80	Hg	—	(1.27) ± 0.18
81	Tl	(0.9)	0.82 ± 0.04
82	Pb	1.90	2.05 ± 0.03
83	Bi	—	0.71 ± 0.03
90	Th	0.02	0.08 ± 0.02
92	U	(≤ − 0.47)	− 0.49 ± 0.03

[a] Abundances are given as $\log(N_{el}/N_{H}) + 12.00$. Values in italics correspond to accurate atomic data and use the solar model of Holweger and Muller (1974). Values in brackets correspond to elements not observable in the solar photosphere (He, Ne, etc) or elements where the oscillator strengths are not accurately known.

[b] Meteoritic abundances correspond to the analysis of Cl chondrites.

particularly for UV lines. NLTE calculations require collisional excitation rates and ionisation cross-sections as well as photoionisation cross-sections.

Some elements (e.g. He, Ne) are not observable in the solar photosphere because they have no line located in the photospheric spectrum (visible, near IR) or because their visible lines such as He I 5876 A require too much excitation energy to be photospheric in origin. They are observable only in the chromosphere or corona. The helium abundance is calculated from the analysis of prominences, the solar wind or solar cosmic rays.

Abundance studies carried out for many stars, cosmic rays, and the interstellar medium show that the Sun's composition is close to what is called *cosmic*, which is that of fairly young objects ($t \leq 5 \times 10^9$ y) in the Universe. Such objects have formed from an interstellar medium which has already been enriched in heavy elements by several generations of supernovae. Deviations from this composition once claimed in hot main-sequence stars have been invalidated by NLTE analysis of the lines. In contrast, very old stars (galactic halo, globular clusters, population II in general) or those in advanced stages of evolution (carbon stars, red giants, barium stars, etc.) show clear abundance differences from the solar composition, for certain elements. The first group is deficient in heavy elements, since they were formed in a region of the interstellar medium poor in metals. The second group have undergone violent events during their evolution which have brought to the surface elements recently formed by nuclear reactions in the centre (see Chap. 8).

In general we assume that the abundance $A = N_A/N_H$ is constant with depth in the atmosphere. There are exceptions: stars with peculiar chemical composition (Ap, Bp), and helium in the solar corona. These peculiar abundances can be explained by diffusion of elements under the effects of gravity, concentration and temperature gradients, and radiative forces (Sect. 5.3).

2.4 Fine Structure of the Quiet Solar Atmosphere

As the Sun is the nearest star, we see its surface in a wealth of detail which will remain hidden for a long time for other stars. However, stellar physics can no more ignore this surface structure than classical physics can ignore the microscopic interpretation of the phenomena it seeks to explain. In fact the Sun's heterogeneous surface structure reveals the interaction between motions of the surface gas, its radiative properties, and its organisation by the magnetic field produced in the deeper convective layers.

We observe solar fine structure not only in white light, UV continua, and spectral bands of various widths but also in the lines of the photosphere, chromosphere, or transition zone. In this section we shall describe the structure of the *quiet* component of the solar atmosphere without going much into the relevant physical processes. This is the component which

is affected, albeit weakly, by the eleven-year cycle which is very noticeable in active components such as sunspots, active regions, prominences, and coronal holes.

The Sun's surface structure is primarily marked by local brightness variations. Various horizontal scales are apparent, corresponding to different values and scales of the observed magnetic field.

2.4.1 The Quiet Photosphere

Lateral structure observed at the photosphere has three scales: *granules* and *mesogranules*, from Mm to 10 Mm, *supergranules*, from 20 Mm to 50 Mm, and *flux tubes* (< 0.3 Mm).

The Granulation.   The granulation is seen in white light (Fig. 2.11) as fairly irregular cells which are brighter than the intergranular space around them. It is no longer visible in the photospheric UV continuum at 2200 Å. The average distance between granules is about $1.8''$. A range of sizes is revealed as the spatial resolution is increased. There is a critical size at about $1.4''$. Roudier and Müller (1986) show that large granules have fractal dimension 2.15, and smaller ones 1.25. This critical size corresponds to a break in the slope of the power spectrum of the granules, the smaller ones
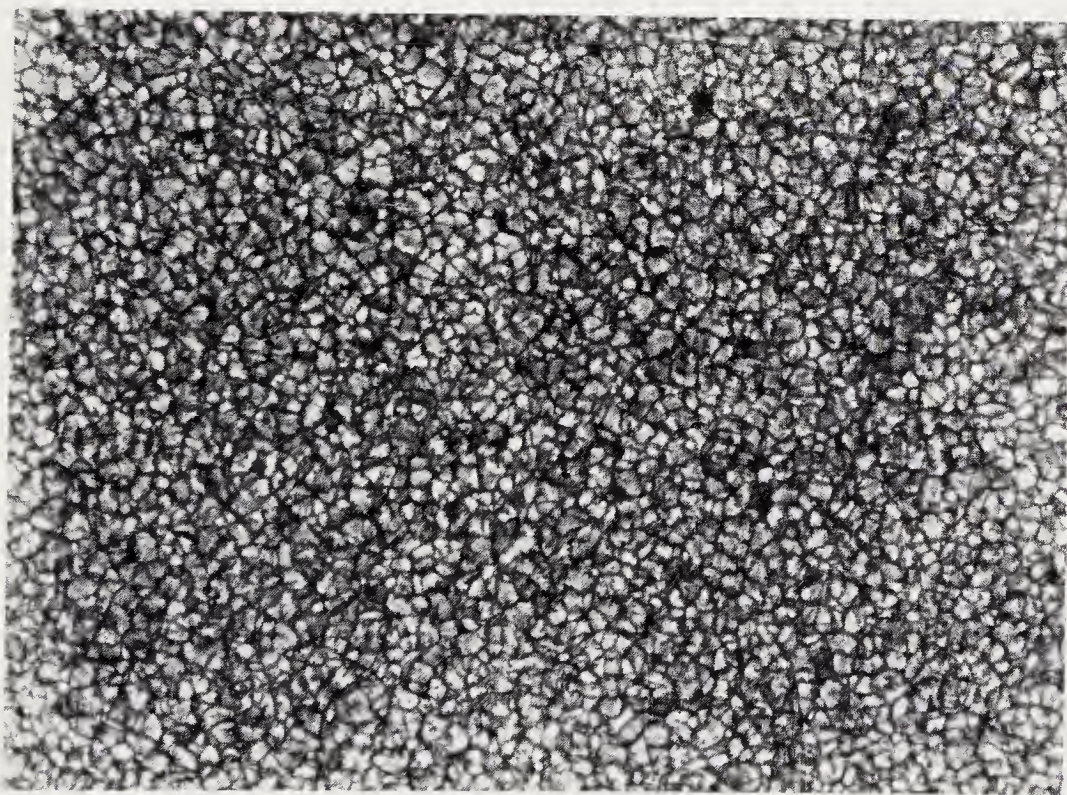

Fig. 2.11. Photograph of the solar granulation taken on 1978 July 9. Wavelength 5750 Å, passband 100 Å; resolution $\sim 0.25''$. (Photograph courtesy of Observatoire du Pic du Midi)

having a Kolmogorov spectrum as expected for the smallest scales of convective turbulence (see Chap. 5, Fig. 7). Granules are small convective cells penetrating into the photosphere from the convective layer (the penetration is about 100 km). Convective cells about 10 times the size of a granule are called *mesogranules*. They are observed as a pattern in brightness and spectral line shift.

Granules last for the order of ten minutes. Cinematography shows that they are born by the mixing or fragmentation of preexisting granules; they disappear through the same processes, although the smaller ones can simply vanish.

Granule motions are deduced from *spectra* taken at high spatial resolution (0.75″ at best from the ground). The vertical velocity is typically about 1 to 2 km s^{-1} at 100 km above the top of the convective zone, while the horizontal component is 2 or 3 times as big. Rising motions (hot granules) are observed amongst a predominance of falling motions. The associated temperature fluctuations are about 100 to 200 K.

The Supergranulation.   This appears at the level of the photosphere because of the tracer provided by the *network bright points*. These are structures at the limit of spatial resolution (typically 0.22″), discovered by observations with a very narrow spectral passband centred on photospheric lines or in the wings of Hα. These bright points never appear inside a granule, but only in the intergranular regions, where they align themselves along the boundaries of huge cells (diameters 30 to 50″) called *supergranules*. The boundaries of the supergranules are also called the *photospheric network*.

The photospheric network is also observed in the UV at wavelengths which probe the temperature minimum ($\sim$ 1600 Å). At this height, greater than that considered above, the network boundaries are bright and very clearly defined. Numerous bright points fill the whole interior of the supergranule, not just the boundaries.

Horizontal mass motions from the centre to the edge are observed in supergranules, with velocities 300 to 400 m s^{-1}. Vertical falling motions of speed 100 m s^{-1} are observed at the boundaries of the cells.

Association with Magnetic Field Regions.   This is an important property of the photospheric network. The magnetic field is concentrated at the edges of the supergranulation cells. The measured field intensity (several hundred gauss) is not typical of its value in the finer structures where it emerges, called *flux tubes*. In fact the spatial resolution of present measurements of the magnetic field does not exceed 1 to 2″. We have to use indirect methods to find the fields in the flux tubes, which are provisionally identified with the bright points of the network (Stenflo 1973). A typical value for the field in a flux tube is 1 kG. The present description of flux tubes will be greatly improved once spatial resolution is improved (e.g. the Themis solar telescope).

Table 2.6. A. Fine structure of the quiet photosphere

Structure	Mean diameter (km)	Mean lifetime	Remarks
Granules	1 000	8 min	Small convective cells
Mesogranules	5 000–10 000	2 h	Larger convective cells
Supergranules	30 000–35 000	20 h	Large convective cells
Network bright points	150?	18 min	A dense chain of bright points is sometime called *filigree*
Cell bright points	?		

B. Fine structure of the quiet chromosphere

Structure	Mean diameter (km)	Mean lifetime	Remarks
Spicules	500–1 500	5 min	
Network	30 000–35 000	20 h	
– Bright mottles	1 500–4 000	11–12 min	Parts of the network boundaries
– Dark mottles	1 000–8 000	5–15 min	Identical to spicules
– Fibrils	1 000–2 000	6–12 min	Elongated horizontal structures
Supergranules cells	30 000–35 000		
– Bright points	1 000	3–4 min	

The three scales of structure observed in the photosphere give essential information about solar convection and its interaction with the magnetic field, which determines the size and number of flux tubes. Table 2.6A summarizes the properties of the photospheric fine structure.

2.4.2 The Quiet Chromosphere

Higher up in the solar atmosphere, the main structures are *spicules* and the *chromospheric network*.

Spicules. These were discovered in white-light observations of the Sun's limb during an eclipse. They are clearer when observed using a Lyot filter in the Hα line, either in the core or wings. Quasi-monochromatic images made without scanning are called *filtergrams*. At the limb, spicules appear like blades of grass, i.e. elongated cylinders, essentially radial, but sometimes inclined. They reach down to the lower chromosphere, although they are difficult to see at lower altitudes as they cover each other. Their vertical extent is not well defined; the most likely mean value is about 9000 km above the photosphere. The presence of spicules clearly shows that the

chromosphere is no longer spherically symmetrical above 2000 km. On the disc, spicules are also called *dark mottles*. They lie along the boundaries of the chromospheric network. The geometrical properties of spicules are summarised in Table 2.6B.

The spicules are remarkable for their dynamical properties. Their average lifetime is 5 min if observed at the limb. Their individual motions can be seen either directly, i.e. by imaging, or spectroscopically through the Doppler effect. The first method gives the velocity component in the plane of the limb and the second the velocity normal to this plane. Spicules show rising and falling motions of average speeds 20 to 25 km s^{-1}, which seem to be directed along the spicule's axis. The motion is predominantly upwards, so that spicules transport mass from the chromosphere to the transition zone. The mass flux is estimated at 10^{15} protons cm^{-2} s^{-1}.

The Chromospheric Network.   This picks out a cell structure at the height of the chromosphere. It appears bright if we observe in the Ca II H and K lines, in the centre of Hα or Lyα, and dark in the wings of Hα. From the temperature minimum, where it occupies about 10 % of the solar surface, it reaches into the chromosphere, occupying 35 to 40 % of the surface observed in transition-region lines.

The main instrument for studying the chromospheric network at moderate resolution is the *spectroheliograph*. Comparing spectroheliograms obtained for photospheric and chromospheric lines of different formation depths shows that the chromospheric network extends the photospheric network in height: its cells fit smoothly over those of the photospheric supergranulation.

All of the network and the enclosed cells are in motion: matter rises at the centre of a cell, with horizontal motions to the edges, where there are falling motions. This *downdraft* has a typical mean speed of 800 m s^{-1} in the middle chromosphere, 4 km s^{-1} in the transition zone. The mass flux is not conserved, being about 10^{-4} to 10^{-2} of its photospheric value in the middle chromosphere and transition region respectively. At the lowest level are horizontal flows feeding the falling motions. In contrast, the descending flux at the transition region is of the same order as the rising flux in the spicules (Athay 1986), so that some of the vertical circulation at the cell boundaries can be seen as the return flow of the spicules.

At high spatial resolution the chromospheric network shows finer structures on the disc:

— *Dark mottles*, often grouped in "rosettes" and "bushes". These are just the spicules.

— *Bright mottles*, the basic resolved bright elements of the network boundaries and thus the extension of the bright points of the photospheric network.

— *Fibrils*, which are horizontal structures, well observed in Hα, extending above the supergranules, perhaps segments of flux tubes.

Inside the cells are *bright points*, seen in the Ca II K line (we have already encountered these near the temperature minimum).

The properties of these various fine-structure elements are summarised in Table 2.6B.

2.4.3 The Quiet Transition Region

The chromospheric network is observable at the formation heights for the lines of C III, C IV, O IV, O VI, up to the coronal line Mg X (see Table 7.1 for the formation temperatures of these lines).

In all the regions of the quiet Sun where the network appears, from the temperature minimum to the lower corona, radiative losses are larger at the edges of the network than at the centre. It thus appears that the non-thermal heating is stronger at the edges. At the photospheric level these boundary regions coincide with concentrations of magnetic flux. Although the field is inferred, not measured, at the chromosphere, transition zone, or corona, it is tempting to conclude that the non-radiative heating in the outer layers of the quiet Sun is connected with the presence of a magnetic field diverging above the chromospheric network. Figure 2.12 shows the field configuration, together with the chromospheric network and the flows caused by convection at the centre of a cell and by spicules at the boundaries.

This section on the fine structure of the solar atmosphere has shown the existence of various lengthscales and the close relation between velocity fields and magnetic flux concentrations. The magnetic regions encountered have open fieldlines, except perhaps for the small bipolar regions associated with the bright points. It is clear that understanding the mass and energy
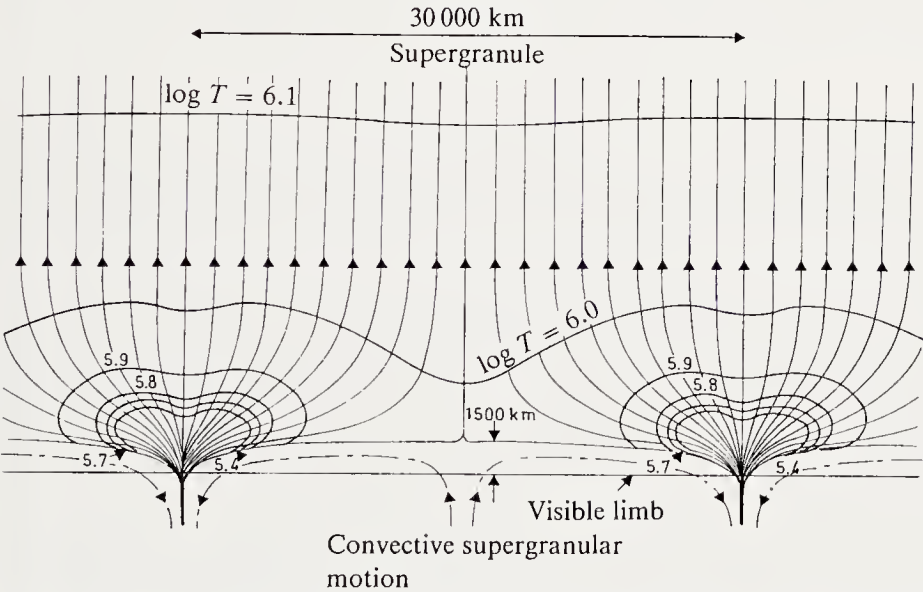

Fig. 2.12. Model of the network showing the magnetic fieldlines, convective flow lines, and isotherms. (From A.H. Gabriel, in *IAU Coll.* 36, ed. by R.M. Bonnet, P. Delache (1976), p. 375)

balance requires a good optical identification and modelling of all the types of lateral structure. The rising mass flux in the spicules is certainly not the origin of the solar wind, which has only about 1 % of the mass flux (3×10^{-11} g cm^{-2} s^{-1} in the wind, compared with 2×10^{-9} g cm^{-2} s^{-1} in spicules) and the matter in the spicules must fall back. There is thus a circulation involving spicules and the edges of supergranules.

2.4.4 The Quiet Corona

One might give this designation to that part of the corona above the weakly magnetised regions of the Sun. However, it is actually composed of active regions only weakly active or of weakly expanding coronal holes. These are the main constituents of the corona as described in Sect. 2.5.

2.5 Resolved Structure in the Active Sun

In this section we shall describe the characteristic regions of the active Sun, reserving the variability linked to solar activity to Chap. 7. We shall emphasise the topologies and lengthscales, relating them to the measured magnetic field. This gives the size of all the structures if $\beta = P_g/P_{\mathrm{mag}} = (\Re \rho T/\mu)(B^2/8\pi) = c_s^2/v_a^2$ is much smaller than 1, which holds for layers above the solar photosphere. Here c_s is the isothermal sound speed and v_a the Alfvén speed. The structuring of the photosphere into flux tubes, where $\beta > 1$, creates bundles of force tubes through magnetoconvective phenomena originating below the photosphere. A hierarchy of magnetic elements thus appears at the surface of the Sun, from the photosphere to the corona. Methods of measuring magnetic fields are discussed in Chap. 7.4.

2.5.1 Sunspots and Active Regions

Sunspots.   These were the first sign of irregularity observed on the Sun's surface, initially with the naked eye (by the Chinese and Greeks), then with a *camera obscura* (Fabricius), and at about the same time (1610–1611) with a refracting telescope (Galileo). A spot has a central region, the *umbra*, surrounded by a brighter region, the *penumbra*, which are clearly separated; the penumbra is itself quite distinct from the neighbouring quiet photosphere. The full diameter of a spot is between 20 and 60 Mm, that of the umbra alone about 10 to 20 Mm. The penumbra is striated by elongated dark radial structures, called *fibrils* (Fig. 2.13).

The magnetic field of sunspots was measured from 1908 by Hale, who discovered intense fields. At the centre of a spot the field intensity B reaches 3000 to 4000 G and is vertical in direction; in the outer penumbra it is about 1000 G, and the fieldlines are inclined, almost horizontal in isolated spots; at the extreme edge of the penumbra B is less than 100 G.
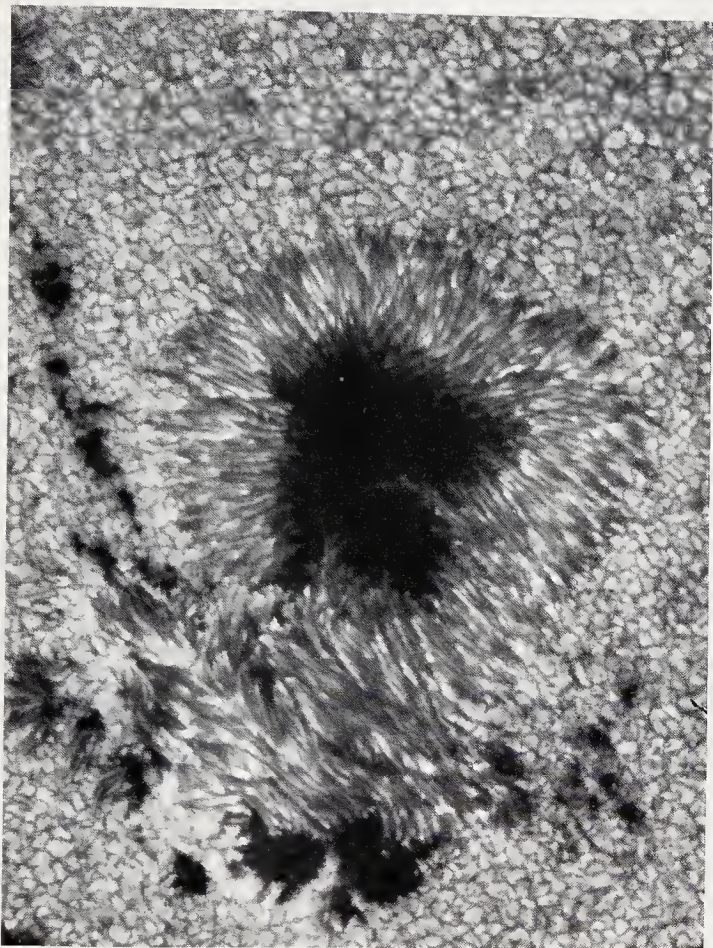
Fig. 2.13. A sunspot and the surrounding granulation. Visible are the umbra (central region of the spot) and the penumbra, formed of radial fibrils. Wavelength: 5280 Å, passband: 100 Å, resolution ∼ 0.3″. (Photograph by Observatoire du Pic du Midi)

In practice most spots are not isolated, but go in pairs, and they are found only in a well-defined latitude region, between +30° and −30°. At the beginning of the solar cycle the spots appear at high latitude; the pairs are oriented east–west and migrate towards the solar equator in time. Spots not only trace the solar magnetic field, but are good indicators of large-scale motions in the Sun. The spots extend downwards below the photosphere and are anchored at the depth where the field is generated. They follow the Sun's differential rotation (the surface rotation period increases from the equator to the poles; see Table 2.1) and move more rapidly than the surrounding photosphere at each latitude (by 1 to 3 %). This suggests that the Sun's rotation increases inwards, at least to the depth of the spots' feet, a result contradicted by helioseismology.

Inside a moving pair of spots, one, e.g. the leading spot, has positive magnetic polarity and the other negative, with the opposite relation across the solar equator. The configuration survives for about eleven years and then reverses (see Chap. 7.1). A sunspot pair thus forms a magnetic dipole;

83

the fieldlines run from one spot to the other. There is a line of inversion of polarity between the two spots, which is the locus of points where the vertical field component changes sign (neutral line).

Spots do not on average survive more than two or three rotation periods. They decay slowly at first, then rapidly; they leave behind a system of bright spots in the umbra (umbral dots).

The basic structure of a sunspot magnetic field is a tight bundle of flux tubes held close together below the photosphere, whose ends float under magnetic buoyancy. In the umbra at presently available resolution only the tubes are visible; if there are regions with weak or absent magnetic field this may be marked by umbral dots. No granules are observed in the spots, so convection does not appear there. The unsolved problem arises of what happens to the flux blocked by the spots, since they are darker than the surrounding photosphere. We do not know the ultimate size of the flux tubes or their filling factor; the present upper limits on their diameters are 0.2 to 0.3″.

The spot equilibrium is dominated by magnetic pressure. At a given height there is a balance between $P_g(\text{int}) + P_{\text{mag}}$ and $P_g(\text{ext})$, so that the gas pressure $P_g(\text{int})$ inside the spot is lower than the exterior gas pressure $P_g(\text{ext})$. The electron temperature is therefore lower: analysis of molecular bands (TiO, SiO, CO) gives values 3000 to 3400 K. Optical depth $\tau = 1$ is 600 to 1000 km deeper inside the spot than in the normal photosphere.

Active Regions. Sunspots are parts of *active regions*, areas where there is considerable evidence of magnetic fields. Thus regions where the magnetic field emerges in a less compressed form are also localised above the spots; they are called *plages* and are composed of *faculae*. These regions are not observed at the level of the photosphere like sunspots, but at the height of the chromosphere or transition region. In a plage the filling factor is less than in the underlying spot, at about 5 to 10 %; the average measured field intensity is about 100 G, so that the local value of B is of the order of 1500 G. At the level of the transition region the pressure and temperature gradient are each fives times larger than in the quiet Sun. Activity linked to the plages will be studied in Chap. 7.

All solar activity is confined to active regions. For example, *flares* occur in regions near sunspots, exactly where the magnetic fieldlines are in tension at the line of polarity inversion between two sunspots. The magnetic properties of the flare site are clearly vital to an understanding of the underlying mechanism. As we shall see, prominences observed on the disc also straddle lines of polarity inversion (or neutral lines).

2.5.2 Prominences

Prominences are thin layers of cold dense plasma within the corona. They are seen either as bright at the limb (Fig. 2.14) or as dark *filaments* against the disc. Their temperature is about 7000 K and the electron density con-
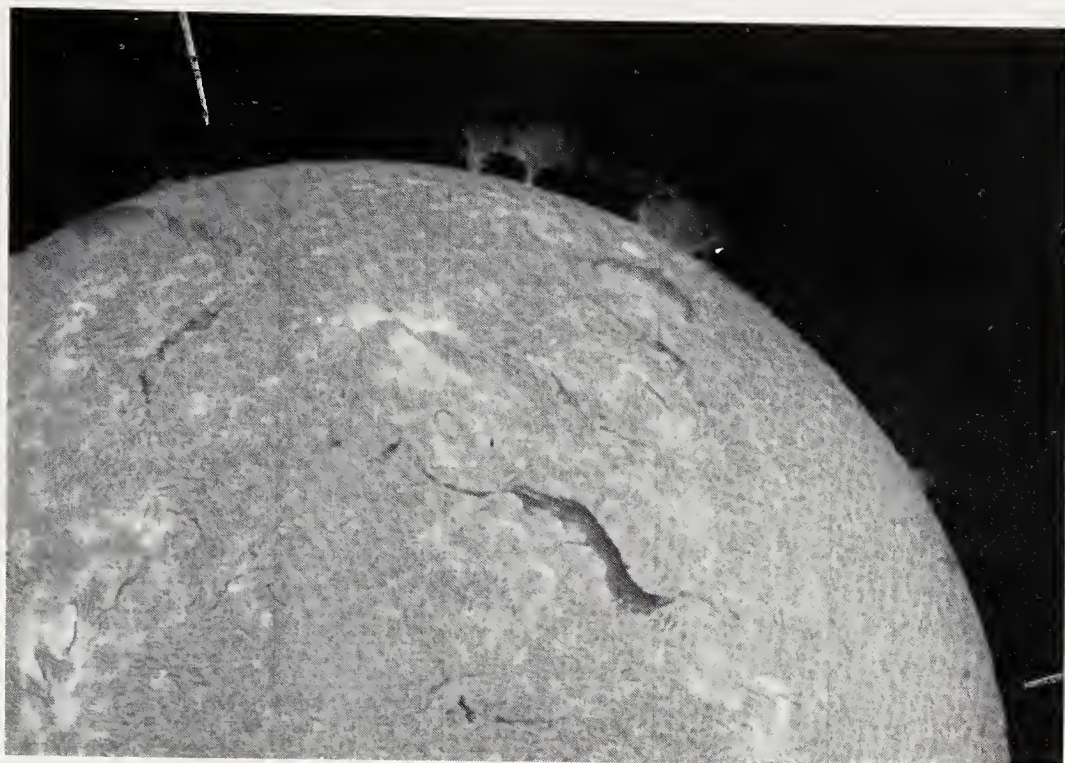
Fig. 2.14. Prominences (bright against the sky background) at the Sun's limb and filaments (dark against the disc). Visible also are bright faculae, regions of strong magnetic field. Image obtained 1980 August 2 in Hα. Resolution about 2″. (Spectroheliogram of the Observatoire de Meudon)
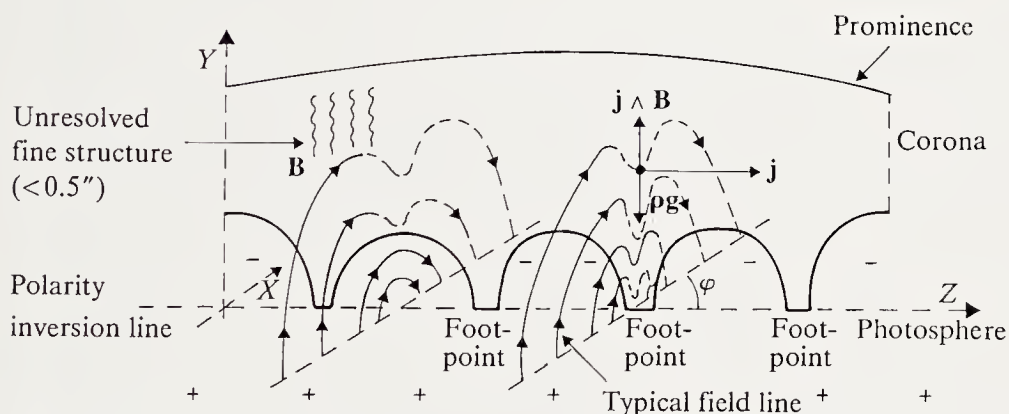
Fig. 2.15. Schematic view of a prominence and associated magnetic field. (From J.M. Malherbe, Thesis, Univ. Paris VII, 1987)

troversial: either $< 10^{10}$ cm^{-3} or ten to twenty times greater. Figure 2.15 shows their shape: the sheet forming the prominence is situated above a neutral line of the magnetic field and enters the photosphere at footpoints. The magnetic field is roughly perpendicular to the sheet, i.e. in the (X, Y) plane. The angle between the filamentary axis (Z) and the field vector is less than 20°.

The equilibrium of the filaments is due to a balance between the $\boldsymbol{j} \times \boldsymbol{B}$ force and gravity $\rho \boldsymbol{g}$, the current $\boldsymbol{j}$ being given by Ampère's law as $(c/4\pi)\,\mathrm{curl}\,\boldsymbol{B}$. We have $j_z B_x = \rho g$, which causes a dip of the fieldlines. It is the B_x field component which supports the prominence. The footpoints occur in regions of the photosphere where the field is sheared. The fieldlines in the (X, Y) plane of Fig. 2.15 are stirred at their bases by photospheric motions.

We can distinguish *quiescent* prominences, situated at the edges of active regions or between them, or in polar regions, and *plage* prominences, associated with active regions.

The typical size of a quiescent prominence is as follows: length $\simeq 200\mathrm{Mm}$, height $\simeq 50$ Mm, thickness $\simeq 2$ Mm. A prominence thus extends about one pressure scaleheight into the corona $(T = 10^6$ K$)$. We note that the length is about the scale of the largest convection cells (see Chap. 5.5).

Prominences have fine structure: unresolved vertical *threads*, or clearly defined vertical arches.

The field can be measured by the Hanle or Zeeman effects. The first method gives the vertical field and shows that the field does not lie in the plane of the prominence. The intensity is about 5 to 10 G in quiescent prominences, reaching 20 to 150 G in plage prominences. It appears to be uniform inside the prominences.

Prominences are in a dynamic state, since rising matter motions are observed, e.g. in Hα or in C IV (speeds a few km s^{-1}). Cine films of this process are spectacular, particularly as the whole prominence begins to disappear, when the motions accelerate and the prominence vanishes abruptly ("disparition brusque"). The origin of the forces acting on the prominences just before their disappearance is still ill-understood.

The problems of formation and of modelling prominences have received much attention. We note first that the mass of a quiescent prominence of volume V,

$$M_{\mathrm{pro}} = N_p m_p V = 10^{11} \times 1.67 \times 10^{-24} \times 2 \times 10^{10} \times 5 \times 10^9 \times 2 \times 10^8$$
$$= 3 \times 10^{15}\ \mathrm{g}$$

is a significant fraction of the mass of the corona

$$M_{\mathrm{cor}} = N_p m_p A H\,,$$

where A is the area of the photosphere, and H the coronal scale height $(5 \times 10^9$ cm$)$. With $N_p = 3 \times 10^8$ cm^{-3} in the corona, we find

$$M_{\mathrm{cor}} = 1.5 \times 10^{17}\ \mathrm{g}\,.$$

Two formation mechanisms for prominences have been proposed: (a) condensation (by thermal instability) of the coronal material, although $M_{\mathrm{pro}}/M_{\mathrm{cor}}$ is a problem (the ratio is too large) if N_p is of the order of
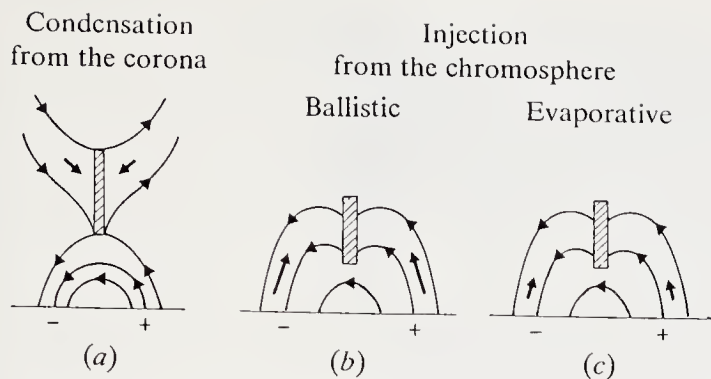
Fig. 2.16. Formation scenarios for a prominence. The prominence is the *hatched region*. (a) The prominence material condenses from the corona; (b) and (c) injection from the chromosphere. (From J.M. Malherbe, Thesis, Univ. Paris VII, 1987)

10^{11} cm^{-3} in the prominence, as assumed above; this objection disappears if N_p is less than 10^{10} cm^{-3}; (b) injection from the chromosphere, with two possible variants: ballistic injection by transient jets caused by rapid variation of the gas or magnetic pressure at the foot of a loop, or continuous evaporation because of a lack of heating of the medium above the neutral line. The two mechanisms are shown schematically in Fig. 2.16, where the lines of force of the magnetic field and the velocity vectors of the matter are indicated: the mechanism of injection from the chromosphere has not yet been observed.

2.5.3 Coronal Holes

These are large-scale regions of the corona observable in X-rays in wavelength bands of a few Å (Fig. 2.17), in EUV lines such as He II (304 Å), Ne VII (465 Å), or Mg IX (368 Å), and in He I (10830 Å). They appear as regions fainter than the neighbouring corona, but their relative contrast decreases in lines formed at lower temperatures: coronal holes are less noticeable in the transition region than in the corona (the line intensities are reduced by 25 to 30 %) and are not discernible in the chromosphere. The He I 10 830 Å line (the only coronal-hole diagnostic observable from the ground) is an exception, since its formation mechanism involves coronal radiation directed towards the chromosphere.

Coronal holes can be localised in polar regions, but are also observable at low latitudes heliographically. They lie over large unipolar magnetic regions of the photosphere, where the fieldlines are open. They appear preferentially at certain longitudes and are associated with new dipolar magnetic regions. It thus appears that large-scale field topology fixes the position and geometry of coronal holes. A coronal hole may cover from 1 to 5 % of the Sun's surface, and a polar hole up to 10 %.

The orbiting platform Skylab (1973 May – 1974 February) supplied the main results on the solar X-ray emission. Coronal holes can be modelled (Fig. 2.8) by the methods described in Sect. 2.2 for the transition region and
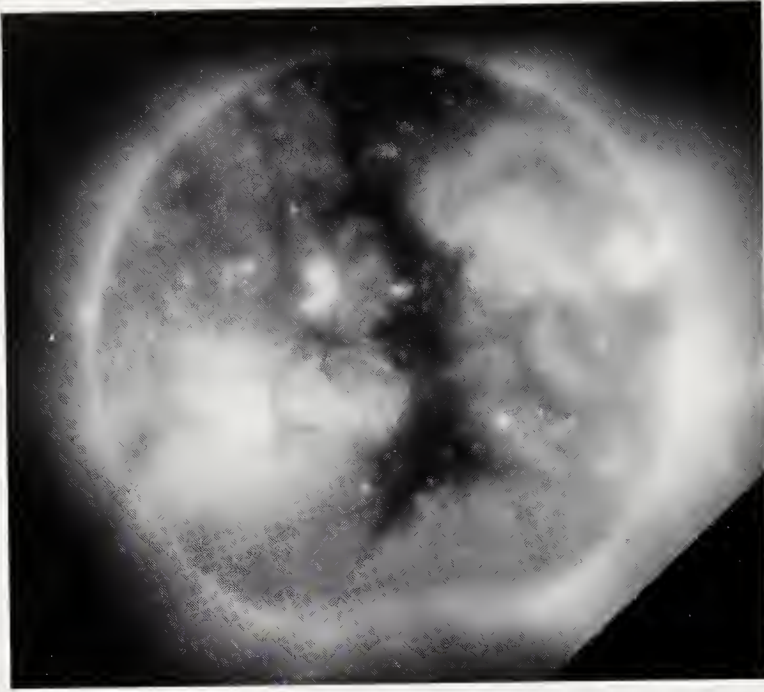
Fig. 2.17. Image of a coronal hole, seen by the American Science and Engineering Telescope on board Skylab, 1973 June 1. It shows the hole stretching from the pole to low latitudes. This hole persisted for the whole Skylab mission. The passband is 3–30 Å, 44–60 Å

the corona. The transition region is about 5 times thicker in a hole than in the quiet corona, the temperature gradient dT_e/dz about 5 to 10 times smaller and the pressure 2 or 3 times smaller in the quiet corona.

Another property of coronal holes is their rotation. Unlike sunspots, whose migrations are the best tracers of the surface differential rotation (Table 2.1), coronal holes rotate *rigidly*; there is no more than a 3 % variation in their rotation speed between the poles and the equator. Their synodic period is 27 days.

Finally, coronal holes are associated with high-velocity flows of the solar wind, as observed near the Earth's orbit. The matter velocity at the base of a coronal hole is of the order of 20 km s^{-1}; Fig. 2.8 shows the expansion velocities in various coronal regions. We note that these velocities vary over the solar cycle, because a coronal hole produces a more rapid flow at solar minimum than maximum. The *Ulysses* space mission, launched in 1990, will for the first time observe the solar wind from outside the ecliptic plane, since it is due to fly over the solar poles. It will provide a 3D view of the solar wind.

The term representing the energy flux in the solar wind dominates the energy equation for a coronal hole (2.9).

Table 2.7 Resolved structure of the active Sun

Structure	Mean diameter	Mean lifetime
Spot		$\sim$ 100 d (2 to 3 rotations)
– Umbra	10 000–20 000	for large spots, $\sim$ 1 d
– Penumbra	20 000–60 000	for small spots (pores)
– Umbral dot	1 000	15–30 min
– Flux tube	< 100	
– Fibril	200	2 h
Plage		$\sim$ 150 d (5 to 6 rotations)
– Facula		
– Ephemeral active region		
Prominence		
– Quiescent		several months
– Active region prominence		several weeks
– Thread	< 300	several mins
Coronal hole		$\sim$ 200 d (6 rotations)
– Bright point		

2.5.4 Coronal Bright Points

These are very obvious on X-ray images of the corona, and cover deeper short-lived active regions, and, deeper still, photospheric bright points.

Table 2.7 summarises the nomenclature, sizes, and lifetimes of the resolved structure of the active Sun.

2.6 Remarks

We should ask what connection there is between the variety of phenomena observed at various heights in the Sun's atmosphere and the models whose computation principle was explained in Sect. 2.2. It is clear that the physics of the Sun's outer layers is very complicated (see e.g. Priest 1982) and modelling can only be seen as a stage, giving local values of N_e, T_e, and the velocity field. Clearly it is only reasonable at present to model individual structures (spots, faculae, flux tubes, etc.) rather than to treat the solar surface as homogeneous. However, we must go beyond this modelling in order to discover the origin of such phenomena. The formation of structures on the active Sun involves the magnetic field and its interaction with the solar plasma. The scales of the magnetic field, the properties of regions situated at foot points of lines of force, instabilities, waves, and ordered motions all result ultimately from mechanisms at work in the upper part of the convection zone. The book by Priest (1982) discusses in greater depth the study of one of the most complex objects in astrophysics.

3. Stellar Evolution

3.1 Basic Internal Structure

3.1.1 Basic Observational Data

It may appear strange to begin a discussion of observational data with abstract questions and concepts dating from the middle of the 19th century (Kelvin 1862; Lane 1868). A star is an object characterised by a number of global data: *spectral properties, luminosity, mass,* etc.; binary systems are fundamental for mass determinations. The idea of evolution arises once we ask about the energy reserves of a star. A star of 10 solar masses, with a luminosity 10^4 times that of the Sun, would transform all its hydrogen into helium in 8 million years, a time much shorter than the age of the Galaxy (10 to 15 billion years). This elementary result immediately implies that stars must be forming in the Galaxy right now and that our view of the sky is a kind of snapshot of objects of very different ages and thus very different stages of evolution. To the global data mentioned above we should add the *age*. We know the Sun's age, under the assumption that it is close to that of the Solar System, i.e about 4.55 billion years (Encrenaz, Bibring 1990, p. 350). Stellar-evolution models are only able to give estimates for the *ages of galactic clusters*, but we are unable to estimate the ages of isolated stars, and scarcely those of binaries (in the rare cases where the evolutionary state of one of the components is identifiable).

The presence of easily identified resonance lines in absorption, blue-shifted with respect to the photospheric lines in red giants (α Ori, α^1 Her, o Cet, ε Peg), has shown since 1951 that stellar envelopes may expand. Such very extended envelopes have also been detected through infrared dust emission, maser lines (OH, H_2O), and emission lines in the millimetre region. These expanding envelopes show the existence of a stellar *wind*, and thus mass loss from the star. From the observations one can obtain estimates of the *mass loss rate* (we shall return to this in Sect. 3.7 and Chap. 4). For the moment we note that mass loss rates can be 10^{-7}–10^{-8} $M_\odot/y$ for red giants and 10^{-6}–10^{-7} $M_\odot/y$ for red supergiants. Among blue stars, the mass loss rates in Wolf–Rayet stars can reach 10^{-5} $M_\odot y^{-1}$. Such high mass loss rates signify relatively brief stages of evolution (roughly a few million years) but must of course be taken into account in the study of stellar evolution.

The goal of stellar-structure theory is to explain the global properties of stars and follow these in time. This requires a knowledge of the distri-

butions of physical parameters such as temperature, density, pressure, and other thermodynamic variables, as well as the chemical composition and the hydrodynamics and magnetohydrodynamics of the stellar plasma.

The full programme is clearly very ambitious and it is essential to break it into stages. The study of internal structure has a basic reference model which is assumed to be stationary, non-rotating, and of constant mass. The description *stationary state* comes from assuming that dynamical terms are negligible. We retain terms involving secular evolution: the radiation of the star itself results from irreversible processes such as changes of chemical composition or radius. The timescales of these processes are so long compared with dynamical timescales that we can regard the reference star as being in equilibrium. This still holds even when we abandon the assumption of constant mass, as the energy $\frac{1}{2}Mv^2$ involved in the mass loss is small compared with the star's luminosity. Under mass loss the star passes thus through a sequence of equilibrium configurations.

3.1.2 Basic Equations

We give here the internal-structure equations in Newtonian form (relativistic effects are considered in Chap. 8).

Poisson's Equation

$$\Delta\Phi = -4\pi G\rho\,,$$

where Φ is the gravitational potential, ρ the density, and G the constant of gravitation.

Hydrostatic Equilibrium

$$\nabla P + \rho\nabla\Phi = 0\,,$$

where the pressure P is given by an equation of state as a function of local variables ρ and T (density and temperature) and chemical composition, described by the mass abundance X of hydrogen, Y of helium, and Z of all other elements ($X + Y + Z = 1$). The detailed abundance of the other elements is only important for problems of radiation transfer and nuclear chemistry.

The Heat-Transfer Equation

$$F = -K\nabla T\,,$$

where K is the thermal conductivity and F the flux (erg cm^{-2} s^{-1}). For the moment it suffices to state that K is a function of local variables (density, temperature) and detailed chemical composition of elements other than hydrogen and helium. The integral $\int F\mathrm{d}S$ over the spherical stellar surface is

the luminosity L. We shall discuss the precise meaning of the term "surface" when considering boundary conditions.

The Energy Equation

$$T\frac{dS}{dt} = \rho\varepsilon - \text{div } F,$$

where S is the entropy and ε the energy production rate per unit mass.

The change of entropy caused by changes of temperature, density, and chemical composition means that we have to write a mass-conservation equation.

Mass-Conservation Equations

— Mass conservation:

$$\frac{\partial\rho}{\partial t} + \nabla \cdot \rho v = 0,$$

where v is the fluid velocity. This is usually small, so that the corresponding inertia terms are completely negligible, and were omitted from the momentum equation, which reduces to the hydrostatic equation.

— Conservation of chemical elements:
We write this symbolically for an element i with mass concentration c_i:

$$\rho\frac{\partial c_i}{\partial t} = \nabla \cdot \rho v_i c_i + \sum_j K_{ij}\rho^2 c_i c_j,$$

where v_i the velocity of the element (due e.g. to diffusion or gravitational settling) and K_{ij} the reaction rate per unit mass of element i with element j, assuming that only two-body reactions occur.

Spherical Symmetry.   The hydrostatic equation shows that pressure is constant on equipotentials (∇P and $\nabla\Phi$ are parallel). The temperature and density are also constant on equipotentials, and a static system has spherical symmetry. We can thus write all the equations in terms of r only:

— Gravity:

$$g = -\frac{GM_r}{r^2}; \tag{3.1}$$

— Mass:

$$dM_r = 4\pi^2\rho\,dr; \tag{3.2}$$

— Hydrostatic equilibrium:

$$dP = -g\rho\,dr; \tag{3.3}$$

— Heat transport:

$$F = -K \frac{\partial T}{\partial r} \, ; \tag{3.4}$$

— Energy: With specific heat ratio γ, and for a perfect gas,

$$\text{div } F = \rho \varepsilon - \frac{1}{\gamma - 1} P \frac{\partial}{\partial t} \log \frac{P}{\rho^\gamma} \, ; \tag{3.5}$$

— Mass conservation:

$$\frac{\partial \rho}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \rho v = 0 \, . \tag{3.6}$$

To solve these equations we need to know the equation of state $P = P(\rho, T, X, Y, Z, ...)$, the thermal conductivity $K = K(\rho, T, X, Y, Z, ...)$, the energy production rate per unit mass $\varepsilon = \varepsilon(\rho, T, X, Y, Z, ...)$, and suitable boundary conditions. At the centre $(r = 0)$ we require finite density, so that the mass has no singularity, and gravity tends to zero there. The total radiative luminosity $\int_0^r 4\pi r^2 \text{div} F \text{d}r$ must vanish at the centre, requiring the temperature gradient $\text{d}T/\text{d}r$ to vanish too. At the surface we often use the approximation $P = \rho = T = 0$. A more correct treatment requires us to match the atmosphere to the interior. Numerically one calculates a model atmosphere and matches it to the interior at a depth such that the outer layers contain a few percent of the total mass.

3.1.3 Comparison with Observation

Observations give us the mass and luminosity of about thirty members of binary systems with reasonable accuracy, with the most accurately measured star being of course the Sun. From parallaxes we can get luminosities for a few hundred stars, the accuracy decreasing with distance (these statements refer of course to the pre-*Hipparcos* era). Observation of star clusters gives Hertzsprung–Russell diagrams which reveal evolutionary effects. Atmospheric chemical compositions may show the imprint of evolution, but also constitute a difficult puzzle.

The differences (dwarfs, giants, supergiants) evident on the HR diagram correspond to different stages of evolution. Binary systems provide important information in the form of the main-sequence *mass–luminosity relation*. It is still the main observational evidence constraining the theory of internal structure (Fig. 3.1).
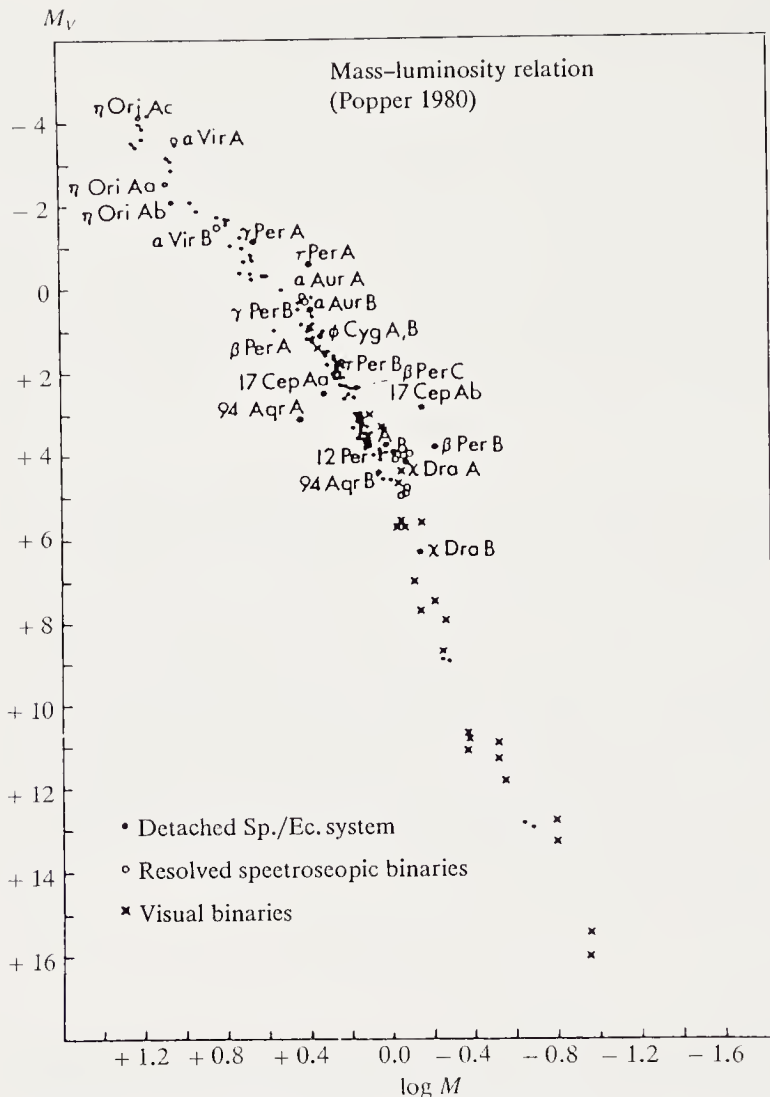
Fig. 3.1. Mass–luminosity relation. Data from the Popper (1980) and McAlister and Hartkopf (1984) lists of binaries. The points ● refer to detached spectroscopic binaries, ○ to resolved spectroscopic binaries, and × to visual binaries. From H.A. McAlister, in *Calibration of Fundamental Stellar Quantities*, ed. by D.S. Hayes et al., Kluwer 1985, p. 99. (Reproduced by kind permission of Kluwer Academic Publishers)

3.2 First Approximations: Orders of Magnitude

Before full numerical solutions by computer became readily available, great efforts went into studying simplified models of stellar structure, if possible analytically. These classical solutions are useful in providing good approximations to the full solutions and give good estimates of internal conditions. Even if not quantitatively exact, they provide an excellent idea of the physical processes at work.

3.2.1 Polytropes

Early results were obtained by avoiding the problem of heat transport and by assuming a relation $P = K\rho^\Gamma$, $\Gamma = $ const. These configurations were called polytropes by Rayleigh. Defining the polytropic index as $n = 1/(\Gamma - 1)$ and setting

$$\rho = \rho_c \theta^n, \qquad r = \alpha\xi, \tag{3.7}$$

where ρ_c is the central density, ξ a dimensionless variable, and α a length scale defined by

$$\alpha^2 = \frac{(n+1)K}{4\pi G}\rho_c^{(1/n)-1}, \tag{3.8}$$

the star's structure is given by the differential equation

$$\frac{1}{\xi^2}\frac{d}{d\xi}\xi^2\frac{d\theta}{d\xi} = -\theta^n. \tag{3.9}$$

The solution must satisfy $dP/dr = 0$ at the centre, implying that $d\theta/d\xi = 0$ for $\xi = 0$. This discards the solution which is singular at the origin, and the conditions $\theta = 1$, $d\theta/d\xi = 0$ define the solution of (3.9) completely. The point where $\theta = 0$, and thus where pressure, temperature, and density vanish, defines the reduced stellar radius ξ_1.

We get the mass M of the polytrope as the integral $\int 4\pi r^2 \rho dr$, and thus

$$M = -4\pi\alpha^3\rho_c\left(\xi^2\frac{d\theta}{d\xi}\right)_{\xi=\xi_1}, \tag{3.10}$$

and the average density $\langle\rho\rangle$ is related to the central density ρ_c by

$$\frac{\langle\rho\rangle}{\rho_c} = -\frac{3}{\xi_1}\left(\frac{d\theta}{d\xi}\right)_{\xi=\xi_1}. \tag{3.11}$$

We also get an expression for the central temperature,

$$T_c = \frac{GM\mu}{\Re R}\frac{1}{-\xi_1(d\theta/d\xi)_{\xi=\xi_1}}, \tag{3.12}$$

when the pressure is that of a perfect gas, $P = \Re\rho T/\mu$, where $\Re$ is the gas constant and μ the mean molecular mass.

Use of the Saha equation shows that under the physical conditions in the interiors of main-sequence stars, matter is highly ionised over most of the star. The mean molecular mass μ is given by counting the free particles in the stellar gas. A good approximation is to assume complete ionisation; this is certainly true of hydrogen and helium in the interior, and introduces negligible error for other elements. Thus we have

Table 3.1. Properties of polytropes

n	ξ_1	$-\xi^2\left(\dfrac{d\theta}{d\xi}\right)_{\xi=\xi_1}$	$\left(-(n+1)\xi_1\left(\dfrac{d\theta}{d\xi}\right)\right)^{-1}_{\xi=\xi_1}$	$(-3/\xi_1)\left(\dfrac{d\theta}{d\xi}\right)_{\xi=\xi_1}$
1	3.14159	3.14159	0.5	3.28987
1.5	3.65375	2.71406	0.53849	5.99071
3	6.89685	2.01824	0.85432	54.1825
3.25	8.01894	1.94980	0.96769	88.153
5	∞	1.73205	∞	∞

$$\frac{1}{\mu} = 2X + \frac{3}{4}Y + \frac{1}{2}Z. \tag{3.13}$$

Estimates of the central density and temperature need the quantities $(-1/\xi_1)(d\theta/d\xi)_{\xi=\xi_1}$ and $-\xi_1(d\theta/d\xi)_{\xi=\xi_1}$. These estimates mark off the region in which we need to make precise estimates of P, K, and ε. Table 3.1 gives the main properties of polytropes of various indices n.

The *Eddington standard model* has particular importance. The total pressure is the sum of gas and radiation pressure:

$$P = P_G + P_R. \tag{3.14}$$

Eddington introduced the simplifying assumption

$$\frac{P_R}{P_R + P_G} = 1 - \beta = \text{const}. \tag{3.15}$$

This implies that $P \sim \rho^{4/3}$ and gives polytropic index $n = 3$. For the Sun this gives the results below as a function of chemical composition ($Z = 0.020$):

X	T_c (Sun)
0.25	14×10^6 K
0.28	13.8×10^6 K

These estimates (very close to the exact results) show that the central temperatures of stars are high enough for them to get energy from nuclear reactions. This gives the nuclear timescale t_N for stellar evolution (Sect. 3.4.2).

3.2.2 The Vogt–Russell Theorem

For stars of uniform chemical composition, an important result, the *Vogt–Russell theorem*, shows that *mass–luminosity* and *mass–radius* relations exist. The system (3.1–4) is one of 4 differential equations with 4 unknown functions of r, i.e $\rho(r), T(r), M(r)$, and $L(r)$. We have to solve them for a specified mass $M(R) = M$ and four other conditions $M(0) = L(0) = 0$ and $\rho(R) = T(R) = 0$. These 5 conditions (including $M(R) = M$) make this

an eigenvalue problem. There is a unique solution, giving $L(M)$ and $R(M)$. This theorem explains the existence of mass–radius and mass–luminosity relations for the main sequence, but fails once the chemical composition depends on radius.

3.3 Stellar Structure: Basic Physics

The basic properties described briefly above show the need to use the most complete and precise physical data in studies of the internal structure.

3.3.1 The Equation of State

Quite precise equations of state are needed for some applications. The temperature and density range is vast, from low densities (e.g. 10^{-5} g cm^{-3}) to nuclear densities ($\rho \simeq 10^{14}$ g cm^{-3}) and temperatures of only a few thousand degrees at the surface of a red giant to hundred of millions of degrees or even more in some short-lived phases.

Pressure Ionisation. When the ionisation degree is calculated using the Saha equation,

$$\frac{N_{i+1,n}N_e}{N_{i,m}} = 2\frac{(2\pi m_e kT)^{3/2}}{h^3}\frac{\sum_n g_{i+1,n}e^{-\chi_n/kT}}{\sum_m g_{i,m}e^{-\chi_m/kT}}e^{-\chi_{\text{ion}}/kT},$$

the summations over states diverge unless a *cutoff* is used to express the fact that the ions are not isolated but in a plasma. The presence of electrons and other ions limits the terms appearing in the state summations. This is called pressure ionisation. We can represent it with varying degrees of refinement. The simplest idea is as follows. An ion of charge je produces an electric field near an ion of charge $(i + 1)e$ (Fig. 3.2). Only excitation levels below the maximum of the perturbed potential can exist, and the state summations are truncated accordingly. More exactly, we have to multiply the Boltzmann probability $\exp(-\chi_0\beta/n^2)$ for the existence of state n by the probability that the state will be disrupted by the nearest ion, $\exp\left\{(-r^3/a^3)\left[1 + [j/(i+1)]^{1/2}\right]^{-3}\right\}$. This corresponds to a state n, defined by

$$\frac{\chi_0}{n^2} = \frac{e^2}{r}\left(\sqrt{i+1} + \sqrt{j}\right)^2$$

and a probability

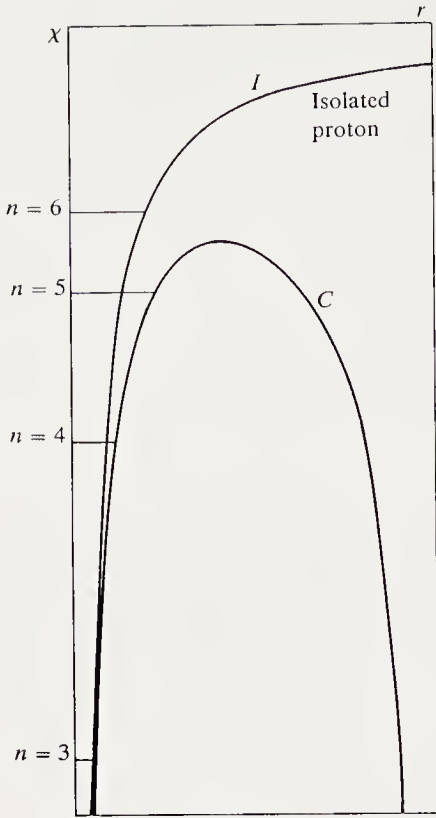$$p_n = \exp\left(-\frac{n^6 e^6}{\chi_0^3}\left(\sqrt{i+1} + \sqrt{j}\right)^6\frac{1}{a_j^3}\right),$$

Fig. 3.2. Electrostatic potential χ of an isolated proton (I) and one in the presence of another ion (C) as a function of their distance r. The energy levels $n = 3, 4, 5$, and 6 are those of an isolated hydrogen atom. Curve C is for a density 4×10^{-6} g cm^{-3} or 2.4×10^{18} atoms cm^{-3}. Levels above $n = 5$ are not populated

where a_j is the average distance of the ion of species j, defined by $(4\pi a_j^3/3)N_j = 1$, where N_j is the number density of ions of charge je. If there are different ionic species present we form the product of the probabilities (assumed independent in a weakly correlated medium), and the exponential contains the summation $\sum \frac{4}{3}\pi N_j \left[(i+1)^{1/2} + j^{1/2}\right]^6$.

As an example, this simplified model of pressure ionisation gives for O^{7+} in a hydrogen–helium medium $X = 0.7, Y = 1 - X = 0.3$ a contribution $\exp(-0.0065n^6\rho)$. The level with principal quantum number $n = 2$ is no longer populated above $\rho = 2.4$ g cm^{-3}, and ionisation is total at about $\rho = 15$ g cm^{-3}.

A full treatment of pressure ionisation requires sophisticated methods of statistical physics.

Electron Degeneracy. The use of Fermi statistics gives the relations below if we ignore relativistic effects (we discuss these later). Defining functions $F_n(\eta)$ by

$$F_n(\eta) = \int_0^\infty \frac{x^n \mathrm{d}x}{\mathrm{e}^{x+\eta} + 1},$$

the electron density, pressure, and internal energy are given by the relations

$$N_e = \frac{4}{\sqrt{\pi}} \frac{(2\pi m_e kT)^{3/2}}{h^3} F_{1/2},$$

$$P_e = \frac{8}{3\pi^{3/2}} \frac{(2\pi m_e kT)^{5/2}}{h^3} F_{3/2},$$

$$U_e = \frac{4}{\pi^{3/2}} \frac{(2\pi m_e kT)^{5/2}}{h^3} F_{3/2},$$

and the total pressure is

$$P = \frac{\Re \rho T}{\mu_A} + P_e + P_R,$$

where μ_A is the mean molecular mass of the *ions* alone, P_e the electron pressure and $P_R = aT^4/3$ the radiation pressure. The degeneracy parameter

$$\eta = -\log 2 \left[N_e h^3 / (2\pi m_e kT)^{3/2} \right]$$

defines the degree of degeneracy of the electron gas. Figure 3.3 shows the curve where electron degeneracy gives a contribution $(\Delta P_e/P_e)_{\deg} = 0.1$, and also the curve $P_R/P = 0.1$. We note that the degeneracy contribution is not negligible at the centre of the Sun.

Collective Effects. In a weakly correlated low-density plasma the Debye wavelength is large compared with the distance between ions. There are two modifications to the equation of state from collective effects.
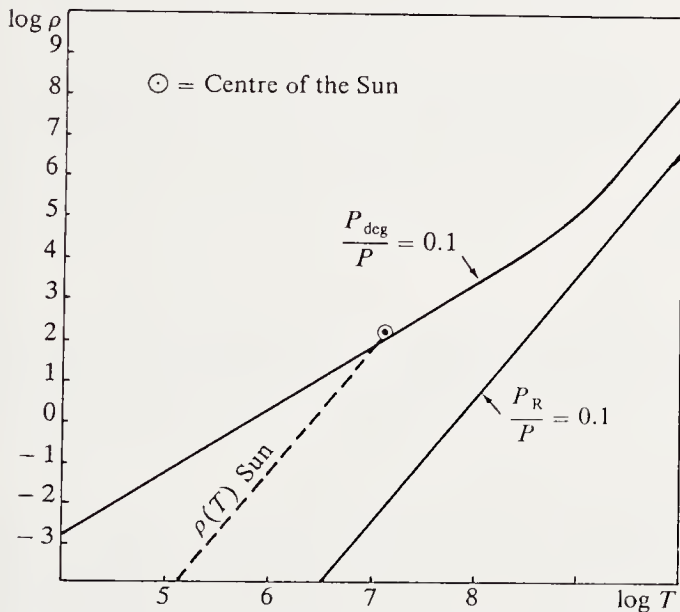

Fig. 3.3. The $(\log T, \log\rho)$ plane. The curves $P_{\deg}/P_G = 0.1, P_R/P_G = 0.1$ are plotted, together with the $\rho(T)$ relation for the central region of the Sun. The electron gas at the centre is weakly degenerate and radiation pressure is negligible

The first correction is due to the electrostatic interaction between charged particles. We have (Landau and Lifschitz, 1959)

$$\delta P_D = -\frac{e^3}{3} \sqrt{\frac{\pi}{kT}} \left(\sum_i \frac{N_i}{V} Z_i^2 \right)^{3/2} ,$$

where N_i is the number of ions of charge Z_i in the volume V.

The second correction comes from the quantum-mechanical interaction between electrons and ions, taking account of bound states of large quantum number and free or scattering states (hyperbolic electron orbits in classical theory). The Planck–Larkin correction for a rarified, almost completely ionised medium, is

$$\delta P_{PL} = 2^{3/2} N_e N_p \lambda_H^3 kT Q_{PL} ,$$

where N_e, N_p are the electron and proton number densities, λ_H the thermal de Broglie wavelength

$$\lambda_H = \left(\frac{\pi \hbar^2}{m_H kT} \right)^{1/2} ,$$

and Q_{PL} is the convergent (Planck–Larkin) sum

$$Q_{PL} = \sum_k g_k \left(e^{+\beta E_k} - 1 - \beta E_k \right) .$$

g_k is the statistical weight of state k, E_k being the ionisation potential. The quantity Q_{PL} is due to Larkin (1960). Its importance comes from the fact that the bound states with energy close to the ionisation potential give a contribution to the total pressure which is compensated by free states (scattering states) of low energy. This modifies the Saha equation, which for pure hydrogen becomes

$$\frac{N_i^* N_e^*}{N_H^*} = 2 \frac{(2\pi m_e kT)^{3/2}}{h^3} \frac{1}{Q_{PL}} .$$

To find the number of hydrogen atoms in state k we should use the Boltzmann formula

$$\frac{N_{Hk}}{N_{H1}} = \frac{g_k}{g_1} \exp(E_k - E_1)$$

and take account of pressure ionisation for high quantum numbers. This becomes important for $n \gtrsim (E_1/e^2 N_i^{1/3})^{1/2}$. This truncation, which only depends on the density, results from a completely different physical process from that assuring the convergence of the sum Q_{PL} for $n \gtrsim E_1/kT$, which depends only on temperature.

The Planck–Larkin formula has not yet been extended to denser media requiring higher-order terms. Astrophysicists use the equation of state obtained by minimising the free energy given by summing over all contributions and calculating the occupation probability for excited states in a more elaborate way than in Sect. 3.1.1. The most important results are those of Mihalas, Hummer, and Däppen (1988) for $\rho < 10^{-2}$ g cm^{-3} and $T < 10^7$ K. The two equations of state agree for low densities, but the free-energy method has the advantage of simultaneously giving the energy-level populations.

3.3.2 Opacity and Thermal Conductivity

In a static medium ($v = 0$) with a temperature gradient, energy is transported by radiation and thermal conductivity. The radiation contribution to the heat transfer is given by the expression for the radiative flux in the quasi-isotropic approximation

$$F_R = -\frac{4}{3}aT^3 \frac{dT}{dr}\frac{c}{\kappa_R \rho} = -K_R \frac{dT}{dr},$$

where κ_R is the *opacity per unit mass* of the material, given throughout the star (except for optically thin parts of the atmosphere) by the *Rosseland mean*, κ_R, with

$$\frac{1}{\kappa_R} = \frac{1}{\kappa + \sigma} = \frac{\displaystyle\int_0^\infty \frac{1}{\kappa_\nu + \sigma_\nu}\frac{\partial B_\nu}{\partial T}\,d\nu}{\displaystyle\int_0^\infty \frac{\partial B_\nu}{\partial T}\,d\nu},$$

where κ_ν is the monochromatic opacity and σ_ν the scattering coefficient (to a first approximation $\sigma = (8\pi/3)(e^2/m_e c^2)^2 N_e = 0.19(1 + X)$.

We must add the contribution from thermal conductivity, K_c, so that the thermal diffusivity is

$$K = \frac{1}{C}(K_R + K_c),$$

where C is the thermal capacity of the medium. One of the characteristic stellar timescales is R^2/K, where R is the radius, which gives the order of magnitude of the heat transfer within the star.

Opacity. Calculations of opacity have been undertaken since the time of Eddington (1926). It is important to remember that it is the central temperature which, via thermonuclear reaction rates, leads to theoretical stellar luminosities.

The main contributions to the opacity are bound–free, bound–bound, and free–free transitions. The first approximation uses absorption coeffi-

cients $\kappa_{j,i,n}(\nu)$ of isolated atoms of species j, ionisation state i, and a quantum state symbolised by n (standing for the full set of quantum numbers specifying the state). We first need the weighting by population p_{jin} of the level n for the atomic species j, i-times ionised, and the abundance N_j of the element j. It is clear that the absorption coefficient $\kappa_{j,i,n}(\nu)$ depends on this abundance. In what we regard as the *usual* case (roughly, main-sequence stars) we take this to be "cosmic" for heavy elements (CNO and above). In practice we choose certain conventional abundances which have to be modified when we consider very old population II stars.

Analytic calculations are of no use in handling the ten or so elements (C, N, O, Ne, Mg, Al, S, Ar, Fe) and numerous atomic levels. The behaviour of the Rosseland mean κ_R is complicated. Apparent jumps in κ_R are caused by the maximum of the weighting function $\partial B_\nu / \partial T$ crossing a discontinuity in the opacity as the wavelength of this maximum changes with temperature. A second effect is due the temperature and density, which together change the level populations.

The second approximation involves collective effects. The electric field of the ions both changes the ionisation potential (similar to that discussed for pressure ionisation) and also, together with collisions, broadens the lines and raises the bound–bound opacity. The implementation of this second approximation is more difficult than it appears at first sight, since we have to change not merely the ionisation potentials but also the atomic wave functions. Now opacity calculations require a knowledge of the atomic dipole moment $\int \psi_i r \psi_f^* dr$ where ψ_i and ψ_f are the initial and final wave functions, and the calculation of wave functions perturbed by collective effects poses difficulties which have not been entirely overcome. There are two methods: one involves calculating the wave function in a sphere having the average volume occupied by an ion, the other involves calculating the wave function in a Coulomb potential screened by a Debye potential. Neither method is completely satisfactory, as they obscure statistical properties of the micro-field that the atom sits in. The micro-field-model method (Frisch and Brissaud 1971) avoids this difficulty and is being systematically applied to line broadening in the framework of the Opacity Project (Seaton).

Thermal Conductivity. Calculating thermal conductivity in a plasma is not difficult, although the screening adopted for the Coulomb potentials requires care. In weakly correlated plasmas the main contribution to the thermal conductivity comes from electrons. With a diffusion coefficient

$$D \cong \frac{1}{3} \frac{1}{\pi} \frac{(kT)^{5/2}}{e^4 N_i \sqrt{m_e}} \frac{1}{\log l_D/a}$$

and $(l_D/a) = \left(\frac{4}{3}\pi N_e\right)^{1/3} (8\pi N_e e^2/kT)^{-1/2}$ where l_D is Debye length, we get the thermal conductivity by multiplying D by the thermal capacity per unit volume $C = 3N_e kT/2$.

For *degenerate electrons* the mean free path is larger because after a collision the electron has to find a free state. This effect of Fermi statistics decreases the cross-section and thus raises the mean free path. Only electrons near the Fermi limit p_F are able to find a free state after a collision. The number of effective electrons is thus equal to the fraction of electrons in the interval $kT/2$, where

$$\frac{N_e(\text{effective})}{N_e} = \frac{4\pi p_F^2 \Delta p_F}{\frac{4}{3}\pi p_F^3} = \frac{3\Delta p_F}{p_F}$$

with

$$p_F \Delta p_F \simeq m_e kT.$$

The mean thermal energy per electron is no longer $3kT/2$ but must be calculated for a degenerate gas. The average energy is $15m_e(kT)^2/p_F^2$ and the mean energy per degree is $Q = 30m_e(k^2T/p_F^2)$. The product DN_eQ is the thermal conductivity.

Estimating the cross-section for electrons of energy $(p_F^2/2m_e)$ and using the probability that the electron finds a free state after the collision, we have

$$\sigma \cong \pi \left(\frac{2m_e e^2}{p_F^2}\right)^2 \frac{m_e kT}{p_F^2},$$

and finally the order of magnitude

$$K_{\text{cond}} = \frac{3}{32\,\pi^2} h^3 \frac{\rho}{Am_H} \frac{k^2 T}{m_e^2 e^4},$$

or

$$K_{\text{cond}} = 6.4 \times 10^{14} \frac{(\rho/10^5)}{A_i} T_7,$$

which agrees with more refined treatments of thermal conductivity ($T_7 = (T/10^7 \text{ K})$).

Approximate Representation of the Opacity. The radiative opacity written as an interpolation formula

$$\kappa_R = \kappa_0(1 + X)Z\rho^\alpha T^{-3-s}$$

with $\alpha \simeq 1$ and $s \simeq 0.5$ is known as *Kramers' law*. This form is often used.

3.3.3 Thermonuclear Reactions

The idea that the energy sources of the stars are nuclear in origin was first clearly stated by H. N. Russell (1919). The first calculations, due to

Atkinson (1931), used the results of Gamow (1928) on barrier penetration. Generally, if N_i, N_j are the number densities of nuclei of species i and j, σ_{ij} the cross-section, and v_{ij} the relative velocities of two nuclei i, j, the number of reactions per unit volume and time is

$$r_{ij} = N_i N_j \langle \sigma_{ij} v_{ij} \rangle \frac{1}{1 + \delta_{ij}},$$

where δ_{ij} is the Kronecker symbol ($\delta_{ii} = 1, \delta_{ij} = 0, i \neq j$). The calculation of $\langle \sigma_{ij} v_{ij} \rangle$ assumes Maxwellian velocity distributions and requires a knowledge of the cross-section σ_{ij} as a function of energy.

Generally we have to consider the following:

— neutron capture (important only for late stages of evolution);
— charged particle captures (protons, alpha particles); the calculation of $\langle \sigma v \rangle$ differs according to whether we consider a resonant or non-resonant reaction (i.e far from resonance);
— beta captures: reactions $^Z_A X + e^- \rightarrow ^{Z-1}_A X + \nu$.

We will examine the various reactions in connection with the evolutionary phases where they occur.

Reaction Rates. Between charged nuclei the main phenomenon is the crossing of the potential barrier. The cross-section thus has the form $S(\pi M)^2 \exp\left(-2 \int \frac{1}{\hbar} (2M|E - V|)^{1/2} \mathrm{d}x\right)$. For a Coulomb potential, $V = (Z_1 Z_2 e^2 / r)$, and $E = (Z_1 Z_2 e^2 / r_0)$ is the kinetic energy. We can write the cross-section as

$$\sigma(E) = \frac{S(E)}{E} \exp(-(E_G/E)^{1/2}),$$

where E_G is the Gamow energy

$$E_G = \left(\frac{2\pi e^2 Z_1 Z_2}{\hbar}\right)^2 \frac{M}{2},$$

where $M = M_1 M_2 / (M_1 + M_2)$ is the reduced mass. $S(E)$ is a slowly varying function of energy which is determined experimentally where possible.

The Maxwellian energy distribution $\exp(-E/kT)$ appears in the calculation of the mean value $\langle \sigma v \rangle$. The product $\left\{\exp(-E/kT)\exp\left[-(E_G/E)^{1/2}\right]\right\}$ has a well-defined maximum for

$$E_{\max} = \left(\frac{k^2 T^2 E_G}{2}\right)^{1/3}.$$

The width of this maximum is

$$\Delta E = \frac{2^{1/3}}{3^{1/4}} (E_G k^5 T^5)^{1/6}.$$

104

It is always small compared with $E_{\max}$ and the function in square brackets above can be replaced by a Gaussian to a very good approximation. The main term in the reaction rate is then

$$
\begin{aligned}
r_{ij} &= \frac{N_i N_j \langle \sigma v \rangle}{1 + \delta_{ij}} \\
&= \frac{N_i N_j}{1 + \delta_{ij}} \left(\frac{2}{M} \right)^{1/2} \frac{\Delta E_0}{(kT)^{3/2}} S_{\text{eff}} \exp \left[-\frac{3E_0}{kT} \right] ,
\end{aligned}
$$

where

$$
E_0 = \left[\pi \alpha Z_1 Z_2 kT \left(\frac{Mc^2}{2} \right)^{1/2} \right]^{2/3} = E_{\max}
$$

is the Gamow peak, and

$$
\Delta E_0 = 4(E_0 kT/3)^{1/2}
$$

is the width of the peak;

$$
E_0 = 0.1220(Z_1^2 Z_2^2 A)^{1/3} T_9^{2/3} \,\text{MeV} ,
$$

$$
\Delta E_0 = 0.2368(Z_1^2 Z_2^2 A)^{1/6} T_9^{5/6} \,\text{MeV} ,
$$

where A is the reduced atomic mass. The deviation from a Gaussian introduces a factor $\left(1 + \frac{5}{12} \frac{kT}{3E_0} - \frac{35}{288} \left(\frac{kT}{3E_0} \right)^2 \ldots \right)$, which is always close to unity.

Collective Effects. For a nucleus in a plasma the potential barrier is no longer described by a Coulomb barrier. For a weakly correlated plasma the potential of an ion of charge Z_i is described by a Debye potential and the interaction energy W is given by

$$
W_{ij} = \frac{Z_i Z_j e^2}{r} \exp \left(\frac{r}{l_D} \right) ,
$$

where l_D is the Debye length, here given by the classical relation

$$
l_D^{-2} = \frac{4\pi e^2 \rho}{m_H kT} \left[\sum_i \frac{X_i Z_i^2}{A_i} + \sum_i \frac{X_i Z_i}{A_i} \right] .
$$

When the Debye length is large we can expand the exponential for small distances to first order, giving

$$
W_{ij} = \frac{Z_i Z_j e^2}{r} - \frac{Z_i Z_j e^2}{l_D} .
$$

In the expression for the cross-section the energy E is replaced by $E + Z_i Z_j e^2 / l_D$. Changing variables in the calculation of $\langle \sigma v \rangle$ yields the factor

$$A = \exp \left(\frac{Z_i Z_j e^2}{l_D \, kT} \right) > 1 \, ,$$

which is an acceleration factor for the nuclear reaction rate. As an example, at the centre of the Sun, with $X = 0.7, Y = 0.3$ (and neglecting other ions), for the proton–proton reaction

$$A = 1.053 \, .$$

If the plasma is not weakly correlated, the main term in the acceleration of the reaction rate, analogous to $(Z_i Z_j e^2 / l_D kT)$, comes from taking the limit as $r \to 0$ of the autocorrelation function (De Witt et al. 1973):

$$g_{12} = \exp \left[-\frac{Z_1 Z_2 e^2}{r kT} + H_{12}(r) \right] \, ,$$

and the classical limit is thus (with $\beta = 1/kT$)

$$f = \exp[H_{12}(0)] = \lim_{r \to 0} \left[g_{12}(r) \exp \left(\frac{Z_1 Z_2 \beta e^2}{r} \right) \right] \, .$$

For a nuclear reaction between two species of charges Z_1, Z_2 giving a composite nucleus of charge $(Z_1 + Z_2)$, we have

$$\exp[H_{12}(0)] = \exp \left(-\frac{1}{kT} \left[\mu(Z_1 + Z_2) - \mu(Z_1) - \mu(Z_2) \right] \right) \, ,$$

where $\mu(Z_i)$ is the chemical potential of the relevant species, taking account of autocorrelation effects. The calculation of these functions for strongly correlated plasmas is given in the literature.

Resonant Reaction. In this case the cross-section is given by the Breit–Wigner formula. If we denote the reaction $i(j, k)l$, the cross-section is

$$\sigma(E) = \pi \lambda^2 \frac{\omega \Gamma_{ij} \Gamma_{kl}}{(E - E_R)^2 + (\Gamma^2 / 4)} \, ,$$

where λ is the de Broglie wavelength for the reduced mass M, $\omega = (2J + 1)/[(2J_i + 1)(2J_j + 1)]$ the statistical factor, in which J is the angular momentum of the composite nucleus, J_i and J_j the angular momenta of the incident particles i and j, and Γ the full width of the resonant level, of energy E_R. If we integrate over a Maxwellian velocity distribution and the resonant energy is much greater than the energy of the Gamow peak, the reaction rate becomes

$$r_{ij} = \frac{N_i N_j}{(1 + \delta_{ij})} \langle \sigma v \rangle$$

$$= \frac{N_i N_j}{(1 + \delta_{ij})} \left(\frac{2}{M\,kT} \right)^{3/2} \sqrt{\pi}\, \hbar^2 \omega \frac{\Gamma_{ij}\Gamma_{kl}}{\Gamma} \exp(-E_R/kT).$$

Obviously each reaction must be examined in detail to see what levels are involved and what is their width. The basic data are from experimental nuclear physics.

3.3.4 Convection

The condition for stability of radiative equilibrium is another basic physical datum, this time macroscopic. If this condition is violated, a convective zone will appear (sometimes a semi-convective zone if there is a gradient of chemical composition). The physics of convection is discussed in Chap. 5.

The standard description of convection is called *mixing-length theory*. This description completely ignores the fact that the boundary described by $\nabla_{\text{ad}} = \nabla_{\text{rad}}$ can be crossed by convective motions. In reality v does not vanish at this boundary and this gives *convective overshooting*, which we consider in Chap. 5.

Example: The Main Sequence. In models of chemically homogeneous (ZAMS) stars the extent and location of convective zones depend on the mass and chemical composition. As an example we consider here a reference model with mass fractions X, Y, Z of hydrogen, helium, and other elements given by 0.71, 0.27 and 0.02. (This composition is called the Russell mixture – see Chap. 2). Figure 3.4 shows the fraction of the mass of the star occupied by one or two convective zones as a function of total mass. Between 0.36
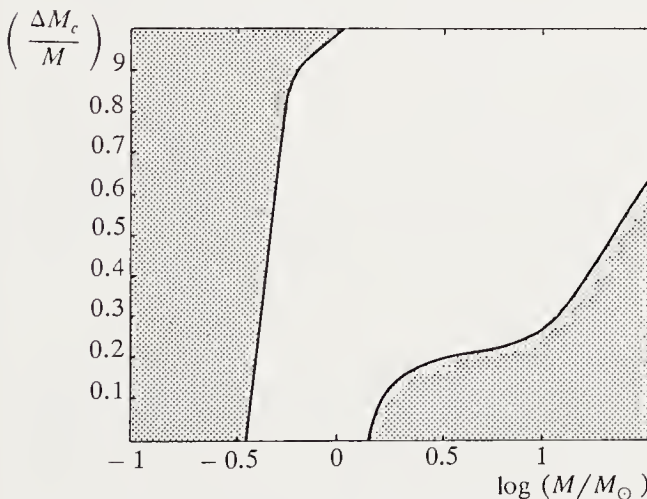

Fig. 3.4. Convective zones. The abscissa shows the mass (logarithmic scale) and the ordinate the convective mass fraction, for main-sequence stars of population I. Below 0.3 $M_\odot$ the stars are fully convective. Above 1.2 $M_\odot$ a convective core appears

and $1.4\,M_\odot$ the stars do not have convective cores. Below about $1.2\,M_\odot$ the outer convective zone extends deeper and deeper into the star as the mass decreases, until all of the star apart from the radiative atmosphere is convective (at about $0.3\,M_\odot$).
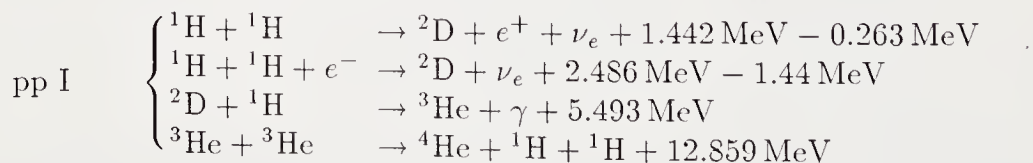
3.4 Stellar Structure

The study of the internal structure of stars is closely linked to that of their evolution. Two major irreversible processes govern the evolution: the change of chemical composition in nuclear reactions, and mass loss, which at certain evolutionary phases can have dramatic importance. Stellar-structure theory begins with the assumption that stars are chemically homogeneous at their formation. This assumption is based on modelling of the pre-main-sequence phase, in which the forming star has not yet ignited thermonuclear reactions and can release energy only by contracting under gravity. These stars are strongly unstable according to the Schwarzschild criterion and are fully convective. It is then reasonable to assume that convective mixing is violent enough to ensure chemical homogeneity. With the one exception of deuterium, which is destroyed above 600 000 K in the reaction $D^2(p,\gamma)\,{}^3He$ (see Sect. 3.4.1), the star arrives on the main sequence with the same composition as the interstellar cloud from which it was born.
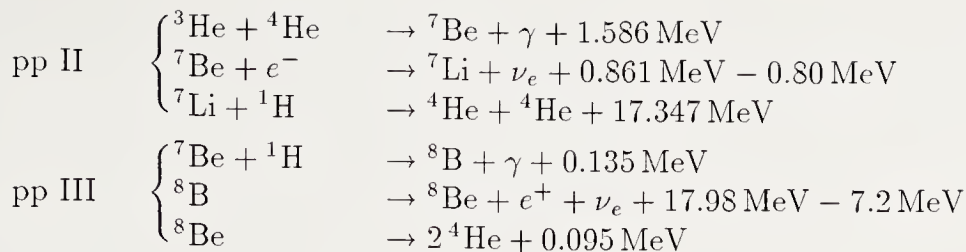
Such homogeneous stars are then at the beginning of their evolution (ZAMS), and we devote the first part of this study to them. First classified by Russell in 1915, the *white dwarfs* represent an endpoint of stellar evolution. We can understand them by use of the equation of state of a degenerate gas, even before studying the evolutionary processes giving rise to them. Stellar-evolution theory makes constant reference to white dwarfs, so we shall discuss their structure at the outset. We treat stellar-evolution theory from Sect. 3.5 onwards.

3.4.1 Thermonuclear Reactions

We consider here the dependence of energy production rates on physical conditions. As we are here concerned with the main sequence we discuss only the two cycles of reactions relevant to it for the moment.

The pp Chain. The reaction networks are

$$
\text{pp I} \quad
\begin{cases}
{}^1\text{H} + {}^1\text{H} & \rightarrow {}^2\text{D} + e^+ + \nu_e + 1.442\,\text{MeV} - 0.263\,\text{MeV} \\
{}^1\text{H} + {}^1\text{H} + e^- & \rightarrow {}^2\text{D} + \nu_e + 2.486\,\text{MeV} - 1.44\,\text{MeV} \\
{}^2\text{D} + {}^1\text{H} & \rightarrow {}^3\text{He} + \gamma + 5.493\,\text{MeV} \\
{}^3\text{He} + {}^3\text{He} & \rightarrow {}^4\text{He} + {}^1\text{H} + {}^1\text{H} + 12.859\,\text{MeV}
\end{cases}
$$

$$\text{pp II} \quad \begin{cases} {}^3\text{He} + {}^4\text{He} & \to {}^7\text{Be} + \gamma + 1.586\,\text{MeV} \\ {}^7\text{Be} + e^- & \to {}^7\text{Li} + \nu_e + 0.861\,\text{MeV} - 0.80\,\text{MeV} \\ {}^7\text{Li} + {}^1\text{H} & \to {}^4\text{He} + {}^4\text{He} + 17.347\,\text{MeV} \end{cases}$$

$$\text{pp III} \quad \begin{cases} {}^7\text{Be} + {}^1\text{H} & \to {}^8\text{B} + \gamma + 0.135\,\text{MeV} \\ {}^8\text{B} & \to {}^8\text{Be} + e^+ + \nu_e + 17.98\,\text{MeV} - 7.2\,\text{MeV} \\ {}^8\text{Be} & \to 2\,{}^4\text{He} + 0.095\,\text{MeV} \end{cases}$$
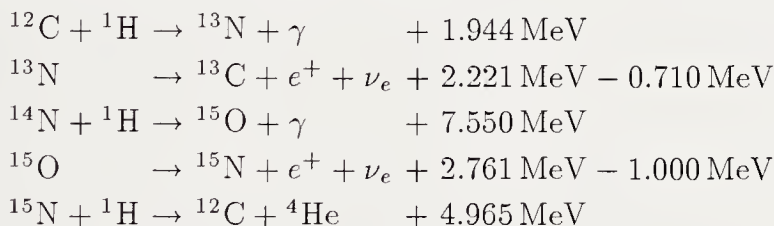
The energy liberated includes the annihilation energy of the positron; the energy given with a minus sign is the average energy lost as neutrinos. As we shall see, energy production by this sequence of reactions dominates up to about $1.3\,M_\odot$. For lower-mass stars the central temperatures are too low for the reactions to proceed beyond the formation of ${}^3\text{He}$.

In the Sun, the reaction network labelled pp I contributes about 85 % of the luminosity, group pp II about 15 %, and pp III only about 0.015 %. The slowest reaction is ${}^1\text{H}(p, e^+\nu_e)^2\text{D}$. If the energy production rate is represented as a power law, $\varepsilon \sim T^n$, the exponent n is given by
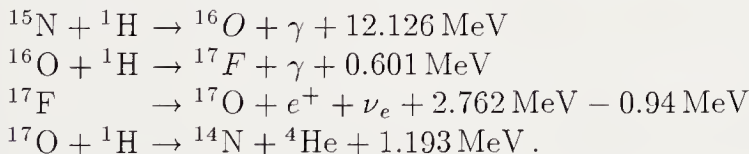
$$n = \frac{\tau}{3} - \frac{2}{3}.$$

For the Sun, with $\tau = (33.3/T_6^{1/3})$ we have for $T_6 = 15$, $n = 3.83$, and the mean liberated energy per reacting proton is 6.541 MeV.

The CNO Cycle. For stars with masses above $1.2\,M_\odot$ the carbon cycle dominates. The reaction network is

$$\begin{aligned} {}^{12}\text{C} + {}^1\text{H} &\to {}^{13}\text{N} + \gamma & &+ 1.944\,\text{MeV} \\ {}^{13}\text{N} &\to {}^{13}\text{C} + e^+ + \nu_e & &+ 2.221\,\text{MeV} - 0.710\,\text{MeV} \\ {}^{14}\text{N} + {}^1\text{H} &\to {}^{15}\text{O} + \gamma & &+ 7.550\,\text{MeV} \\ {}^{15}\text{O} &\to {}^{15}\text{N} + e^+ + \nu_e & &+ 2.761\,\text{MeV} - 1.000\,\text{MeV} \\ {}^{15}\text{N} + {}^1\text{H} &\to {}^{12}\text{C} + {}^4\text{He} & &+ 4.965\,\text{MeV} \end{aligned}$$

(von Weizsäcker 1938; Bethe 1939). Supplementary proton captures can occur:

$$\begin{aligned} {}^{15}\text{N} + {}^1\text{H} &\to {}^{16}O + \gamma + 12.126\,\text{MeV} \\ {}^{16}O + {}^1\text{H} &\to {}^{17}F + \gamma + 0.601\,\text{MeV} \\ {}^{17}F &\to {}^{17}O + e^+ + \nu_e + 2.762\,\text{MeV} - 0.94\,\text{MeV} \\ {}^{17}O + {}^1\text{H} &\to {}^{14}\text{N} + {}^4\text{He} + 1.193\,\text{MeV}. \end{aligned}$$

The slowest reaction of the CNO cycle is ${}^{14}\text{N}(p,\gamma)^{15}\text{O}$. This fixes the reaction rate. The branching ratio between the two ${}^{15}\text{N}$ reactions giving ${}^{12}\text{C} + {}^4\text{He}$ and ${}^{16}\text{O}$ is $1/1200$ and gives weak oxygen production. Carbon reappears at the end of the cycle and behaves like a catalyst.

A power-law representation, with $\tau = 151.92 T_6^{-1/3}$, gives $n = 19.7$ for $T_6 = 15$. Comparing the two approximate expressions

$$\varepsilon_{pp} = 27 \left(\frac{\rho}{100}\right) X^2 \left(\frac{T}{14 \times 10^6}\right)^4 \text{ erg g}^{-1}\text{ s}^{-1}$$

and

$$\varepsilon_{CN} = 20 \left(\frac{\rho X_H}{100}\right) \left(\frac{Z_{CN}}{0.01}\right) \left(\frac{T}{15 \times 10^6}\right)^{20} \text{ erg g}^{-1}\text{ s}^{-1},$$

we see that the energy production rates are equal at about $(M/M_\odot) = 1.2$ on the main sequence.

3.4.2 The Nuclear Timescale

This is the characteristic time spent on the main sequence. For stars between about 0.3 and 10 $M_\odot$ it is reasonable to take this as the time to exhaust the central hydrogen content. If Q is the energy production per reaction and γ_N the ratio $\varepsilon_{\text{centre}}/\langle\varepsilon\rangle$, the characteristic time is

$$t_N = \frac{M}{L} Q \frac{X_1}{\gamma_N},$$

where X_1 is the initial hydrogen content. Transforming hydrogen into helium via the CNO cycle produces about 25 MeV. The energy from the pp and related reactions depends on the temperature. If the $^3\text{He}(^3\text{He},2p)^4\text{He}$ reaction dominates, the yield is 13.1 MeV per pp reaction. At higher temperatures, where pp II dominates, the yield per pp reaction is 25.7 MeV. Finally, when pp III dominates, neutrino losses result in a yield of 19.1 MeV. However, on the main sequence the carbon cycle takes over before pp II comes into operation. For an $n = 3$ polytrope and a reaction rate $\propto T^n$, we have $\gamma_N(n = 4) = 9.75$ and $\gamma_N(n = 16) = 3.10$. In the carbon cycle $Q = 0.0268$ mass units per reaction, and for pp I, $Q = 0.0140$ mass units per reaction, the energy yield per reaction being Qc^2.

For the Sun, $t_N = 15.8 \times 10^9$ y. Using the mass–luminosity relation (see Sect. 3.4.4) we find for stars with masses $M < 1.4 M_\odot$

$$t_{pp}^{MS} \cong 1.6 \times 10^{10}(M/M_\odot)^{-2} \text{ years}$$

and for $M > 1.2 M_\odot$

$$t_{CN}^{MS} \simeq 10^{10}(M/M_\odot)^{-3} \text{ years}.$$

3.4.3 Homology

The idea of *homology* is that near a given mass, stellar models obey scaling laws involving powers of the physical quantities $(L, M, R, \kappa_0, \varepsilon_0, \mu)$. This is essentially a form of dimensional analysis.

Even though homology results are only approximate and cannot in any way replace full numerical solutions of the structure equations, they give

valuable insight into the physical meaning of the solutions. They are mainly applied to chemically homogeneous (i.e. ZAMS) stars.

The equations of hydrostatic equilibrium, transfer, mass conservation, and nuclear luminosity become proportionality relations:

$$P \sim \frac{M^2}{R^4}, \qquad L \sim \mu^{7.5} M^{7+s-\alpha} R^{3\alpha-7-s} \kappa_0^{-1},$$

with opacity $\kappa = \kappa_0 \rho^\alpha T^{-3-s}$,

$$\rho \sim (M/R^3), \qquad L \sim \varepsilon_0 \mu^n M^{n+2} R^{-3-n},$$

n being the exponent of T in the thermonuclear reaction rate, expressed as a power law. Eliminating R gives a mass–luminosity relation. The agreement is reasonable for spectral types from B to G on the main sequence. It is poor for low and high-mass stars, where large parts of the structure are dominated by convection and the energy transport is not radiative as assumed above.

Conversely, we can try to use the observed $L(R), R(M)$ relations to find the exponent n in the nuclear-energy generation rate. For spectral types G4 to M, with $L = 0.41 M^{2.3}$ and $R = M^{0.5}$, we find $n = 3.6$; for spectral types B–G , with $L = 1.12 M^{3.9}, R = M^{0.75}$, we find $n = 16.6$. These results can be seen as confirming the thermonuclear nature of main-sequence energy sources.

3.4.4 The Mass–Luminosity Relation

Anticipating a result of evolutionary theory, we note that stars spend a long time very close to their starting point on the HR diagram. For a given spectral type this defines an interval of about one magnitude in luminosity and constitutes the *main sequence*. The stars for which we know L, R, and M simultaneously are not necessarily zero-age stars if $M \gtrsim 0.8\, M_\odot$, even if they are on the main sequence. It is clear that there is little possibility of detailed observational checks of the theory in this case. For masses less than about $0.8\, M_\odot$, however, the timescale t_N for evolution is appreciably longer than the age of the Galaxy.

Despite these restrictions, the quantitative predictions of the ZAMS $L(M), R(M)$ relations constitute one of the most important tests of stellar-structure theory and a primary method of determining cosmic abundances. Comparison of evolutionary predictions with the data for star clusters gives a second major test of the theory.

Low-Mass Stars. Stellar models using the standard treatment of convection and the best theoretical opacities and equations of state are in good agreement with observation (Fig. 3.5). The models have conventional population I composition ($X = 0.73, Y = 0.25, Z = 0.02$).

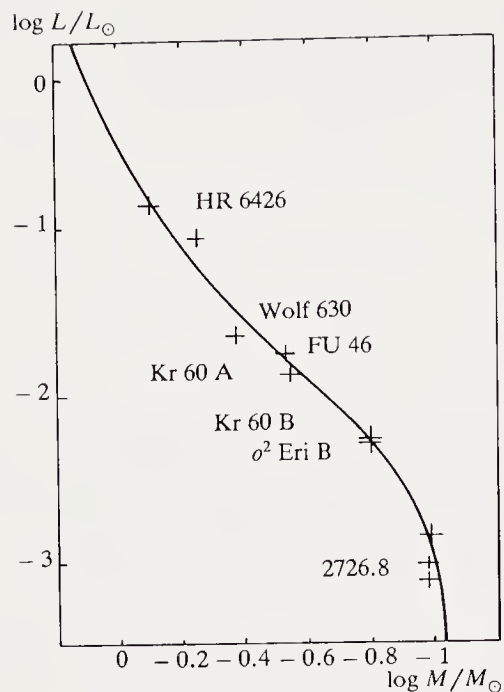**Fig. 3.5.** Mass–luminosity relation for low-mass stars. The curve gives the theoretical relation of H.C. Graboske et al. (1973)

Stars of Intermediate Mass. We mean here stars of 1–$10\,M_\odot$. They are dominated by radiative regions but have small convective cores which grow with mass and surface convection zones which disappear above about $2\,M_\odot$.

With a suitable power-law representation of the opacity, homology relations give a reasonably accurate mass–luminosity relation. Within the present errors on $L, R,$ and M, stellar-structure theory accounts well for the observed ZAMS for intermediate-mass stars (Fig. 3.6). Forthcoming data (e.g. stellar seismology, *Hipparcos* data) will provide more stringent tests.
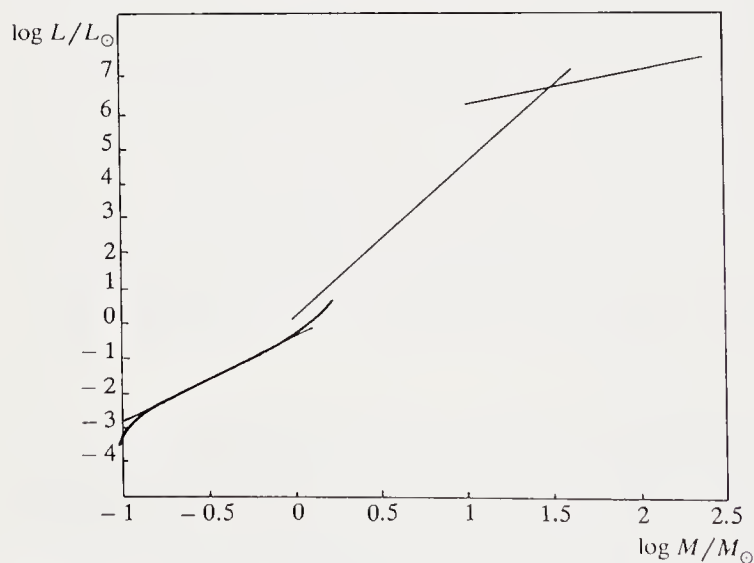

Fig. 3.6. Mass–luminosity relation for the whole main sequence. There are three homology domains $(0.2\,M_\odot < M < 1\,M_\odot; 1\,M_\odot < M < 30\,M_\odot; 30\,M_\odot < M)$

High-Mass Stars. High-mass stars are almost entirely convective and dominated by radiation pressure. The mass–luminosity relation follows from the condition for convective equilibrium in the form $\nabla_{\text{ad}} = \nabla_{\text{rad}}$:

$$\frac{1}{16\pi c} \frac{P}{P_R} \frac{\bar{\kappa}L}{GM} = \frac{1}{4},$$

where $\bar{\kappa} = 0.19(1 + X)$ is the opacity, which here reduces to electron scattering. With $P = P_R$ we get

$$(L/L_\odot) \simeq 4.6 \times 10^4 (M/M_\odot).$$

The relation $L \sim M$ is approximately correct (Fig. 3.6), but high-mass stars have such short lifetimes ($t_N = 6 \times 10^5/\gamma_N$ years) that the concept of the ZAMS is scarcely meaningful.

3.4.5 White Dwarfs

Russell (1915) pointed out the existence of a peculiar class of stars of faint magnitude and spectral type similar to A. The explanation of their structure had to await the discovery of the equation of state of degenerate matter (1924) as an application of Fermi–Dirac statistics. White dwarfs have been the objects of systematic study (number counts, luminosity function, spectral classification) since Russell's discovery.

There are essentially two classes of white dwarfs: DA, with pure hydrogen envelopes, and DB, having little or no hydrogen in their atmospheres and possibly some trace heavy elements. White dwarfs in general are one of the endpoints of stellar evolution, the others being neutron stars and black holes. White dwarfs provide the classic application of the equation of state of a degenerate electron gas. Because reference to white dwarfs occurs constantly in every discussion of stellar evolution we shall treat them as a parallel to main-sequence stars, where the first approximation to the pressure is provided by adding perfect gas and radiation.

Zero-Temperature Stars. For the moment we ignore the white dwarf's atmospheric envelope (which produces its spectrum). Then the equation of state gives a mass–radius relation. Degeneracy dominates even quite close to the surface, and the size of the non-degenerate envelope is negligible compared with the stellar radius. If p_F is the radius of the Fermi sphere in electron momentum space at zero temperature, setting

$$x = \frac{p_F}{m_e c}$$

gives relations defining the equation of state at $T = 0$:

$$\rho = \frac{8\pi p_F^3}{3h^3} \frac{A}{Z} M_0 = 9.75 \times 10^5 \mu_e x^3 \text{ g cm}^{-3} = Bx^3,$$

$$P = \frac{\pi m_e^4 c^5}{3h^3} f(x) = 6.01 \times 10^{22} f(x) = Af(x)$$

with

$\mu_e = (A/Z),$

$f(x) = x(2x^2 - 3)(x^2 + 1)^{1/2} + 3 \arcsin x.$

Using

$$\frac{df}{dr} = 8x^4(1+x^2)^{-1/2} \frac{dx}{dr}$$

and setting

$y^2 = x^2 + 1,$

$y = y_0 \psi \quad (\psi = 1$ at the centre of the star$),$

$\alpha = \left(\frac{2A}{\pi G}\right)^{1/2} \cdot \frac{1}{By_0}, \qquad G$ being the gravitation constant

$r = \alpha\eta,$

we get

$$\frac{1}{\eta^2} \frac{\partial}{\partial \eta} \left(\eta^2 \frac{d\psi}{d\eta}\right) = -\left(\psi^2 - \frac{1}{y_0^2}\right)^{3/2}.$$

The star's surface is defined by $x = 0$, or $\psi = (1/y_0)$, giving the *mass–radius* relation for white dwarfs (Fig. 3.7). There are two asymptotic regimes. For y_0 close to 1, or $x_0 \ll 1$, the structure is that of an $n = 3/2$ polytrope, with

$x = x_v \psi, \quad (\psi = 1$ at the centre$),$

$\alpha = \left(\frac{A}{\pi G}\right)^{1/2} \frac{1}{Bx_0^{1/2}},$

$r = \alpha\eta,$

and

$$\frac{1}{\eta^2} \frac{d}{d\eta} \left(\eta^2 \frac{d\psi}{d\eta}\right) = -\psi^{3/2},$$

which gives a mass–radius relation (for $\mu_e = 2$)

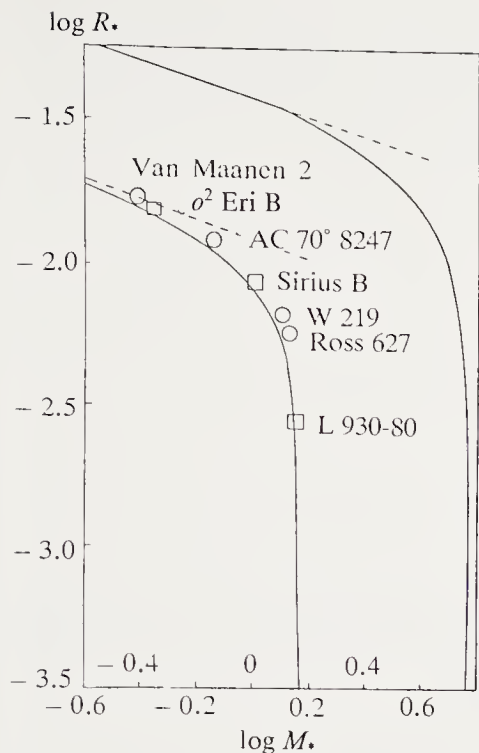$$\frac{R}{R_\odot} = 0.01275 \left(\frac{M}{M_\odot}\right)^{-1/3}.$$

Fig. 3.7. Mass–radius relation for fully degenerate white dwarfs. The curves are for $(A/Z) = \mu_e = 1$ and 2. (From E. Schatzman, *White Dwarfs*, North Holland Amsterdam 1958)

When y_0 approaches infinity (very dense stars) the star becomes a polytrope of index 3. In this limit the mass is independent of the radius, which itself tends to zero:

$$ M = \left(\frac{2A}{\pi G} \right)^{3/2} \frac{1}{B^2} \left(-\xi^2 \frac{d\psi}{d\xi} \right)_S = \frac{5.76}{\mu_e^2} M_\odot \,. $$

This mass (which takes no account of general relativistic effects, Chap. 8) is called the *Chandrasekhar limiting mass*. For $\mu_e = 2$, $M = 1.44\,M_\odot$. The existence of this limiting mass is extremely important. It means that any degenerate star of mass above $1.44\,M_\odot$ must collapse, in practice towards neutron-star or black-hole configurations.

White dwarfs are below the main sequence in the Hertzsprung–Russell diagram (Fig. 1.10). Figure 3.8 shows the lines $R = \text{const}$, with an indication of the corresponding mass, using Chandrasekhar's $R(M)$ relation.

Non-zero Temperatures. The temperature of white dwarfs is in reality non-zero, but the degree of degeneracy increases so rapidly towards the interior that we do not need to consider very deep layers before $T = 0$ becomes a very good approximation.

In the outer layers gravitational settling is extremely effective and occurs in a time short compared to the star's age. In a DA white dwarf, hydrogen floats to the surface in the resulting equilibrium. The DB stars have very little hydrogen, and white dwarfs with metal lines have a convec-
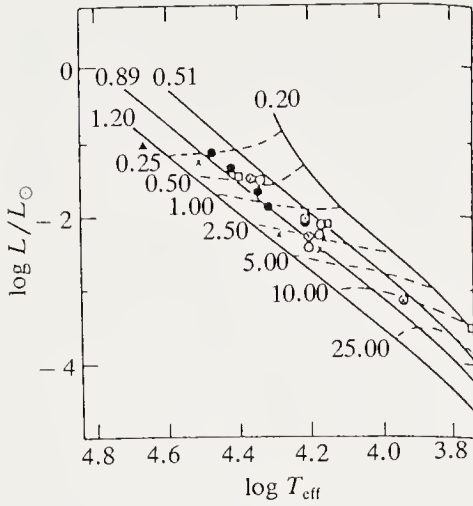
115

Fig. 3.8. Position of white dwarfs on the Hertzsprung–Russell diagram. The curves $R = $ const are shown. (From M.A. Sweeney, *Astron. Astrophys.* **49**, 375 (1976))

tive zone reaching to the degenerate interior, where gravitational settling is ineffective. This dredges up the metals and brings traces to the surface. Gravitational settling is discussed in Chap. 8.

As an example we discuss a white dwarf with a pure hydrogen atmosphere in radiative equilibrium before showing how effective gravitational settling is.

The absorption coefficient follows a Kramers law:

$$\overline{\kappa} = \kappa_0 \rho T^{-3.5}, \quad \kappa_0 = 7.4 \times 10^{22} \text{ cgs}.$$

The outer layers can be taken to have constant gravity:

$$\frac{dP}{dr} = -g\rho,$$

and we approximate the pressure as that of a perfect gas, $P = \Re \rho T / \mu$ down to the point where it must be replaced by that of a non-relativistic degenerate gas, $P = K_1(\rho/\mu_e)^{5/3}$. The heat transfer equation is

$$\frac{d}{dr}\left(\frac{1}{3}aT^4\right) = \frac{-\kappa_0 \rho^2 T^{-3.5} L}{4\pi R^2 c},$$

and we find the solution

$$\rho = \left(\frac{16}{3}\frac{g\mu}{\kappa_0 \Re T_{\text{eff}}^4}\right)^{1/2} T^{3.25}.$$

We assume that degeneracy pressure takes over once it becomes equal to the gas pressure:

$$K_1(\rho/\mu_e)^{5/3} = \Re T.$$

This gives a relation between luminosity and internal temperature for hydrogen-rich (DA) white dwarfs,

116

$$L = \frac{64\pi}{3} \frac{\mu}{\kappa_0 \Re} \left(\frac{K_1}{\Re}\right)^3 \frac{1}{\mu_e^3} T^{3.5} \sigma_{\text{Stefan}} GM \,,$$

and the thickness δr of the non-degenerate layer:

$$\delta r = 4.25 R^2 (\Re/GM\mu)^{9/7} (\Re/K_1)^{6/7} \mu_e^{10/7} \left(\frac{3\kappa_0 L}{64\pi\sigma_{\text{Stefan}}}\right)^{2/7} \,,$$

or numerically,

$$\frac{L}{L_\odot} = 1.2 \times 10^{-6} T_6^{3.5} (M/M_\odot) \,,$$

$$\frac{\delta r}{R} = 8.217 (R/R_\odot)(M/M_\odot)^{9/7}(L/L_\odot)^{2/7} \,.$$

For a typical white dwarf such as Sirius B, $R/R_\odot \simeq 0.01$ and $L/L_\odot \simeq 10^{-3}$, giving the non-degenerate thickness as $\delta r/R \simeq 0.01$ and justifying a posteriori the assumption $g = \text{constant}$ in the outer layers.

In the white dwarf, interior heat is transported by thermal conductivity. In degenerate matter the mean free path of electrons is very long and the conductivity is high; the interior of a white dwarf is essentially isothermal. This gives an estimate of the cooling time t_C. We assume the available internal energy is $Q = M\Re T/\mu_i$ (the ions are a perfect gas: see Chap. 8) so that

$$t_c = 1.76 \times 10^{-3} (L/L_\odot)^{-5/3} 10^6 \text{ years} \,,$$

which is of the order of 8 billion years for $L = 10^{-4} L_\odot$.

3.5 Stellar Evolution (The Standard Model)

We shall denote as the *standard model* the evolutionary model of a non-rotating star with no mass loss and chemical mixing only in unstable regions. This model (which for some evolutionary phases is completely unrealistic) has two main uses; it is the reference model, and it is also of great help in studying models with mass loss.

There are a number of important phases of evolution, and it is important to consider their physical significance rather than simply giving accurate numerical results.

3.5.1 The Beginning of Stellar Evolution

Starting from a uniform chemical composition, the transformation of hydrogen into helium in the central regions forces changes in the star's structure.

To follow these changes we have to study stars over the full range of masses and initial chemical compositions, from population I to extreme population II.

For a star of mass less then $1.25\,M_\odot$, having a radiative core, the hydrogen concentration decreases, and a near-isothermal core gradually forms. If only thermonuclear energy were available, the disappearance of hydrogen would imply $\varepsilon = 0$ and thus zero flux, and hence an isothermal region. In reality we have to allow for entropy production (3.5). To within a fraction of a per cent nevertheless the core becomes isothermal and energy production becomes progressively confined to a thin layer around it.

For stars above $1.25\,M_\odot$ an isothermal core also forms, by a rather different route (core convection dies out as the Schwarzschild criterion fails).

There have been many calculations of the various evolutionary phases, and the results are now well established. The close relation of the various quantities makes it quite difficult to identify the physical causes of the evolution, and such understanding is fairly recent.

Evolution Close to the Main Sequence.   Here we mean evolution during the formation of the isothermal core and its initial growth. In the early stages the core density increases so that the radius r enclosing a given mass m_r shrinks. Hydrostatic balance at the boundary of the isothermal core requires pressure continuity there. In a simplified model where the chemical composition is discontinuous at the boundary, the density must be discontinuous too. For an initial population I composition $X = 0.73, Y = 0.25, Z = 0.02$, the mean molecular mass is $\mu_{fi} = 1.34$ in the core and $\mu_{fe} = 0.60$ outside, a change by a factor 2.23. Simultaneously we move from zero temperature gradient (isothermal case) to one differing from zero. In the simple discontinuous model the exterior temperature gradient is

$$(\nabla T)_{fe} = \left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right)_{fe} = \frac{3}{16\pi acG}\,\frac{\mathcal{R}T_f}{\mu_{fe}}\,\frac{\kappa\rho_{fe}L}{M_rT_f^4} = \frac{1}{n_e+1}\,,$$

where n_e is the local polytropic index (outside the core). The density gradient $\nabla\rho = \mathrm{d}\log\rho/\mathrm{d}\log P$ also has a discontinuity, from $(\nabla\rho)_{fi} = 1$ to $(\nabla\rho)_{fe} = 1 - (\nabla T)_{fe}$. This decrease in the density gradient at the boundary produces an increase in the star's radius. As the mass of the isothermal core grows the gravity at its surface increases and the scaleheight $H_P = -\mathrm{d}r/\mathrm{d}\log P$ decreases. The mass of the exterior layers, reduced in volume by the temperature changes, decreases too. If qM is the core mass, a point arrives when the requirement $M_{\text{core}} = qM = M - M_{\text{ext}}$ can no longer be satisfied. One might try to arrange a larger density at the core boundary in order to provide a suitable mass for the outer layers. But this would require a larger pressure and violate hydrostatic equilibrium. This signifies the start of a new phase of evolution: the contraction of the core on a thermal (Kelvin–Helmholtz) timescale. The transition occurs at $q = 0.13$

for stars of about $1\,M_\odot$, and near $q \simeq 0.10$ for stars above $3\,M_\odot$; the difference is caused by the onset of electron degeneracy in the cores of low-mass stars (Schönberg and Chandrasekhar 1942).

Evolution Towards the Giants.   Core contraction raises the temperature at the boundary radius r_f and thus raises the energy production rate. This higher luminosity and the opacity properties of the outer layers produce an expansion of the star, which evolves towards the giant branch. If radiative equilibrium were maintained the radius would tend to infinity.

Comparison of models of different compositions and opacities shows clearly that it is not the growth of the isothermal core to the Schönberg–Chandrasekhar limit which causes this evolution to the giant branch. We give three examples here to show this:

(1) Models with $(Y, Z) = (0.28, 0.01)$ and M running from 3 to $7\,M_\odot$. These models all evolve similarly. After core hydrogen exhaustion the surface luminosity rises to a maximum and then decreases as the envelope expands. A significant fraction of the thermonuclear energy is used to produce the expansion, which involves a kind of thermal runaway. A similar runaway occurs after core helium exhaustion.

(2) Models with $(Y, Z) = (0.28, 0.001)$ and M running from 3 to $7\,M_\odot$. The $3\,M_\odot$ behaves as above. There is a thermal runaway and core helium burning starts when the star becomes a red giant. In the $5\,M_\odot$ model the thermal runaway starts, but is quenched by core helium burning, and only proceeds once core helium is exhausted. In the $7\,M_\odot$ model there is no runaway until core helium is exhausted.

(3) In models of massive stars with non-conventional opacity, thermal runaway and expansion of the envelope occur even while hydrodgen is still burning in the core.

We define the *critical luminosity* as that for which the thermal runaway begins. We can estimate it by neglecting the core radius R_c in comparison with the star's radius.

Assuming a Kramers law, the conductivity is

$$K = \frac{4acT^3}{3\kappa\rho} = K_0 T^{6.5}\rho^{-2} .$$

This gives a $\rho(T)$ relation

$$\rho = \left[\frac{16\pi}{17} \frac{G\mu K_0}{L} \frac{\mu m_H}{\Re}\right]^{1/2} T^{13/4} .$$

For very large radii there is a power-law solution

$$P = P_c(r/R_c)^{-42/11} ,$$
$$T = T_c(r/R_c)^{-10/11} ,$$

$$\rho = \rho_c(r/R_c)^{-32/11} \, ,$$
$$M = M_c(r/R_c)^{+1/11} \, ,$$

with a mass that diverges very slowly. But for the implied polytropic index of 3.2 one finds

$$\rho_c = \frac{1}{44\pi} \frac{M}{R_c^3} \, ,$$
$$T_c = \frac{11}{42} \frac{GM_c}{R_c} \frac{\mu m_H}{k} \, ,$$

and thus a critical luminosity

$$L = 9.443 K_0 \left(\frac{G\mu m_H}{k} \right)^{15/2} \frac{M_c^{11/2}}{R_c^{1/2}} \, ,$$

or in solar units

$$(L/L_\odot) = 2\,490 \left(\frac{M_c}{M_\odot} \right)^{11/2} \left(\frac{R_\odot}{R_c} \right)^{1/2} \, ,$$

which is within about 5 % of the luminosity given by numerical calculations. Once the mass of the envelope becomes negligible we have the solution

$$T = \frac{4}{17} \frac{GM\mu m_H}{k} \left(\frac{1}{r} - \frac{1}{R_*} \right) \, ,$$

where R_* is the stellar radius. Using the relation for the density, we find for $R_* \to \infty$ a luminosity

$$L_{\max} = 0.615 K_0 \left(\frac{Gm_H\mu}{k} \right)^{15/2} \frac{M^{11/2}}{R_c^{1/2}} \left(\frac{M}{M_c} \right)^2 \, ,$$

or in solar units

$$L_{\max} = 162 \left(\frac{M}{M_\odot} \right)^{11/2} \left(\frac{R_\odot}{R_c} \right)^{1/2} \left(\frac{M}{M_c} \right)^2 L_\odot \, ,$$

which is within about 30 % of the numerical result.

The star actually evolves rapidly to lower temperatures until the extended radiative zone becomes superadiabatic (hydrogen ionisation) and the outer layers are thus convective.

Population I Stars: The Hertzsprung Gap.   In the HR diagram $(B-V, V)$ for a cluster (open or globular), the distribution of the stars outlines the main sequence, the departure from the main sequence, evolution to the giant branch and first ascent of the giant branch. The least massive stars evolve slowly and are still on the main sequence, while the most massive stars in
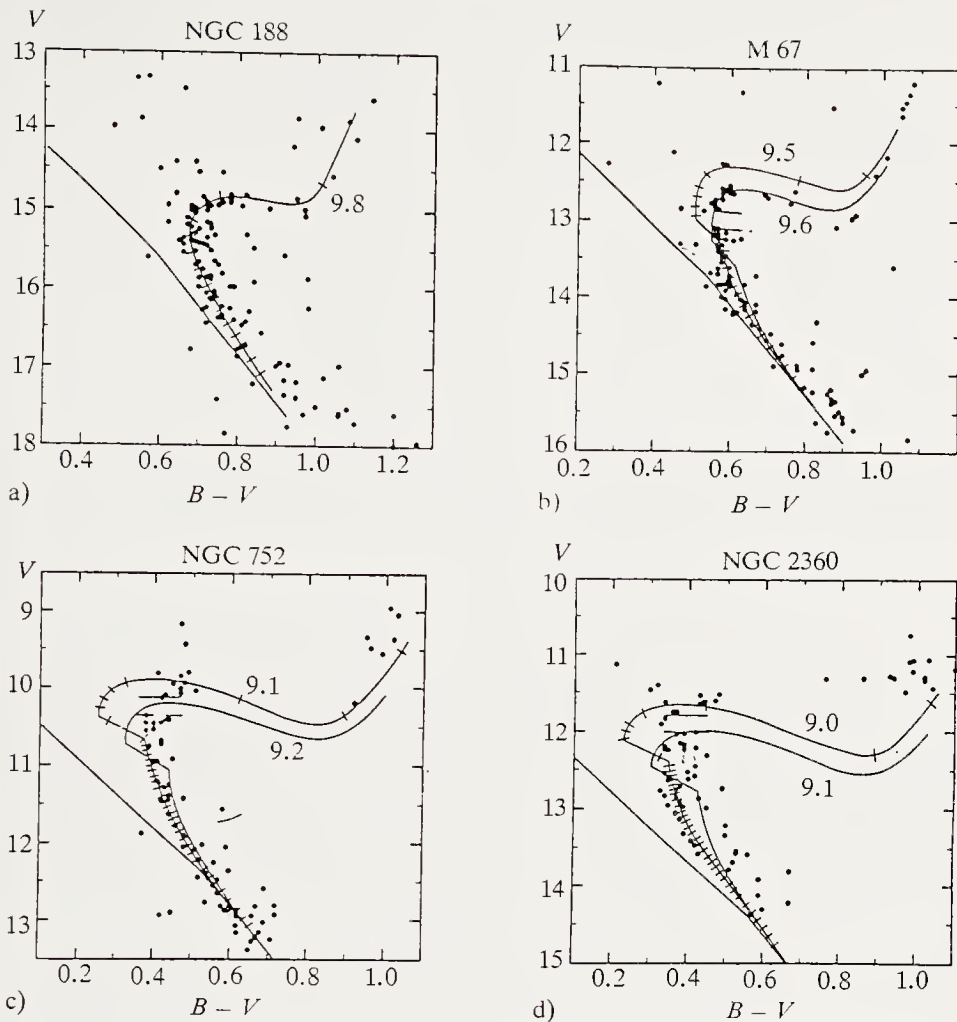
Fig. 3.9. Hertzsprung–Russell (B − V, V) diagram for four clusters: NGC 188, M 67, NGC 752, and NGC 2360. (From A. Maeder, *Astron. Astrophys.* **32**, 177 (1974).) The logarithmic age in years is given against the isochrones. The marks on the isochrones are separated by mass intervals of 0.02 $M_\odot$ and give an idea of the frequency of stars for a constant initial mass function. Data: NGC 188 (Eggen and Sandage 1969); M 67 (Racine 1971), NGC 752 (Grenon and Mermillod 1973), NGC 2360 (Eggen 1969)

the cluster evolve rapidly and have had time to evolve further. Figure 3.9 shows the HR diagrams for four clusters, NGC 188, M 67, NGC 752, and NGC 2360, whose ages are about $6.3 \times 10^9, 3.5 \times 10^9, 1.2 \times 10^9$ and 10^9 y.

The distribution of stars in main-sequence evolution depends simultaneously on the mass distribution along the sequence and the speed of evolution. The density of stars on the diagram is lower for more rapid evolution. At a given age, such as the cluster age t_C, a low-density region implies great sensitivity to small mass differences ($\partial V/\partial M$ or $\partial(B - V)/\partial M$ large)) which is equivalent to rapid evolution. All of the HR diagrams show a zone of avoidance, the *Hertzsprung gap*, corresponding to core hydrogen exhaustion in models and the formation of an isothermal core. The star becomes more luminous and its effective temperature increases.
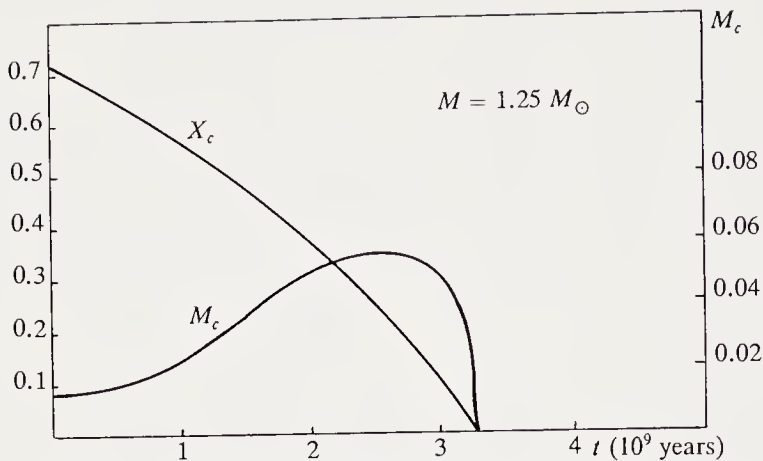
Fig. 3.10. Evolution of the convective core of a 1.25 $M_\odot$ star. Central hydrogen concentration X_c and mass M_c of the convective core as functions of time. The convective core disappears once all the hydrogen is burnt. (After A. Maeder)

This process occurs in stars of mass $M > 1.25\, M_\odot$, which retain a convective core when near to the main sequence. Once the hydrogen content has dropped sufficiently, the radiative gradient $(\nabla T)_{\mathrm{rad}}$ approaches $(\nabla T)_{\mathrm{ad}}$ and the convective zone shrinks rapidly. The core thus changes from an $n = 3/2$ polytrope, corresponding to adiabatic convection, to an isothermal sphere. This rapid change only takes a few per cent of the age of the cluster (Fig. 3.10). As the convective zone disappears the first effect is a slight expansion of both the core and the star, the adjustment of radiative equilibrium increasing the luminosity somewhat.

The Hertzsprung gap has some characteristic quantitative features (Fig. 3.11): (1) distance of the start of the gap above the main sequence; (2) distance of the end of the gap above the main sequence; (3) vertical extent
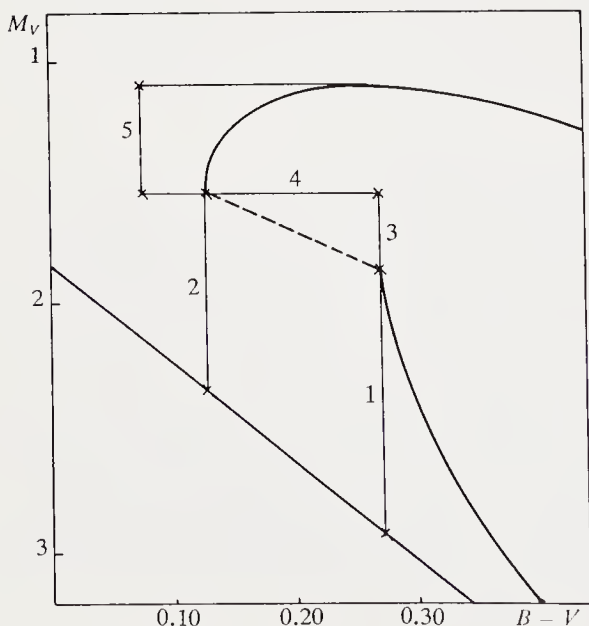

Fig. 3.11. The five quantities characterising the Hertzsprung gap in a galactic cluster (see text)

of the gap; (4) horizontal extent of the gap; (5) position of the maximum of the subgiant branch. The properties of the Hertzsprung gap are a valuable indicator of the structure of stars with mass $M > 1.25\,M_\odot$. Standard models with $X = 0.70, Y = 0.27, Z = 0.03$ are not in accord with observation. They have too wide a gap with an upper edge too close to the main sequence. This seems to arise from strict use of the Schwarzschild criterion to define the convective zone. Taking account of convective overshooting and thus allowing a larger chemically homogeneous central region gives an isothermal core with a bigger mass. Shell burning begins more quickly, explaining the narrower width of the Hetzsprung gap.

In a phenomenological model, Maeder (1976) assumes that overshooting occurs over a distance of about 15% of the mixing length. Taking $l/H_P = 0.5$ this implies overshooting of the order of $0.07H_P$. Clearly this approximation is valid only when H_P is smaller than the convective core radius.

The Age of Galactic or Open Clusters.   The onset of evolution to the giant branch is very visible on all cluster HR diagrams. The corresponding $(B-V)$ colour index fixes the age of the cluster. The result naturally depends on the assumed chemical composition. Determination of the chemical composition is possible if the *absolute magnitude* of the cluster stars are known. With $Z \simeq 0.02, X \simeq 0.70, Y \simeq 0.28$ we get the following values (Maeder 1974):

Cluster	$\log_{10}$ age (years)
Praesepe	8.9
NGC 2360	9.0
NGC 752	9.1
NGC 3680	9.1
M 67	9.6

Evolution Towards the Giant Branch.   As the surface temperatures decrease, hydrogen tends to recombine. The adiabatic gradient $\nabla_{\mathrm{ad}} = \mathrm{d}\log T/\mathrm{d}\log P_{\mathrm{ad}}$ falls below the radiative gradient ∇_{rad}. The Schwarzschild criterion is satisfied and a convective zone forms, whose size grows rapidly as the core contracts.

The properties of this convective envelope are essentially determined by conditions at its surface: gravity and effective temperature, and the opacity. The most important point is that the opacity $\kappa = \kappa_0 \rho^\alpha T^\beta$ increases very strongly with the temperature ($\beta \simeq 10$).

Homology arguments allow us to follow the evolution of a star to the tip of the first giant branch. The reasoning is in three stages: the envelope is decomposed into two regions: an exterior one where ionisation is low and radiative transport dominates convection and an ionised interior one where convection dominates; these are then matched.

(a) *Interior region.* As convection dominates, the star is close to being adiabatic, and thus a polytrope of index 3/2 describes this region if we neglect the mass of the hydrogen-depleted core. The pressure–temperature relation is

$$P_I = K_I T_I^{5/2} .$$

The homology relations $P_I \sim M^2/R^4, T_I \sim M/R$ imply the homology relation for K_I

$$K_I \sim M^{-(1/2)} R^{-(3/2)} .$$

(b) *Exterior region.* Radiative transport dominates. For constant mass and flux we can integrate the radiative equilibrium equation

$$\frac{\mathrm{d}T_E}{\mathrm{d}P_E} = \frac{3}{16\pi acG} \kappa_0 P_E^\alpha \left(\frac{\mu}{\Re}\right)^\alpha T_E^{\beta-\alpha} \frac{L}{M}$$

so that

$$\frac{1}{\alpha+1-\beta}(T_{\mathrm{eff}}^{\alpha+1-\beta}) - T_E^{\alpha+1-\beta} = \frac{3}{16\pi acG}\kappa_0 \left(\frac{\mu}{\Re}\right)^\alpha \frac{P_E^{\alpha+1}}{\alpha+1}\frac{L}{M} ,$$

and we see that the pressure tends to a constant (as long as radiative equilibrium holds). This comes from the rapid rise in the radiative gradient with temperature caused by the large value of the exponent β. Using $L/M \sim T_{\mathrm{eff}}^4/g$ we see that the pressure P_{BE} at the base of the outer radiative zone is given by the homology relation

$$P_{BE} \sim T_{\mathrm{eff}}^{-(\beta-3+\alpha)/(\alpha+1)} g^{1/(\alpha+1)} .$$

(c) *Matching the inner and outer envelopes* is achieved by assuming that the specific entropy is constant across the boundary. This gives another homology relation. Thus

$$S_I = kN_I \log \frac{P_I}{T_I^{5/2}} + \mathrm{const} ,$$

$$S_E = kN_E \log \frac{P_E}{T_E^{5/2}} + \mathrm{const} ,$$

where the ratio N_I/N_E is taken as the ratio of the numbers of free particles per unit mass. For the change from an almost completely ionised medium to an almost neutral one we have

$$\frac{N_I}{N_E} \cong \frac{8X + 3Y}{4X + Y} = i ,$$

and we can write the homology relation

$$\frac{P_{BI}}{T_{BI}^{5/2}} = \left(\frac{P_{BE}}{T_{BE}^{5/2}}\right)^{(1/i)} .$$

For $X = 0.7, Y = 0.30$ we have $1/i = 0.48$. Taking $\alpha = 1$ and $\beta = 10$ and eliminating, we find the homology relations

$$T_{\text{eff}} \sim M^{0.073} R^{-0.055} ,$$
$$L \sim M^{0.291} R^{1.782} ,$$
$$L \sim M^{2.667} T_{\text{eff}}^{-32.585} .$$

This extremely rapid rise in luminosity for decreasing temperature is essentially due to the response of the convective zone to the change of opacity when T_{eff} decreases. This behaviour agrees exactly with numerical results.

The Tip of the Giant Branch. As the core contracts, its temperature increases. We have to consider two cases (Fig. 3.12).

For *masses below* $2.25\,M_\odot$, the effects of degeneracy are dominant and the core evolves towards a strongly degenerate state. The temperature continues to rise during the contraction until a new thermonuclear reaction is triggered, namely
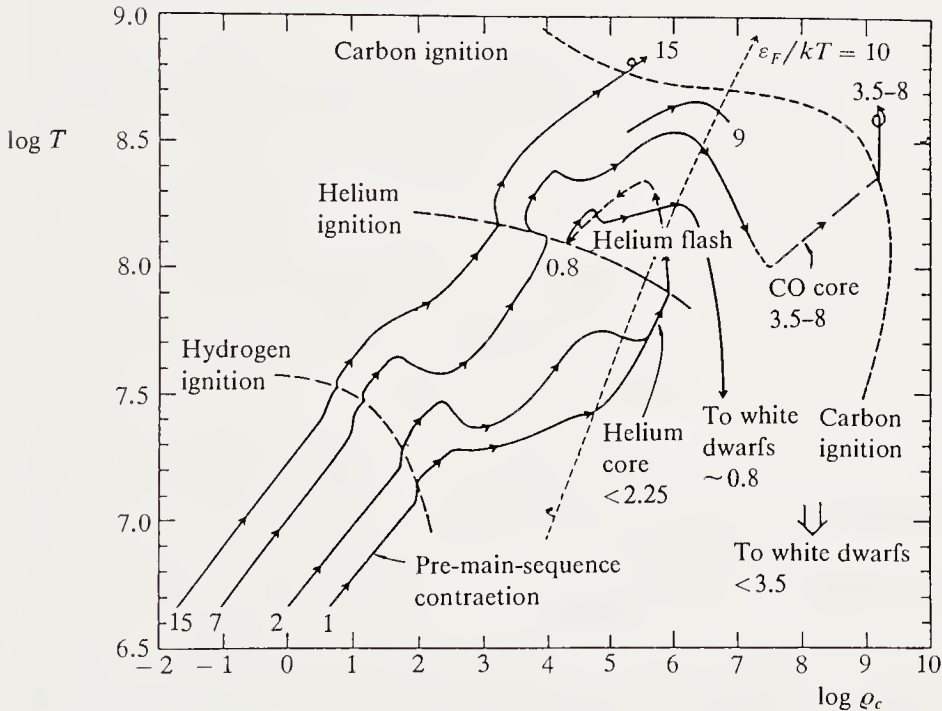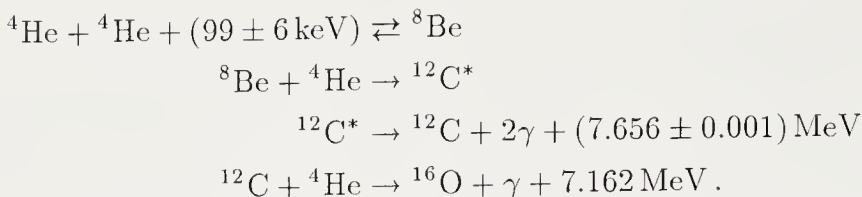

Fig. 3.12. Evolution of the core of a star in the log ρ, log T plane. Curves for $M = 1\,M_\odot$ and $M = 2\,M_\odot$: the core becomes degenerate before helium ignition. The curves for $M = 7\,M_\odot$ and above lead to the carbon flash

$$^4\text{He} + {}^4\text{He} + (99 \pm 6\,\text{keV}) \rightleftarrows {}^8\text{Be}$$
$$^8\text{Be} + {}^4\text{He} \rightarrow {}^{12}\text{C}^*$$
$$^{12}\text{C}^* \rightarrow {}^{12}\text{C} + 2\gamma + (7.656 \pm 0.001)\,\text{MeV}$$
$$^{12}\text{C} + {}^4\text{He} \rightarrow {}^{16}\text{O} + \gamma + 7.162\,\text{MeV} .$$

At the high densities and temperatures $T \sim 10^8$ K prevailing in the core, this reaction runs away. The reason is that the liberated energy raises the temperature without significantly affecting the equation of state, which is dominated by electron degeneracy pressure. For a helium density of 10^6 g cm^{-3}, degeneracy is not lifted until the temperature reaches about 10^9 K. Thus the density remains high, and the temperature grows because the thermal conductivity is too low to remove the heat at the rate it is produced. If K is the thermal conductivity and R_c the helium core radius, the energy equation is

$$\rho C_p \frac{\mathrm{d}T}{\mathrm{d}t} \cong \rho\varepsilon - \frac{KT}{R_c^2} ,$$

where we have reduced $\nabla^2 T$ to its order of magnitude T/R_c^2. Once $\rho\varepsilon$ exceeds the diffusion losses the temperature runs away.

(a) *Reaction rate.* Beryllium ^{8}Be formed in the ground state is 99 keV above the rest-mass energy of 2 alpha particles. The ^{8}Be nucleus is highly unstable and gives 2 alpha particles with a lifetime of order 10^{-17} s. A very small fraction of beryllium has the time to react and produce carbon.

Carbon itself is formed in an excited state at 7.656 ± 0.008 MeV above the ground state[1] (Fig. 3.13).

The probability of returning to the state ^{8}Be + ^{4}He is much larger than the de-excitation of $^{12}\text{C}^*$ to the ground state. We may thus write two Saha equations for the populations of ^{8}Be and $^{12}\text{C}^*$, which when combined give

$$N({}^{12}\text{C}^*) = N\,{}^3({}^4\text{He}) \frac{h^6}{(m_{44} m_{48})^{3/2}} \frac{1}{(2\pi kT)^3} \exp\left(-\frac{(\chi_1 + \chi_2)}{kT}\right) ,$$

where $m_{44} = m_4^2/(2m_4)$ and $m_{48} = (m_4 m_8)/(m_4 + m_8)$, with m_4 and m_8 the masses of the helium and beryllium nuclei respectively. The transition rate to the ground state is then

$$\frac{\mathrm{d}N({}^{12}\text{C})}{\mathrm{d}t} = N({}^{12}\text{C}^*)\frac{\Gamma_\gamma + \Gamma_{\pm\beta}}{\hbar} ,$$

where Γ_γ and $\Gamma_{\pm\beta}$ are the widths of the levels for transitions to the ground state via 2-photon emission or e^+e^- pair emission. $(\Gamma_\gamma + \Gamma_{\pm\beta})$ is of the

---

[1] The need for a resonant reaction ^{8}Be + ^{4}He was pointed out by F. Hoyle (1956) in order to explain the presence of ^{12}C in the Universe. The resonance was found a little later. This is a rare case of a physical prediction originating from astrophysics.

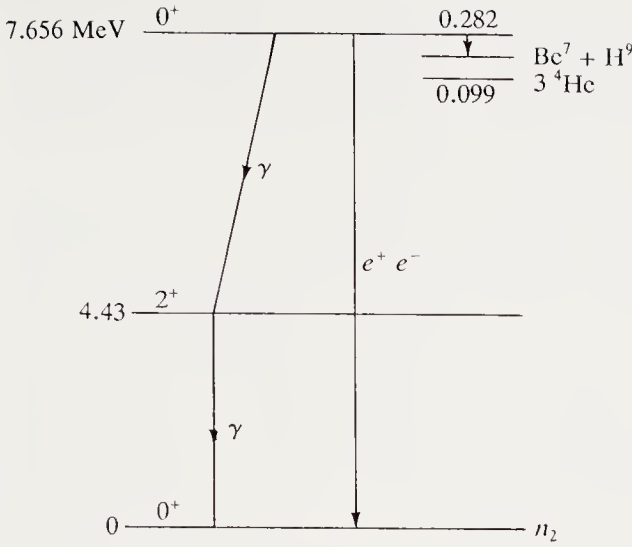Fig. 3.13. Nuclear energy levels involved in the formation of ^{12}C

order 2.5×10^{-3} eV. We still have to consider displacements of these levels by collective effects.

For weakly correlated plasmas collective effects displace the levels through the Debye potential, with the full effect

$$f = \exp(2.25 \times 10^{-3} \rho^{1/2} T_8^{-3/2})$$

for pure helium.

For a strongly correlated plasma we can use the approximate expression for the displacement of the levels

$$\frac{U_0}{kT} = 0.205[(Z_1 Z_2)^{5/3} - Z_1^{5/3} - Z_2^{5/3}] \left(T \frac{X_i Z_i}{\rho_i} \rho \right)^{1/3} T_6 .$$

and combine this expression for the reactions $2\alpha \rightarrow {}^8\text{Be}$ and ${}^8\text{Be} + \alpha \rightarrow {}^{12}\text{C}^*$, giving

$$\frac{U_0}{kT} = 1.6738(\rho/T_6^3)^{1/3} .$$

We note that for the reaction $2\alpha \rightarrow {}^8\text{Be}$ the binding energy of beryllium, 0.99 keV for a pure helium density $\rho = 7.1 \times 10^7$ g cm^{-3}, is reached. Above this density beryllium ^{8}Be is stable and the calculation of the carbon formation rate is modified.

(b) *Thermal conductivity.* We gave above on p. 103 the thermal conductivity of a degenerate medium. We note here that the timescale for heat transport is short compared with the nuclear timescale. The assumption of a degenerate isothermal core is justified.

(c) *Neutrino emission.* Four main processes produce $(\nu, \bar{\nu})$ pairs. Neutrinos and antineutrinos pass freely through stellar matter (except during the formation of neutron stars; Chapter 8). Consequently neutrino production always constitutes an energy loss from the star.

The four processes are:

(1) Production of photoneutrinos in the interaction of a photon with an electron in the presence of an ion:

$$\gamma + e^- + (Z) \rightarrow \nu + \tilde{\nu} + \bar{e} + (Z).$$

This is a modified Compton effect.

(2) Electron–positron annihilation:

$$e^+ + e^- \rightarrow \nu + \tilde{\nu}.$$

(3) Plasmon–neutrino interaction: transverse plasmons (photons constantly interacting with the plasma) or longitudinal plasmons (plasma waves) spontaneously transform themselves into $(\nu, \bar{\nu})$ pairs:

$$\text{plasmon} + \text{plasma} \rightarrow \nu + \tilde{\nu} + \text{plasma}.$$

(4) Bremsstrahlung: a decelerated electron produces a $(\nu, \bar{\nu})$ pair:

$$e^- + (Z) \rightarrow e^- + (Z) + \nu\tilde{\nu}.$$

Figure 3.14 shows where the various processes dominate in the $\log \rho$, $\log T$ plane for a fully ionised Fe plasma.
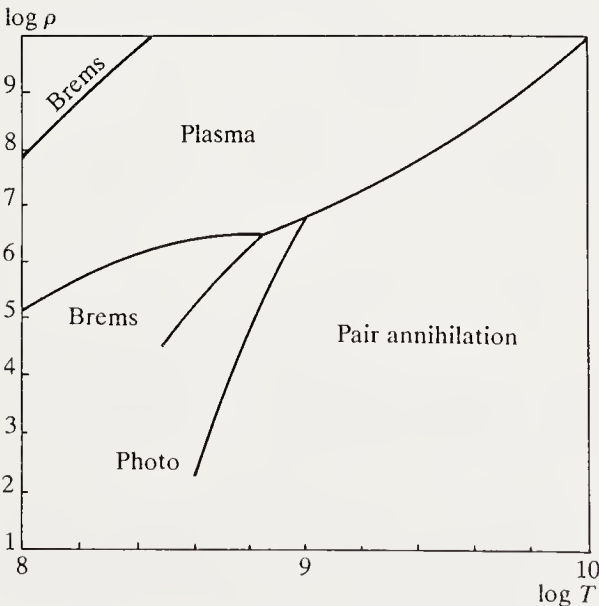

Fig. 3.14. Neutrino production domains in the $\log T$, $\log \rho$ plane for a fully ionised Fe plasma (After G.G. Festa and M.A. Ruderman, *Phys. Rev.* **180**, 1227 (1969))
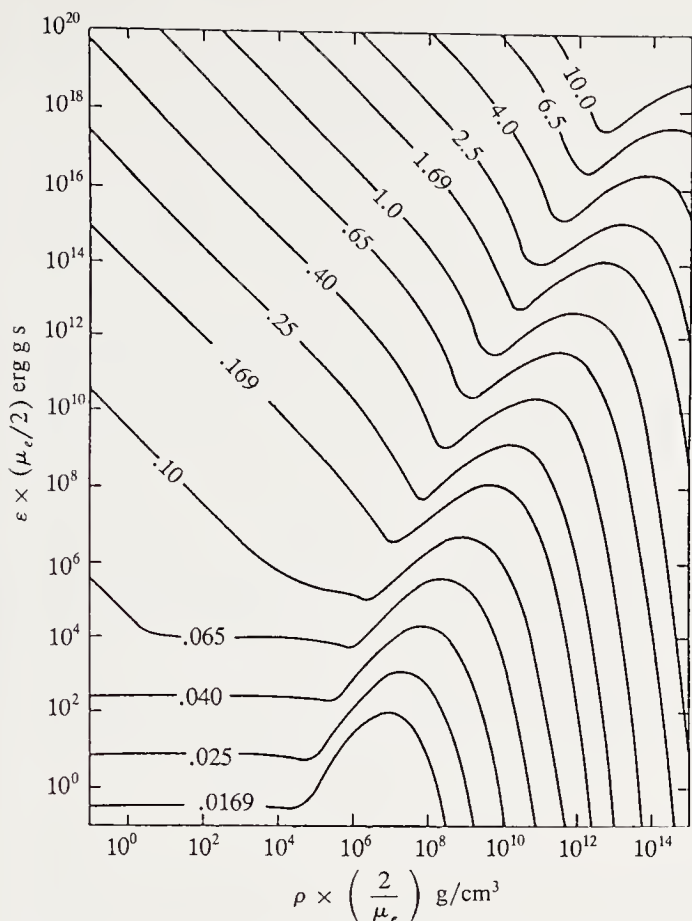
Fig. 3.15. Energy loss rate via neutrino emission. (After G. Beaudet et al., *Astrophys. J.* **150**, 979 (1967). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1967 The American Astronomical Society). The curves are labelled by the value of $kT/m_e c^2$

At high temperatures and low densities the electron–positron annihilation process dominates, with

$$\varepsilon \, (\text{pairs}) \simeq 10^{10} \frac{T}{\rho} \, .$$

At high density and moderate temperatures bremsstrahlung dominates, with

$$\varepsilon \, (\text{brems}) \simeq 0.76 \frac{Z^2}{A} \left(\frac{T}{10^8 \, \text{K}} \right)^6 .$$

The plasmon–neutrino contribution in the intermediate domain has no simple analytic representation (Fig. 3.15).

These energy losses are significant at some epochs of stellar evolution, essentially after the formation of a carbon–oxygen core. The neutrino loss rate completely determines the conditions for the runaway of the $^{12}\text{C} + {}^{12}\text{C}$

reaction but is totally negligible for deciding the stability of the $3\alpha \rightarrow {}^{12}C$ reaction. Runaway occurs if $\rho\varepsilon R_c^2 = KT$.

Replacing R_c by the order of magnitude estimate

$$\tfrac{4}{3}\pi R_c^3\rho = qM \,,$$

where q is the mass fraction of the helium core, we can calculate approximately the helium ignition curve (Fig. 3.12).

At maximum, the luminosity due to the $3\alpha \rightarrow {}^{12}C$ reaction reaches $L \simeq 10^{11} L_\odot$. This energy does not leave the star but remains in a convective zone which reaches almost to the hydrogen-burning shell. The temperature rises until degeneracy disappears. The connection between temperature and pressure now permits a new equilibrium: the helium core expands and cools, and helium burning continues as in more massive stars.

For *masses above* $2.25\,M_\odot$ the temperature and density of the helium core rise without any onset of degeneracy. Evolution then proceeds with core helium burning and shell hydrogen burning. The main effect of helium
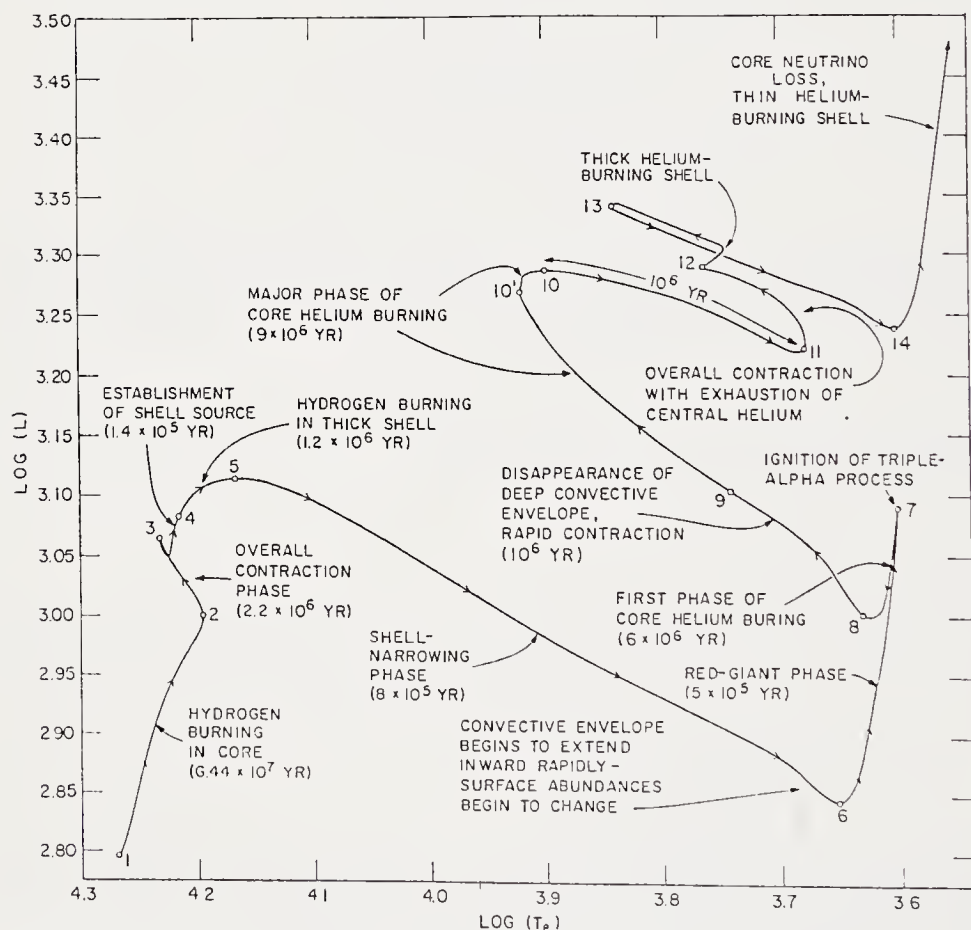

Fig. 3.16. Evolutionary track for a $5\,M_\odot$ star, without mass loss or mixing, in the HR diagram. (From I. Iben, Jr., *Ann. Rev. Astron. Astrophys* 5, 573 (1967). Reproduced by kind permission of *Annual Reviews of Astronomy and Astrophysics*; © 1967 Annual Reviews Inc.)

130

burning is an expansion of the core which modifies the conditions in the thin hydrogen-burning shell and reduces the luminosity. Helium burning expands the core, but the main luminosity source is always hydrogen shell burning. To maintain equilibrium between the luminosity produced and that emitted the outer layers contract as the core expands, causing the outer radius of the star to decrease and the effective temperature to rise (from about 4000 to 8000 K).

Core helium exhaustion produces the same type of left–right–left oscillation across the HR diagram, followed by a new expansion towards the red supergiants. The main energy source is now helium shell burning. The core becomes degenerate, and neutrino losses stabilise the $^{12}C - ^{12}C$ reaction (Fig. 3.16).

The Horizontal Branch. After the *helium flash* and the formation of a non-degenerate helium-burning core, a star of mass $M \lesssim 2.25\,M_\odot$ returns to a lower luminosity, about $30 - 100\,L_\odot$, but its position in effective temperature depends strongly on its chemical composition.

The beginning of the *zero-age horizontal branch* (ZAHB) is easily visible (Fig. 3.17) in the HR diagrams of globular clusters (population II). For young population I stars (Hyades type) it overlaps the ascending giant
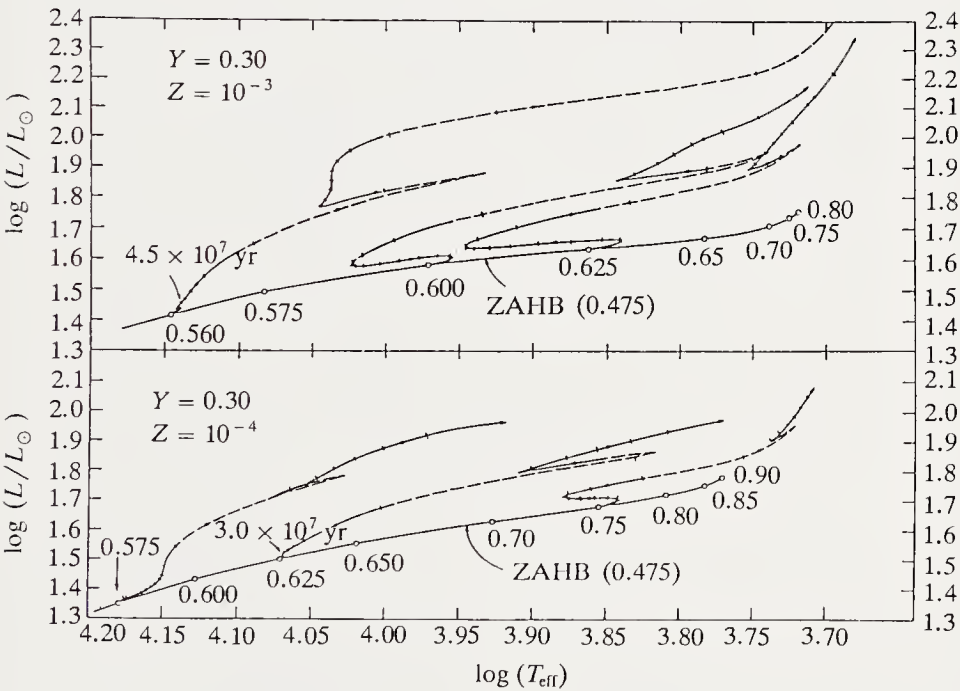

Fig. 3.17. Horizontal and asymptotic branch for two population II chemical compositions. The solid curve is the ZAHB, with masses marked along it in solar units. The mass of the helium core is taken equal to the mass of degenerate helium *before* the helium flash. The later evolution is indicated for some masses by a *continuous curve* (slow evolution) or *dashes* (very rapid evolution). In the second slow phase of evolution helium burns in a thick shell and hydrogen continues to burn in a thin shell. The marks on the tracks are at intervals of 10^7 y. (From S.E. Strom et al., Astron. Astrophys. **8**, 243 (1970))

branch. A simple homology argument at constant luminosity relates the effective temperature to the heavy-element abundance. With opacity $\propto \rho T^{-\beta}$ we have $T_{\text{eff}} \sim Z^{-(\beta+3)/2}$. For $Z = 0.0001$ (old population II) to $Z = 0.02$ (young population I) we find $T_{\text{eff}}(\text{I})/T_{\text{eff}}(\text{II}) = 0.64$ and 0.41 for $\beta = 3$ and $\beta = 0$ respectively. Temperatures differing by about a factor of 2 are indeed found between the ZAHB for young population I and and that for old population I1.

Intermediate-Mass Stars. This mass range is conventionally defined as that of stars which *in the absence of mass loss* would reach carbon or oxygen ignition in a degenerate core. For population I stars this means stars of initial masses between 1.4 (the Chandrasekhar limit) and $10 \, M_\odot$.
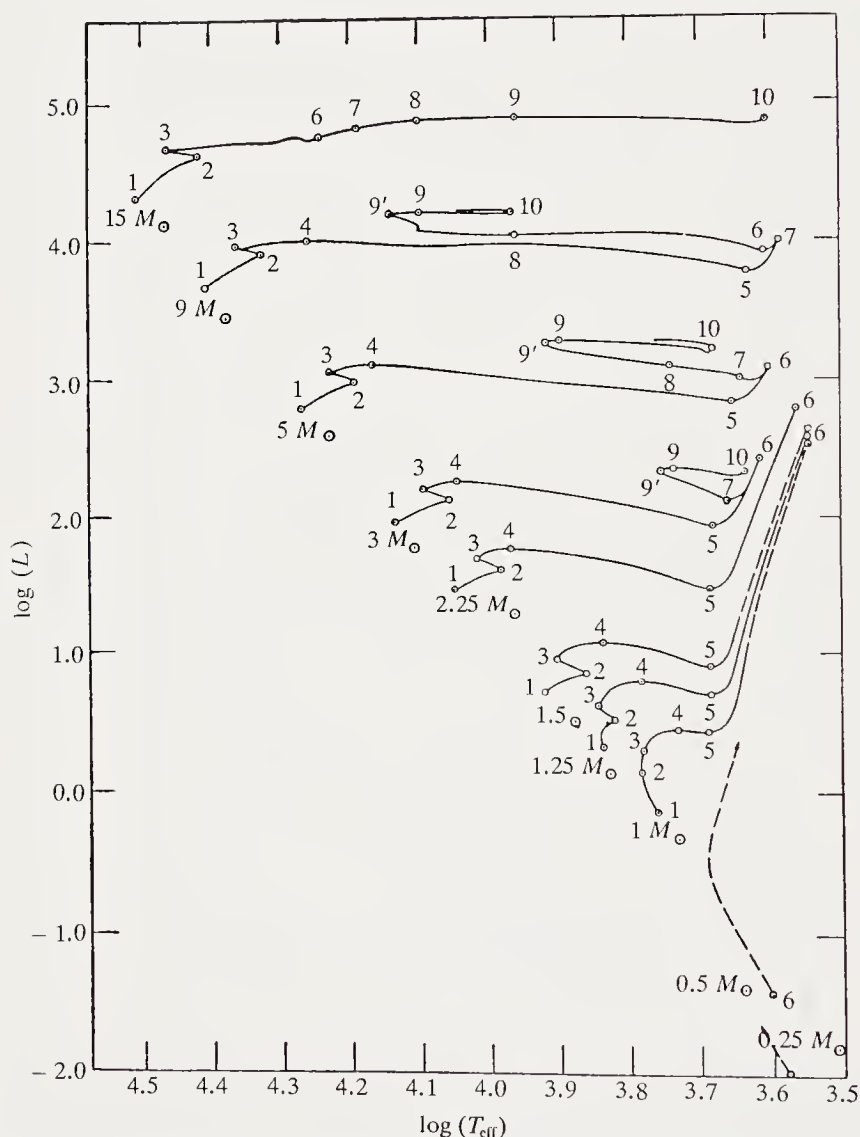

Fig. 3.18. Evolution to the giant branch for stars from $0.5 \, M_\odot$ to $15 \, M_\odot$. (From I. Iben, Jr., *Ann. Rev. Astron. Astrophys.* 5, 584 (1967). Reproduced with the kind permission of *The Annual Review of Astronomy and Astrophysics*; © 1967 Annual Reviews Inc.)

As we have seen, for stars of mass $M < 2.25\,M_\odot$ the helium core becomes degenerate before helium burning starts. Once a composition discontinuity and hydrogen shell burning appear, *all* stars evolve to the giant branch (Fig. 3.18). The reason is the same in each case: the shell source causes a reduction of the radiative gradient, equivalent to an increase in the polytropic index, which results in an expansion of the outer layers.
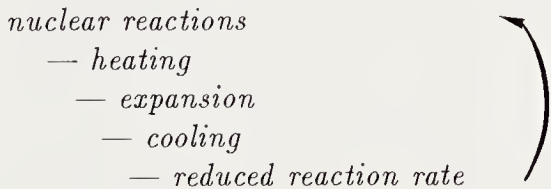
After the excursion to temperatures $\sim$ 4000–5000 K, the evolution becomes very complicated. The main features are:

(1) The occurrence of thermal instabilities.

(2) Crossing of regions of pulsational instability (Cepheids for population I stars with masses above about $5\,M_\odot$; RR Lyrae stars for low-mass population II stars).

(3) Mass loss in a *stellar wind*. This is an important phenomenon to which we shall return (Chap. 4); it is poorly understood. Simple reasoning shows its importance if we refer to Fig. 3.12. After core helium burning, stars between $2.25\,M_\odot$ and $\sim$ 8–10 $M_\odot$ form a degenerate carbon–oxygen core. As it contracts this core enters a regime where a runaway of the ^{12}C – ^{12}C reaction can occur because of screening, because the neutrino losses cannot cope with the thermonuclear energy production rate.

In the absence of mass loss this would mean that all stars of mass above $1.4\,M_\odot$ would explode at the end of their evolution. Let us assume that $20\,M_\odot$ currently forms into stars each year in the Galaxy. With a Salpeter initial mass function $dN = KM^{-2.35}dM$ and a lower mass cutoff of $0.01\,M_\odot$ we find a supernova rate for stars of mass above $1.4\,M_\odot$ of $dN_{\mathrm{expl}}\,dt = (2.8)^{-1}$ per year. If on the other hand we suppose that supernovae only occur for masses above $10\,M_\odot$, this rate becomes one per 39 years, close to the usual estimate of one per 50 years in the Galaxy.

We shall return to the evolution of intermediate-mass stars in Sect. 3.7 on mass loss.

Massive Stars ($M \gtrsim 10\,M_\odot$).   The cores of these stars are *non-degenerate* at carbon ignition. The effect of the feedback

nuclear reactions
$\quad$ — *heating*
$\quad\quad$ — *expansion*
$\quad\quad\quad$ — *cooling*
$\quad\quad\quad\quad$ — *reduced reaction rate*
$\Big)$

stabilises carbon burning. These stars also produce stellar winds which influence their evolution even before they reach the giant branch.

3.6 The Sun

The importance of studies of the Sun's internal structure lies not merely in its intrinsic interest but also in the possibility of testing some of the fundamentals of the theory, both physical and astrophysical. This is because of the precision of the data (mass, luminosity, radius, age, and results specific to the Sun such as helioseismology and neutrinos), as well as the very accurate spectroscopy which is possible (abundances of ^{3}He, ^{7}Li, ^{9}Be). It is also important to compare solar results with those for stars of similar mass (especially in the Hyades), providing a stringent test of the theory.

3.6.1 The Basic Model

The basic solar model requires the best input data for opacities, equation of state, nuclear reaction rates, and a model of the outer convection zone involving the parameter $\alpha = l/H_P$ giving the mixing length. We take initially uniform chemical composition and calculate a model assuming no mixing of chemical elements. We let it evolve up to $t = 4.6 \times 10^9$ y (the Sun's age) and compare calculated and observed quantities. The chemical composition (X, Y, Z) may be modified, as well as the mixing length.

We have to adjust the three parameters (X, Y, Z) so as to obtain the present mass, luminosity, and radius for the Sun. However, the primordial heavy-element abundance Z cannot be chosen freely because of the "cosmic" abundance of elements, such as results from the abundance ratio (Si/H) in the Sun and (other elements/Si) on the Earth and in meteorites. The choice of Z is thus restricted to a very small range. Near $M = 1.25 M_\odot$ (log $M \simeq 0.1$) we have $\Delta \log L = -(2.5 \Delta X + 13 \Delta Z)$, allowing Z-variations of the order of 0.001 only.

Lebreton and Maeder (1986) chose $Z = 0.02$ and got the values below for a model with $\alpha = 2.2$ (this value is discussed later) and no mixing.

3.6.2 Solar Neutrinos

The reactions of the pp III chain (Sect. 3.4.1) are very sensitive to the temperature. The temperature dependence of the ^{3}He, ^{7}Be abundances and the production rate of boron (^{8}B) shows that the main contribution to the high-energy neutrino flux (^{8}B radioactivity) varies like T^{20}. A very small temperature difference changes the neutrino flux considerably.

The first estimate of the neutrino flux (1957) was very large. An attempt to measure these neutrinos through the reaction

$$^{37}\text{Cl} + \nu \rightarrow {}^{37}\text{A} + e^-$$

followed by measurement of the radioactivity of argon ^{37}A,

$$^{37}\text{A} \rightarrow {}^{37}\text{C} + e^+ + \nu,$$

Table 3.2. Solar evolution calculated for $Y = 0.287$, $\alpha = 2.2$, $Z = 0.02$ (Lebreton and Maeder 1986)

Age (10^9 y)	$\log \dfrac{L}{L_\odot}$	$\log T_{\text{eff}}$	$\log T_c$	$\log \rho_c$	X_c
0.000	$-.191$	3.745	7.134	1.879	.693
0.118	$-.154$	3.748	7.130	1.902	.688
0.198	$-.145$	3.750	7.132	1.918	.683
0.326	$-.139$	3.751	7.133	1.928	.675
0.838	$-.123$	3.752	7.138	1.952	.640
1.350	$-.108$	3.753	7.144	1.977	.602
1.862	$-.092$	3.755	7.151	2.004	.563
2.374	$-.076$	3.756	7.158	2.032	.524
2.886	$-.060$	3.758	7.165	2.062	.484
3.398	$-.043$	3.759	7.173	2.094	.443
3.654	$-.034$	3.760	7.177	2.111	.423
3.910	$-.025$	3.760	7.181	2.128	.402
4.140	$-.017$	3.761	7.185	2.144	.383
4.371	$-.008$	3.762	7.189	2.160	.363
4.601	$.000$	3.762	7.193	2.178	.344

was made underground, using chlorine in the form of C_2Cl_4, the standard industrial solvent. The first detector weighed 2 tons. Improvements in solar models soon produced a better estimate of the expected neutrino flux. With a detection threshold $Q = 0.816$ MeV, the use of chlorine only allows detection of beryllium and boron neutrinos. Davis's experiment, installed in 1964 at a depth of 1500 metres in the Homestake gold mine at Lead (South Dakota) started to get significant results during the 1970s. If $\sigma(E_\nu)$ is the ^{37}Cl cross-section as a function of energy and $\phi(E_\nu)\mathrm{d}E_\nu$ the flux of neutrinos, the integral $\phi_\nu = \int \sigma(E_\nu)\mathrm{d}E_\nu$ gives the capture rate per atom. The cross-section σ is very small, of the order of 10^{-44} cm^2, leading to the introduction of the solar-neutrino unit (SNU) which is one neutrino per second per 10^{36} atoms. Davis's result is $\phi_\nu = 2.1 \pm 0.3$ SNU [2] while Bahcall's models predict $\phi_\nu = 7.5$ SNU, and the most recent model of Lebreton and Maeder predicts 8 SNU. These predictions are clearly affected by uncertainties in the opacities (which fix the Sun's central temperature) and cross-sections. Without going into details it appears impossible to adjust the models so as to predict fluxes below 5 or 6 SNU. We shall see that helioseismology gives insight into conditions in the solar interior and allows a check of the models.

3.6.3 The Solar-Neutrino Deficit

The deficit of solar neutrinos poses a difficult and important problem for physicists and astrophysicists. Discussions of it start from the temperature sensitivity of the reaction $^7\text{Be} + p \rightarrow {}^8\text{B} + \gamma$, so that the neutrino flux

---

[2] Davis's 1988 measurements gave a flux of 4 SNU, but the flux appears to have decreased in 1989. The variations of the neutrino flux are not understood.

$\phi_\nu(^8\text{B})$ goes as T^{20} near $T = 1.5 \times 10^7$ K. In the standard model this reaction gives only 0.015 % of the solar energy production. Thus the assumption of a different beryllium reaction rate has no effect on the solar model. We have the choice of two strategies: we can assume the central temperature correct or not.

If T_c is correct then *either* the cross-section for $^7\text{Be}(p, \gamma)^8\text{B}$ is overestimated *or* something happens to the neutrinos between the Sun and us.

If T_c is wrong there are various possibilities: opacities, pp reaction rates, various structural effects or (see below) new physics.

We summarise the discussion below, stressing the essential objections.

Hypothesis I: The Models Give the Correct Central Temperature. (a) The $(^7\text{Be}, p)^8\text{Be}$ cross-section is too large. This seems unlikely: we are extrapolating experimental data to low energies, and a change by a factor 3 is ruled out.

(b) The neutrino deficit is caused by a property of these particles. In 1967 Pontecorvo suggested that neutrinos might have a mass, allowing oscillations between different species, for example electron neutrinos ν_e and muon neutrinos ν_μ. This would be analogous to the oscillation of K-mesons between the two varieties K_1, K_2 because of a small mass difference $((m_1 - m_2) \simeq 5 \times 10^{-5}$ eV, compared with the K meson mass 497.8 MeV). We shall examine the consequences of this idea below. Whatever the cause, the origin of the neutrino deficit in Davis's experiment can only be found by other experiments with different detectors. The capture of ν_e by the reaction

$$^{71}\text{Ga} + \nu_e \rightarrow {}^{71}\text{Ge} + e^-$$
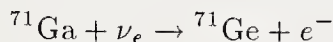
with a threshold of 0.25 MeV will allow detection of neutrinos from the pp I chain, whose number is directly related to the solar luminosity and thus *model-independent*. The standard solar model predicts a flux of about 130 SNU. If the deficit in the Davis experiment *is not due* to neutrino properties there should be a pp I contribution of about 90 SNU. Underground experiments using gallium for neutrino counting have been started (SAGE: a Soviet–American cooperation in Baksan, Caucasus, and GALLEX, a European collaboration in Gran Sasso, Italy. The results of the two gallium experiments came recently into agreement, giving a neutrino flux of 83 ± 18 (error bar) ±8 (systematic errors) SNU (published by the GALLEX cooperation, 1992). This represents about 2/3 of the predicted flux.

The underground Kamiokande experiment in Japan detects neutrinos above 8 MeV, making use of the Cerenkov effect produced in a pool of 2000 tons of water by electrons projected by collisions with solar neutrinos. The measured flux is of the order of one half of the predicted one.

Hypothesis II: The Central Temperature of the Models Is too High.
A temperature error could have one of several causes:

(a) *Opacities.* Lower opacity gives a lower central temperature. For a Kramers law $\kappa = \kappa_0 \rho T^{-3.5}$ a homology argument gives

$$\phi_\nu(^{37}\text{Cl}) \sim \kappa_0^{8/3} \, .$$

The desired factor 3 implies a reduction in the opacity by about 35 %, whereas all improvements in opacity calculations over the last 20 years have led to *increases.*

(b) *Thermonuclear reactions.* An increase in the cross-section for $p(p, \nu e^-)^2 D$ would reduce the central temperature. The 1/3 reduction in the neutrino flux would need an increase of this cross-section by about 50 %. The theoretical cross-section is well known, and its only uncertainty is the neutron lifetime; this does not allow such a large change in the cross-section.

(c) *Magnetic fields.* Strong magnetic fields in the Sun's centre would reduce the gas pressure and thus the central temperature. To reduce the central temperature by 5 % requires a field of about 10^8 G. Such a field would prduce many types of instability, rendering its presence doubtful.

(d) *Diffusion.* Turbulent diffusion, more efficient than atomic diffusion, could bring more nuclear fuel into the core, such as ^{3}He and ^{1}H. This would reduce the temperature and thus the boron neutrino rate as required. However, this process of turbulent diffusivity is easily inhibited by a composition gradient, and (see Chap. 6) the resulting composition distribution is incompatible with data from helioseismology.

New Physics. (a) *New particles.* Particle physicists have proposed new weakly interacting particles on the basis of symmetry arguments. "Dark matter" has been suggested as a manifestation of them. A small number of them in the Sun would transport heat and reduce the central temperature, and hence the boron neutrino flux. Such new solar models have several free parameters (cross-section, mass, lifetime, etc. of the new particles). With the parameters adjusted to fit the solar-neutrino observations these new particles should be detectable in underground experiments at present in preparation.

(b) *Massive neutrinos.* Measurement of neutrino fluxes as functions of distance from nuclear reactors should reveal the oscillations. Neutrinos produced in the reactor are electron neutrinos; during propagation some of these transform into muon neutrinos, which are undetectable. If we assume two eigenstates ν_1, ν_2 of masses m_1, m_2 which differ from the ν_e, ν_μ states, the oscillations are caused by phase mixing between the two states, or in atomic units

$$\frac{d\varphi_\nu}{dt} = E_1 - E_2 = \sqrt{p^2 + m_1^2} - \sqrt{p^2 + m_2^2} \cong \frac{m_1^2 - m_2^2}{2p} \, .$$

In the usual units the characteristic length is given by

$$l_\nu = \frac{2hE}{\Delta m^2 c^3} = 248 p \left(\frac{\text{MeV}}{c} \right) / \Delta m^2 (\text{eV}^2) \, \text{cm} \, .$$

In propagation in vacuo the transformation probability $\nu_e \to \nu_\mu$ is given by

$$P(\nu_e \to \nu_e) = 1 - \sin^2 2\theta \sin^2 \frac{\pi x}{l} \, ,$$

where θ is the *mixing angle* of the two species. Only an upper mass limit of about 10 eV has been found.

In matter the interactions $(\nu_e e)(\nu_\mu e)$ alter the length of the oscillation. In a constant-density medium with

$$l_0 = \frac{2\pi}{\sqrt{2} \, G N_e} = 1.624 \times 10^9 \, \text{cm}/\rho \, ,$$

where G is the Fermi coupling constant, N_e the electron number density in cm^{-3}, and ρ the electron density in units of Avogadro's number, the new oscillation length is

$$l_M = l_\nu \left[1 - 2 \frac{l_\nu}{l_0} \cos 2\theta + \left(\frac{l_\nu}{l_0} \right)^2 \right]^{-1/2} \, ,$$

and we define a new mixing angle

$$\tan 2\theta_M = \frac{\sin 2\theta}{(\cos 2\theta - l_\nu/l_0)} \, .$$

The most interesting case for solar neutrinos is that for $m_2 > m_1$. In this case there is a sort of resonance if the central density exceeds a certain critical value; this resonance is important if the gradient is not too steep, and in this case the transformation of the ν_e is conserved. There is then some suppression of the neutrino flux for neutrino momenta in

$$7.5 \times 10^4 \left(\frac{\text{MeV}}{c} \right) \cos 2\theta \left(\frac{\Delta m^2}{1 \, \text{eV}^2} \right) < p$$

$$< 2 \times 10^8 \left(\frac{\text{MeV}}{c} \right) \sin 2\theta \tan 2\theta \left(\frac{\Delta m^2}{1 \, \text{eV}^2} \right) \, ,$$

corresponding to masses near $\Delta m^2 \simeq 10^{-6} \, \text{eV}^2$.

From these calculations we can use the representation $(\sin^2 2\theta, \Delta m^2)$ to plot the fluxes $\phi_\nu(^{37}\text{Cl})$, $\phi_L(^{71}\text{Ga})$. The experimental results for gallium could determine the *mixing angle* and the *mass difference* for the two neutrinos (Figs. 3.19a and b). It is easy to see in Fig. 3.19a that two small areas in the plane $(\sin^2 2\theta, -\Delta m^2)$ fit the constraints of the experimental data, with a mass of neutrinos of the order of $(10^{-5} \, \text{eV}^2)^{1/2}$. However, one should keep in mind that we do not yet have a final solar model and we cannot exclude the possibility of explaining the observations with zero-mass neutrinos.
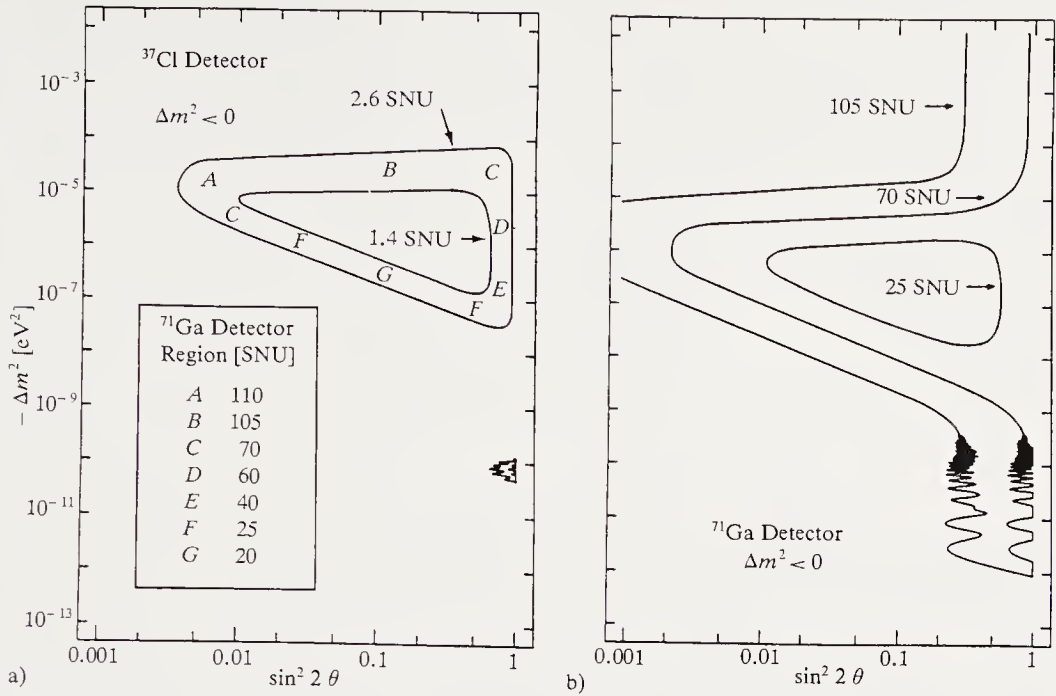
Fig. 3.19. The solar-neutrino flux in the mass-difference–mixing-angle plane (logarithmic scales for Δm^2 in eV^2, and for $\sin^2 2\theta$). (**a**) gives the curves of 2.6 SNU and 1.4 SNU for the chlorine experiment, and at A–G the fluxes for the gallium experiment; (**b**) gives the predictions for the curves of fluxes 25 SNU, 70 SNU, and 105 SNU for the gallium experiment. (After J. Bouchez et al., *Particle Physics and Fields*, **32**, 499 (1986))

3.7 Evolution with Mass Loss

In Chap. 4 we shall discuss the observational data for *stellar winds* and *mass loss* and their theoretical interpretation, which can have considerable importance. In the following we shall assume that mass loss occurs at a rate determined by the star's parameters.

The basic fact is that the wind's mechanical luminosity $\frac{1}{2}\dot{M}v_\infty^2$ is always small compared with the radiative luminosity. This means that we can use the results already found for static models (e.g. luminosity, radius, envelope mass as a function of core mass) to follow at least approximately the trajectory of a star in the HR diagram. Clearly the aim is to bring theory into agreement with observation. We shall distinguish between the evolution of low-mass stars (particularly population II stars, where we wish primarily to determine the ages of globular clusters), the evolution of intermediate-mass stars (particularly to the white-dwarf state) and the evolution of massive stars (Wolf–Rayet stars and evolution to a supernova).

3.7.1 The Age of Globular Clusters

Before attempting any age determination, it is important to identify the evolutionary states of the various families on the colour–magnitude diagram.
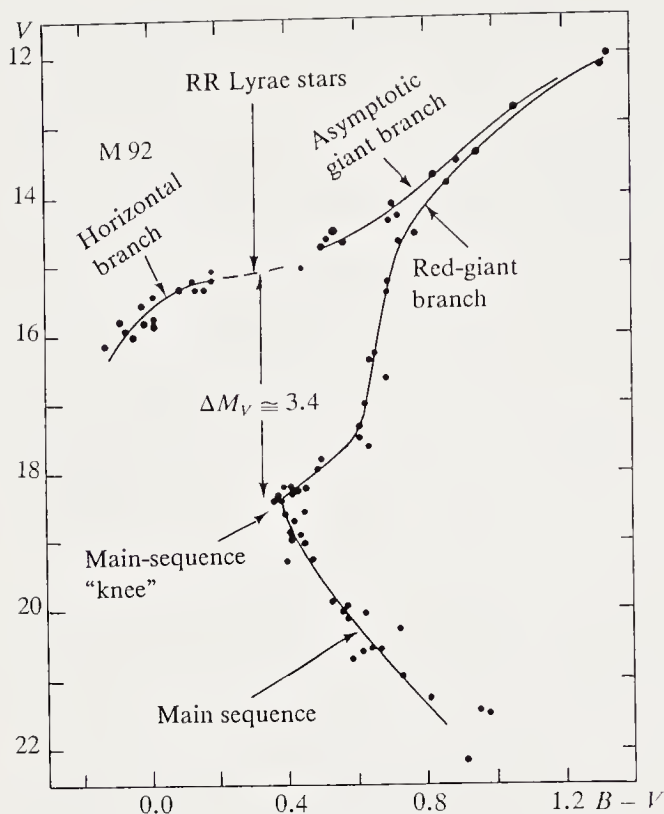
Fig. 3.20. HR diagram for globular clusters. Names of the various evolutionary stages. (From A. Renzini, Course at Saas-Fee 1977, ed. by A. Maeder and P. Bouvier, Observatoire de Genève)

Figure 3.20 defines the relevant nomenclature for M92. We distinguish the main sequence, the *knee* where evolution to the giant branch begins, the red-giant branch (RGB), asymptotic giant branch (AGB), the horizontal branch (HB), and the instability strip containing the RR Lyrae stars. Clearly these regions are not straightforwardly related to the chronological order of evolution.

At first sight it appears that the method used to find the age of open or galactic clusters will work also for globular clusters; we could calculate the isochrones and compare them with the observed $(B - V, V)$ HR diagram. The mass of the stars at the knee of the main sequence where evolution to the giant branch begins should give the cluster age. The problem with this method is that the colour indices of stars with large convective zones are very sensitive to both the modelling of the convective zone and the metallicity. Adjusting the isochrones does not simply involve a vertical shift (in magnitudes), because the horizontal position (in $B - V$) is uncertain.

Because of the great age ($> 10^{10}$ y) of globular clusters, which are population II objects, the knee of the main sequence corresponds to low-mass ($< 1 M_\odot$) stars. Comparison of the observational data with models thus involves low-mass stars having few heavy elements but probably a primordial helium content which is the same for all clusters. Because we know little of the circumstances under which the clusters formed we cannot take them to have the same metallicity. Given their present compositions we assume
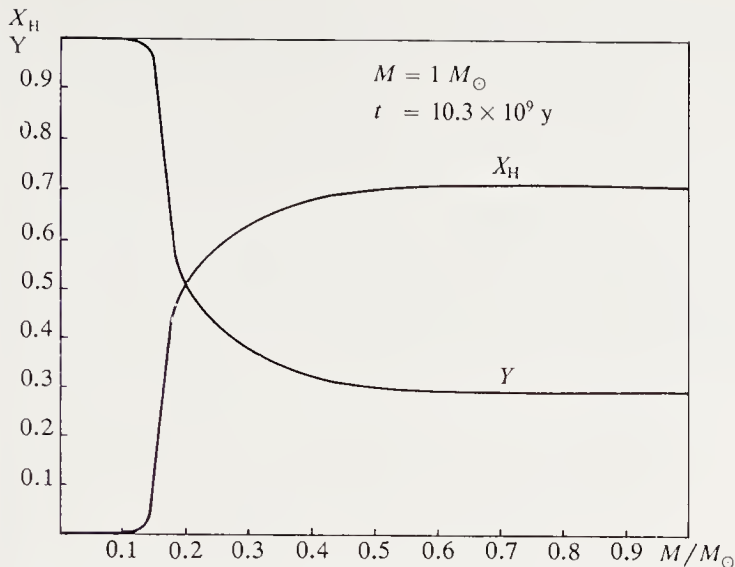
Fig. 3.21. Hydrogen and helium content as a function of mass fraction for a 1 $M_\odot$ star at the end of its main-sequence lifetime

that they were formed from a medium which was only weakly enriched by nucleosynthesis: the abundance of CNO and heavier elements (formed in primordial nucleosynthesis) being extremely small.

Before discussing low-mass stellar evolution we summarise here the main features of the evolution of a low-mass star.

(a) After core hydrogen exhaustion the star evolves towards the giant branch, and then along the RGB. At the end of its main-sequence evolution the helium concentration has changed in a significant fraction of the star (Fig. 3.21). On the RGB the surface convection zone fills more and more of the star, while the hydrogen-burning shell moves slowly outwards (Fig. 3.22). The convection dredges up the helium formed on the main sequence (*first dredge-up*) and enriches the helium content of the outer layers, which is important for the interpretation of the RR Lyrae stars (below).

An important quantity is the luminosity at the knee of the isochrones. A wide grid of models yields the interpolation formula

$$\log L_{\text{knee}} = (0.019 \log^2 Z + 0.065 \log Z + 0.41 Y - 1.179) \log t_9$$
$$+ 1.246 - 0.028 \log^2 Z - 0.272 \log Z - 1.073 Y ,$$

where t_9 is the age in billions of years. This formula holds for $-4 \le \log Z \le -1.4$; $0.2 \le Y \le 0.3$; $0.2 \le t_9 \le 25$.

(b) The central density grows as the star evolves towards the tip of the giant branch. Energy loss via neutrinos from the central regions increases, lowering the central temperature slightly. Once the temperature and density allow the $3\alpha \rightarrow {}^{12}C$ to become unstable, the helium flash occurs at a certain radius where $M(\text{ignition}) \simeq 0.2 M_\odot$. The helium flash partly lifts the degeneracy of the core layers outside this radius (Fig. 3.23). There follows
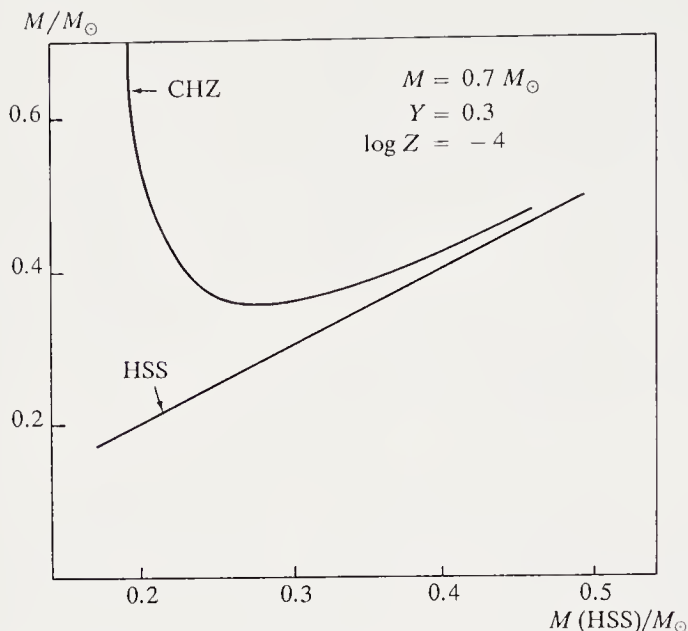
141

Fig. 3.22. Lower boundary of the convective hydrogen zone (CHZ), and the radius of the hydrogen shell source (HSS) around the isothermal core, as functions of the isothermal core mass for a $0.7\,M_\odot$ population II star with $Z = 10^{-4}$
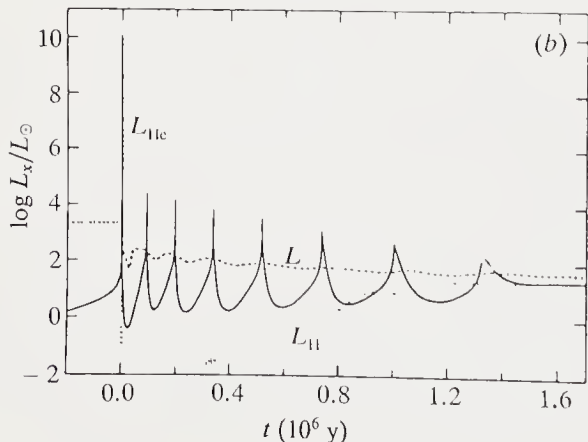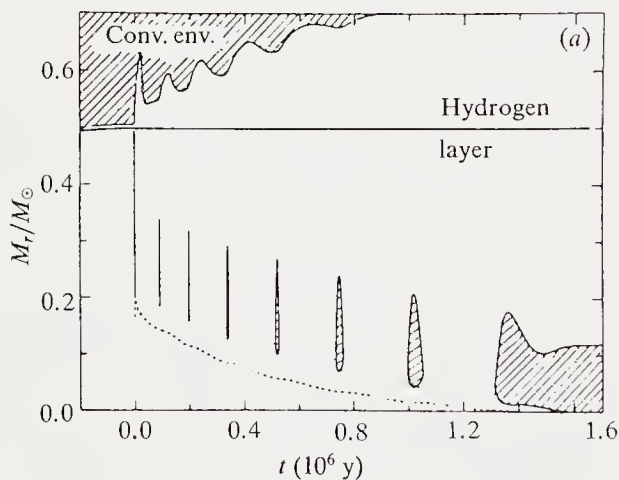

Fig. 3.23. Helium flash, assumed quasi-static, calculated by Mengel and Sweigart. **(a)** Mass of the convective zones (*hatched*) as functions of time; the first convective layer (after the strongest helium flash) is too short ($\sim 10^3$ y) to be resolved; **(b)** surface luminosity L, hydrogen shell luminosity L_H and helium-burning luminosity L_{He} as functions of time. (From I. Iben and A. Renzini, *Physics Reports* **105**, No. 6, 329–406 (1984). Reproduced by kind permission of Elsevier Science Publishers)

a series of flashes until one occurs at the centre and completely lifts the degeneracy (Fig. 3.23). The order of magnitude of the luminosity at the tip of the RGB is given by

$$\log L\,(\text{flash}) = 3.31 + 0.09(3 + \log Z) + 0.45(X - 0.7) - 0.16(M - 0.8).$$

At this point about 5 % of the helium has been transformed into carbon. During evolution to the tip of the RGB moderate mass loss has started. For a $0.8\,M_\odot$ star at the base of the RGB the mass loss is about $0.2\,M_\odot$. The time spent by a star on the RGB above luminosity L is

$$t_{\text{RGB}} = 1.47 \times 10^9\,L^{-0.84}(10^3 Z)^{-0.04}\left(\frac{X}{0.7}\right)^{1.36}$$

$$\times \left(\frac{M}{0.8}\right)^{-0.27}\left(1 - \left(\frac{L}{L_E}\right)^\beta\right)\ \text{years}$$

and a Reimers-type mass loss rate (Sect. 4.3.4) has parameter

$$\eta_R = 0.35\,,$$

which is fully compatible with the observed mass loss.

(c) Core helium burning signals the beginning of horizontal-branch evolution, at the blue side of the RR Lyrae instability strip.

Helium burning in the convective core raises the carbon and oxygen content, and thus the opacity. Defining the edge of the radiative zone by the stability condition $(\nabla_{\text{rad}} - \nabla_{\text{ad}})_e \leq 0$ at the outer boundary of the convective zone, we see that the chemical discontinuity as one crosses towards the interior of the convection zone implies a discontinuity $(\nabla_{\text{rad}})_i - (\nabla_{\text{rad}})_e > 0$ when one crosses the boundary of the convective gone towards interior. This produces a new instability in the radiative region just outside the convective core; a *semi-convective* region arises, with energy transported via radiation, but a very slow adjustment of the chemical composition occurs which maintains the relation

$$\left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\text{ad}} = \left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\text{rad}},$$

or

$$\left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right)_{\text{ad}} = \left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P} - \frac{\mathrm{d}\log\mu}{\mathrm{d}\log P}\right)_{*}.$$

This extra mixing zone ensures a larger nuclear fuel reserve for the star, which stays longer on the HB as a result of this semi-convective instability.

There are various interpolation formulae for the time spent on the HB. From the tables of Sweigart and Gross (1978) we find for t_{HB} in millions of years

$$\log t_{\rm HB} = 2.04 - 3.0(M_c - 0.475) - 0.36(Y - 0.3) + 0.03(3 + \log Z) \,.$$

Using the relation between the helium core mass at the start of the HB and the stellar mass,

$$M_c = 0.4706 + 0.26(X - 0.7) - 0.01(3 + \log Z) - 0.035(M - 0.8) \,,$$

we find

$$\log t_{\rm HB} = 2.053 + 0.42(Y - 0.3) + 0.06(3 + \log Z) + 0.105(M - 0.8) \,,$$

which gives a typical HB lifetime of 100 million years.

(d) After the HB the star moves on to the AGB. Its maximum luminosity is approximately

$$L = 2.5 \times 10^4 (2M_c - 1) \,,$$

where M_c is the core mass on the AGB. The corresponding luminosity function is given by

$$dt_6 = 2.7 d \log L \,,$$

where t_6 is in millions of years.

The maximum luminosity is reached for $M_c = M_*$. Starting from stars with $M_{\rm RGB} = 0.8 \, M_\odot$, we would have $\log L_{\rm max} = 4.2$ without mass loss. Let us examine the consequences for the globular cluster ω Cen, which has about 4000 HB stars. At this luminosity, and a lifetime of 2.4 million years after reaching the AGB (see p. 143), we *should* expect to see about 80 bright stars at the tip of the AGB. We do not see these stars simply because of mass loss. Starting from $0.8 \, M_\odot$ and reaching the HB with $0.6 \, M_\odot$, they have to lose about $0.1 \, M_\odot$ on the AGB. This can be achieved with a Reimers parameter $\eta_R \simeq 0.41$. This result is compatible with $\log L_{\rm max} \simeq 3.2$.

(e) As it makes its excursion to the AGB, the star *may* cross the RR Lyrae instability strip. This depends mainly on the mass of the star (Fig. 3.24). Stars with too low a mass ($M_{\rm HB} < 0.52 \, M_\odot$) do not reach the AGB or the instability strip and evolve towards the white dwarfs. Slightly more massive stars ($0.52 \, M_\odot < M_{\rm HB} < 0.55 \, M_\odot$) cross the instability strip and then move to the blue, crossing it once again before going to the AGB.

Pulsating variable stars with periods less than 8 d (BL Her type) are found in globular clusters which are rich in blue horizontal-branch stars. Their properties show that the excursion to the blue is sufficiently marked that these stars cross the instability strip and spend about a million years there. Variables with periods longer than 12 d (W Vir type) would then correspond to a re-crossing of the instability strip by stars near to the maximum luminosity evolving to the white-dwarf state. The masses of these stars, about $0.5 \, M_\odot$ according to their evolutionary tracks, agree with the masses deduced from the pulsations (see Chap. 6).
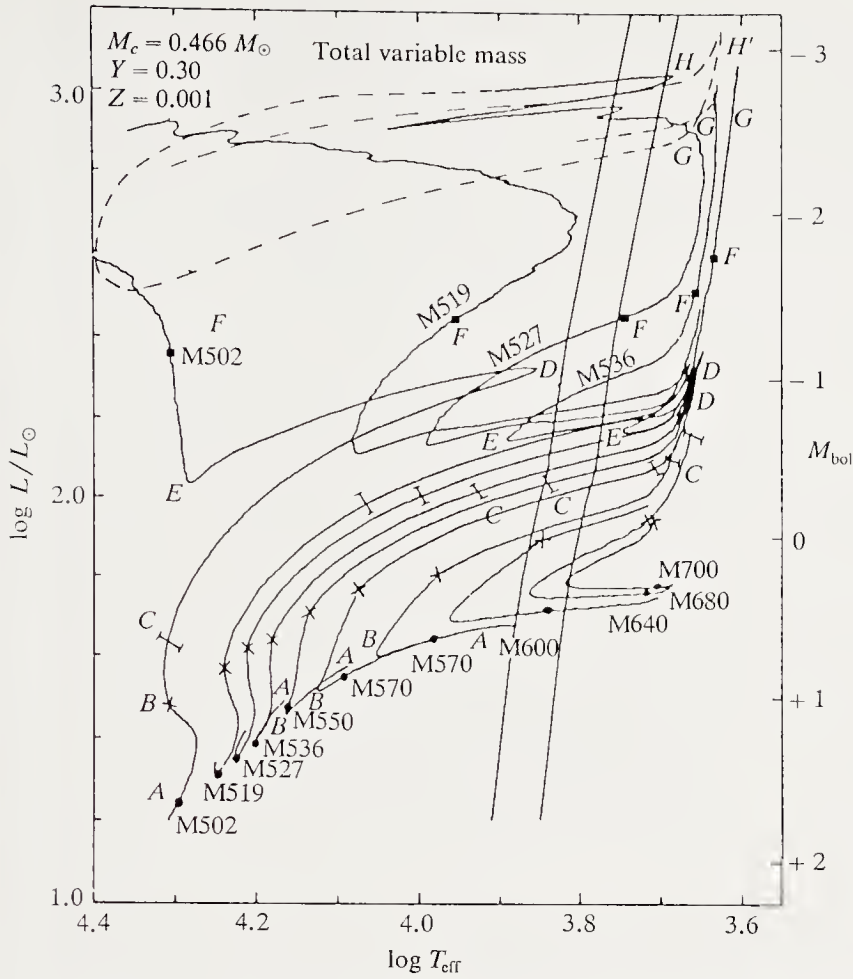
Fig. 3.24. Evolutionary tracks on the HR diagram for population II stars of various masses. The evolutionary stages A–H are explained in the text. (From R.A. Gingold, *Astrophys. J.* **204** 116 (1976). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1976 The American Astronomical Society)

(f) The various stages of evolution labelled A, B, ... in Fig. 3.24 are explained in Table 3.3.

(g) From these data we can find the ages of globular clusters. The most reliable method compares the luminosity at the knee and the HB luminosity for the same effective temperature (here $\log T_{\text{eff}} = 3.85$), avoiding the errors in the theoretical prediction of the colour index $B - V$, the luminosity difference being directly related to age. Using the empirical relation between the luminosity, at the turning point, the HB luminosity and that of the RR Lyrae stars,

$$\log L_T = \log T_{\text{HB}} - 0.40 \Delta M_T^{\text{RR}}(\text{Bol}),$$
$$\Delta M_T^{\text{RR}}(\text{Bol}) = M_T(\text{Bol}) - M_{\text{RR}}(\text{Bol}),$$

we get the age

Table 3.3. Evolutionary stages for population II stars of 0.5 to 0.6 $M_\odot$ (Fig. 3.24)

Letter	Lifetime	Evolutionary stage
A B		Zero-age HB Maximum size of semi-convective zone
	2×10^6 y	
C		Core helium exhaustion Helium burns in a thick shell
	a few times 10^5 y	
D		Core expansion reaches the hydrogen shell source. The luminosity drops and the star's movement to the blue finishes at E
	2×10^6 y	
E		Hydrogen shell burning increases
	8×10^6 y	
F		Luminosity L_{II} passes the helium luminosity L_{He}
	2–4 $\times 10^6$ y	
G		First helium flash. Star evolves in a series of pulses and cycles
H		Continuous curve shows slow loop and dashed curve
	4×10^4 y	the rapid loop corresponding to the last cycle
H'		before evolution towards the white-dwarf state

$$\log t_9 = -0.352 - 1.88(Y - 0.3) - 1.44(Y_{\mathrm{HB}} - Y)$$
$$- 0.088(3 + \log Z) + 0.41\Delta M_T^{\mathrm{RR}}.$$

If we assume $Y_{\mathrm{HB}} - Y = 0.2$, $\log Z = -3.89$ (from the surface abundances), $Y = 0.23$ (perhaps slightly low, but coherent for the system of models used), $\Delta M_T^{\mathrm{R}}(\mathrm{Bol}) = 3.35 \pm 0.2$, we get

$$t_9 = (16 \pm 3.5) \times 10^9 \text{ years}.$$

The compatibility of these evaluations of globular cluster ages with cosmological and abundance data is discussed in the book by F. Combes et al. (1993).

3.7.2 Intermediate-Mass Stars of Population I

The main problem is that of the *critical mass*. As we have seen, stars of mass $M < M_{\mathrm{crit}}$ evolve into white dwarfs after significant mass loss.

We can calculate the evolution using the approximate $L(M_c)$ relations for the RGB and the AGB, and the relation

$$\frac{\mathrm{d}M_{\mathrm{core}}}{\mathrm{d}t} = \frac{Lc^2}{Q},$$

where Q is the energy yield per unit mass of nuclear fuel, together with the mass loss (e.g. Reimers). This gives the final mass ($M_* = M_c$). To satisfy the constraint on the critical mass requires $\eta_R = 4$, about 10 times higher than values encountered earlier, and completely incompatible with observation. In particular, stars of mass less than $1.5\,M_\odot$ would lose their entire envelopes while on the RGB, and helium ignition could not occur. It is possible that a final *superwind* phase, with a very high mass loss for a short time, might resolve this difficulty.

Another problem is the mass of the Cepheids. From the evolutionary tracks (Fig. 3.25) we see that a star can cross the instability strip several times, with a secular period change (increasing or decreasing) given by crossings to the left and right respectively. Cepheids are stars in the phase of core helium burning. During the RGB phase, dredge-up of the deeper layers by convection modified the surface composition, increasing nitrogen and decreasing carbon, revealing the effects of the main-sequence CNO cycle on the central composition.

The period–luminosity relation for the Cepheids has a certain dispersion, according as the pulsation occurs at the first harmonic (extreme blue limit of the instability strip) or the fundamental (blue to red edges of the instability strip). The comparison with theory has long been known to present a problem, in that the mass deduced from the period–luminosity relation (M_{puls}) and that deduced from evolution (M_{evol}) do not agree. Modest mass loss on the RGB as during helium burning, such as would result from a Reimers parameter $\eta_R \simeq 1/2$ (about 10^{-4} to $10^{-5}\,M_\odot$) forces a correction of about 0.2 magnitudes for the Hyades distance modulus above what is given by the convergent-point method. This would bring the two Cepheid mass estimates into agreement, although there are still difficulties for individual Cepheids.

In comparing the problem of the RR Lyrae stars with that of the Cepheids we have to remember that the role of heavy-element opacity differs in metal-poor stars (RR Lyrae) and in Cepheids. A distinct increase (factors of 2–3) of the opacity for elements heavier than helium ("metals") in stellar envelopes would remove the difficulty represented by the conflict between the mass estimate M_{evol} from stellar-evolution calculations and the estimate M_{puls} given by the relation to the period, particularly for beat Cepheids and Cepheids with secondary maxima. For main-sequence stars, and the Sun in particular, the corresponding increase in the opacity would be no more than 10 or 20 % at the base of the convection zone, and very small in the centre, thus being compatible with the constraints on solar models.

The data now coming from *Hipparcos* will doubtless remove the uncertainty over the distance modulus to the Hyades.

3.7.3 Massive Stars ($M \gtrsim 8\,M_\odot$)

Massive stars have intense stellar winds ($10^{-5}\,M_\odot/\mathrm{y}$) and strong mixing effects which can greatly increase the fully mixed mass fraction. We shall

regard the mass loss rate as an empirical parameter. Evolutionary tracks are calculated assuming various mass loss rates and compared with observation so as to estimate the mass loss. Definite conclusions are difficult, because the effects of mass loss and turbulent diffusion are difficult to disentangle.

Study of the evolution of massive stars involves the following parameters:

— *Mass loss rate.* For example Lamers (1981) gives

$$\log \dot{M} = -4.83(\pm 0.28) + 1.42(\pm 0.40) \log \left(\frac{L}{10^6} \right)$$
$$- 0.99(\pm 0.97) \log \left(\frac{M}{30} \right) + 0.61(\pm 0.13) \log \left(\frac{R}{30} \right)$$

and assumes a dependence $\dot{M} \sim (M/R)^{0.6}$. The dispersion in (M/R) appears mainly responsible for the spread in $\dot{M}$.

— *Convective overshooting.* This increases the chemically uniform volume. Maeder (1975) chooses an overshooting depth $(d_{over}/H_P) \simeq 0.25$.

— *Turbulent diffusion.* This also extends the chemically uniform region. Its efficiency is related to the angular velocity of the star, if a parametrisation is required.

In addition the high central temperatures achieved in the course of the evolution oblige us to follow chemical evolution up to the formation of silicon (Fig. 3.26). The elements thus formed are expelled into the interstellar medium by the stellar wind and change its chemical composition. The neon

---

Fig. 3.25. Evolutionary tracks for population II stars of 3, 5, and 7 $M_\odot$ ($Z = 0.001$ and $Z = 0.01$) (without mass loss). The *dashed lines* denote from left to right the blue edge of the first harmonic, the blue edge of the fundamental, and the red edge of the fundamental pulsation in the instability strip. The *tick marks* correspond to intervals 10^6 y (from the main sequence) for 3 $M_\odot$, 2×10^5 y for 5 $M_\odot$, and 10^5 y for 7 $M_\odot$. The ages at the numbered points are given below.

Ages in units of 10^7 y

Points	Z = 0.01 M/$M_\odot$			Z = 0.001 M/$M_\odot$		
	3	5	7	3	5	7
1	0.073	0.0217	0.0155	0.248	0.060	0.017
2	18.791	6.268	3.325	20.793	6.448	3.335
5	20.066	6.442	3.377	24.519	7.146	3.590
7	24.656	–	–	26.850	7.993	3.914
9	–	7.950	3.892	29.729	8.204	–
13					8.276	4.044

(From S.A. Becker et al., *Astrophys. J.* **218**, 633 (1967). Reproduced with the kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1977 The American Astronomical Society)

isotope ^{22}Ne found in cosmic rays was probably formed inside massive stars.

The theory of evolution explains a certain number of observed properties.

(a) *Maximum luminosity of supergiants on the HR diagram.* The brightest stars have absolute magnitude $M_{Bol} = -12$ for type O3; the boundary of the the O-star region moves to decreasing luminosity until type B. The boundary of the supergiant region is then roughly horizontal (Fig. 3.27).

(b) *Statistics of massive stars.* The relative number of red supergiants and Wolf–Rayet stars depends on the mass function and is an indicator of the speed of evolution, depending on the metallicity. The ratio (red supergiants)/(Wolf–Rayet stars) varies from 0.53 in the solar neighbourhood to 23 in the Small Magellanic Cloud, with metal abundances going from 0.03 to 0.003.

(c) *Chemical composition of Wolf–Rayet stars.* These stars (WR) consist of the remnant core of a star surrounded by an expanding envelope with
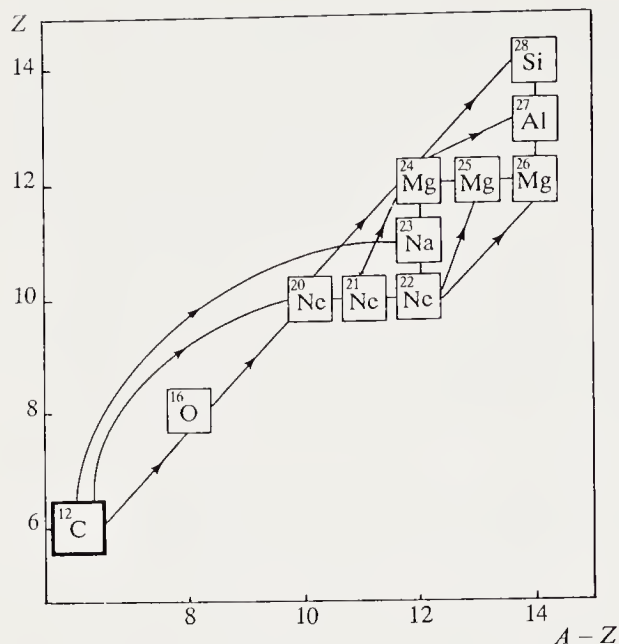

149
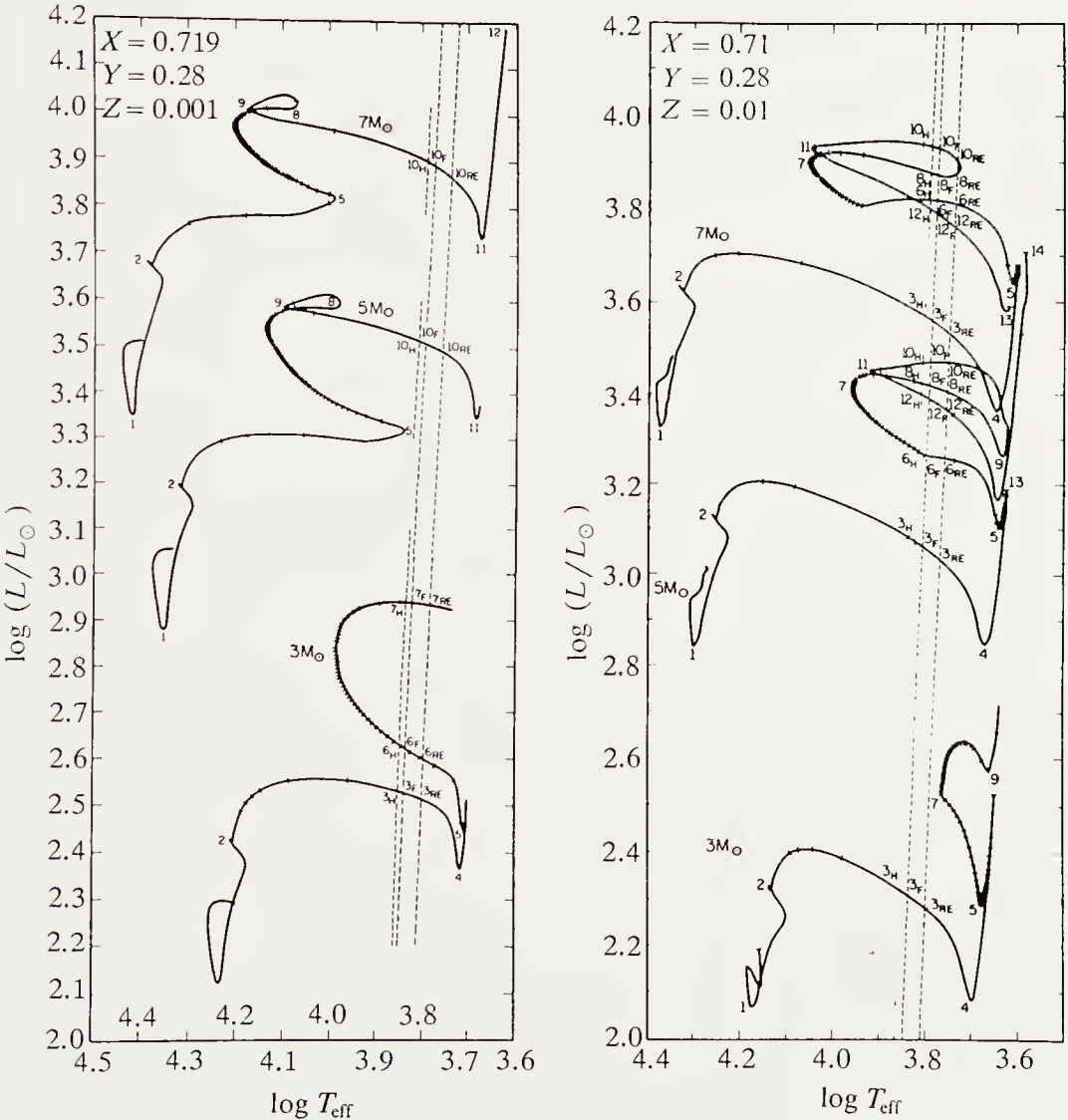
Fig. 3.26. Reactions $X(\alpha,n)Y$, from carbon to silicon. The ordinate is the atomic number Z; the abscissa the number of neutrons $A-Z$ in the nucleus
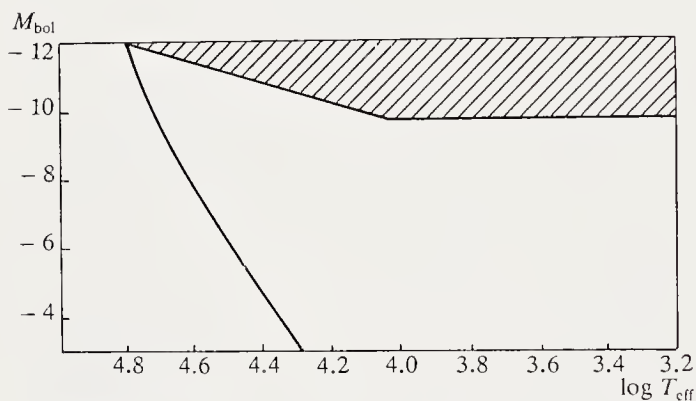

Fig. 3.27. The HR diagram. Main sequence and zone of avoidance

a very large mass loss rate. We distinguish WN stars, in which helium and nitrogen lines are strong, and WC stars, in which helium, carbon, and oxygen lines are strong.

Without going into great detail (Fig. 3.28a) the evolutionary tracks on the HR diagram (Fig. 3.28b) can be classified according to mass.

For $M \gtrsim 40\,M_\odot$

$$\text{BSG}$$
$$\text{O} \rightarrow \text{O}f \rightarrow \textit{Blue supergiants}$$

$$\text{BBV}$$
$$\rightarrow \textit{Bright blue variables} \rightarrow \text{WR} \rightarrow \text{SN}$$

For $40\,M_\odot \gtrsim M \gtrsim 30\,M_\odot$

$$\text{O} \rightarrow \text{BSG} \rightarrow \text{YSG} \rightarrow \text{RSG} \rightarrow \text{WR} \rightarrow \text{SN}$$
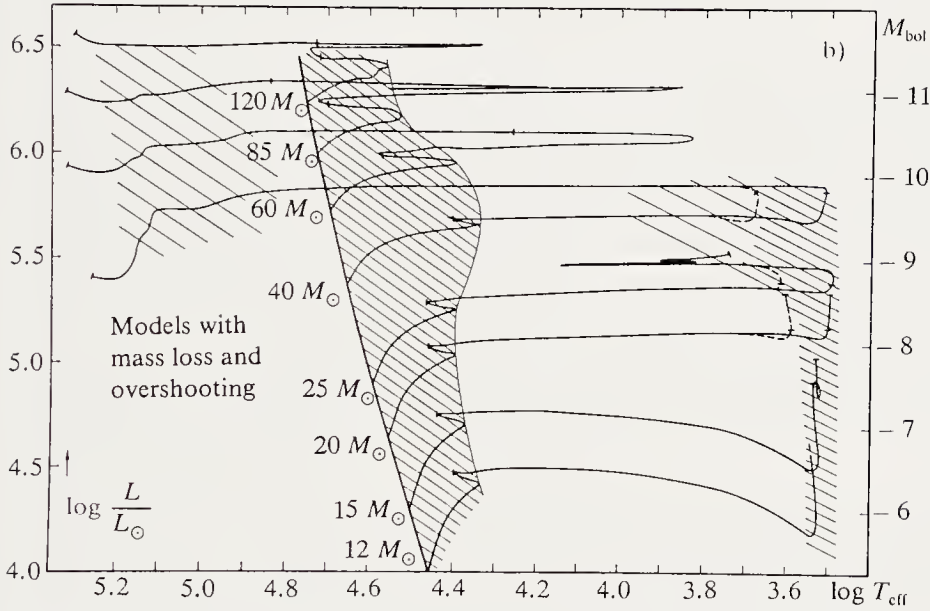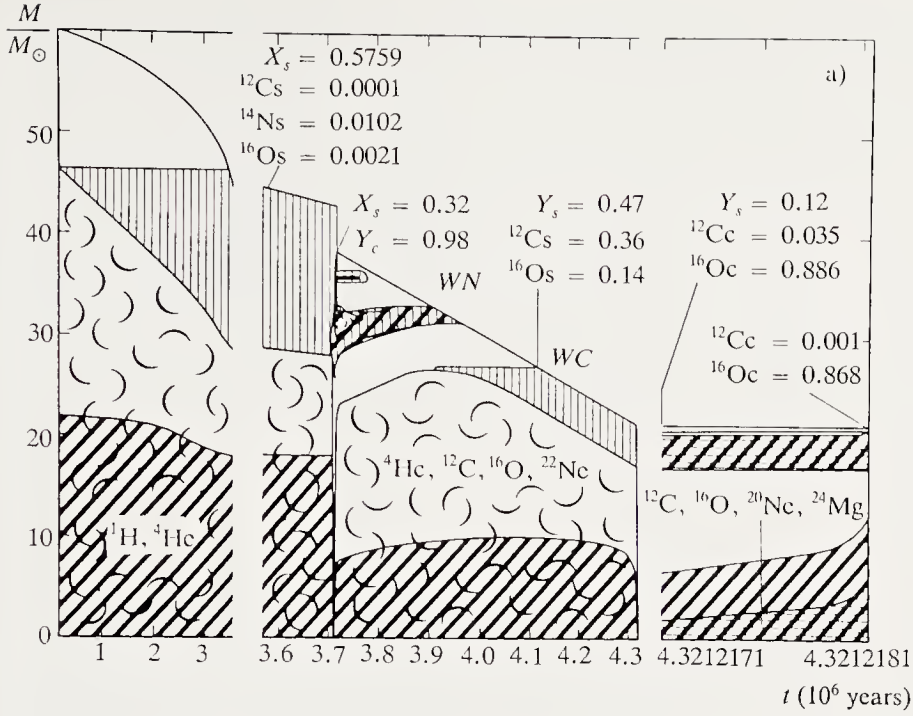
Fig. 3.28. (a) Evolution of the internal structure of a 60 $M_\odot$ star until core carbon exhaustion, with mass loss and convective overshooting. The regions marked with eddies are convective: *heavy diagonals* show regions of strong energy production ($\varepsilon > 10^3$ erg/g/s); *vertical hatching* shows regions of non-uniform H and He composition; and the *horizontal hatching* shows regions of non-uniform ^{12}C, ^{16}O and ^{20}Ne. The surface evolution and observational status of the star are shown: O, WR of types WN then WC. The index s is related to the surface composition, index c to the central composition. (From A. Maeder, *Astron. Astrophys.* **182**, 243 (1987) Fig. 3.) (b) Evolutionary tracks for massive stars ($X = 0.70; Z = 0.02$) with mass loss and convective overshooting. The *hatched regions* show the main-sequence and the helium-burning phase. The first mark on the tracks shows core hydrogen exhaustion and the second core helium exhaustion. (After A. Maeder, *Astron. Astrophys.* **182**, 243 (1987) Fig. 9)

Table 3.4. Lifetime of the various nuclear phases and of the WR phase (in units of 10^6 y, Maeder and Meynet 1987)

M_i	H burning	He burning	C burning	WR
120	2.94	0.51	0.0011	0.51
85	3.32	0.50	0.0018	0.49
60	3.71	0.61	0.0019	0.54
40	4.79	0.64	0.0049	0.49
25	7.09	1.17	0.0072	
20	8.81	1.26	0.0091	
15	12.11	1.63	0.0179	

For $30\,M_\odot \gtrsim M \gtrsim 15\,M_\odot$

$$\text{O} \rightarrow \text{RSG} \left(\begin{array}{c} \text{with or without passage} \\ \text{through the Cepheid instability strip} \end{array} \right) \rightarrow \text{SN}$$

The WN stars correspond to the appearance at the surface of the products of the CNO cycle; the WC stars to the appearance at the surface of alpha-particle reactions ($3\ ^4\text{He} \rightarrow\ ^{12}\text{C}$; $^4\text{He} +\ ^{12}\text{C} \rightarrow\ ^{16}\text{O}$). The lifetimes calculated for the various evolutionary phases give a basis for interpreting the abundances of the different spectral types (Table 3.4).

The rotation parameter introduces varying degrees of turbulent mixing. Above a certain equatorial velocity v_{eq} (~ 325 km s^{-1} for an O star) the mixing speed is larger then the evolution speed. The star remains chemically uniform as it evolves. It evolves to the left of the main sequence and goes directly to the WR(WC) branch, crossing the WN type very rapidly. Study of the association OB Per 1 seems to confirm this bifurcation of the evolution of massive stars, which occurs for $v_{\text{eq}} > v_{\text{crit}} \simeq 35$ km s^{-1}.

The isotopic anomalies of cosmic rays can be explained by assuming that Wolf–Rayet stars inject isotopes of ^{13}C, ^{22}Ne, ^{25}Mg, ^{26}Mg, ^{29}Si, and
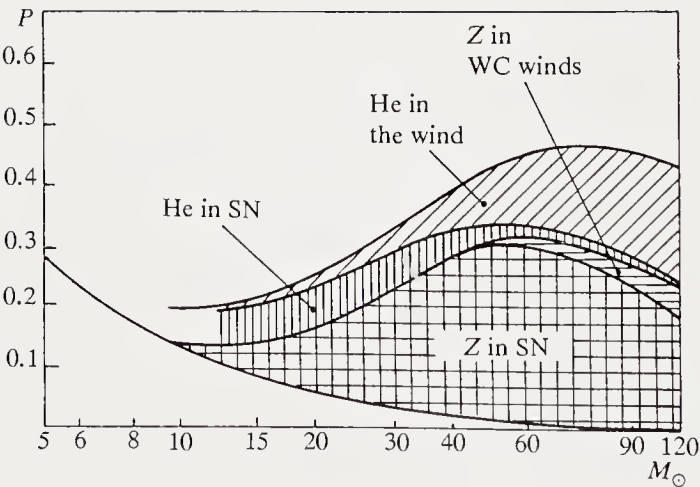

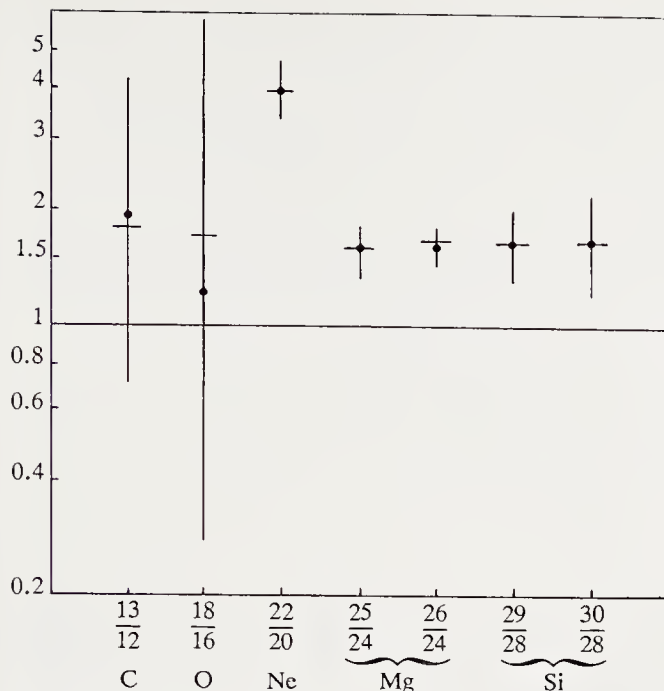**Fig. 3.29.** Yields of various chemical elements (in mass fraction). (After A. Maeder)

Fig. 3.30. Relative abundances of cosmic-ray isotopes

^{30}Si into the Galaxy. We have to take into account gradients of chemical composition and the distribution of WR stars. Using the yields (as mass fractions) for the production of different elements (Fig. 3.29) it is possible to get the results of Fig. 3.30 which appear to give a good explanation of the cosmic-ray anomalies.

3.8 Evolution of Binary Stars

3.8.1 Introduction

Binary stars are so numerous that it is impossible to discuss stellar structure and evolution without taking them into account. According to Jaschek and Gomez (1970), 50 % of main-sequence stars are in binaries; Van Albada (1968) estimates that 60 % of early-type stars are in binaries, with a mass ratio larger than 0.2. Some classes of objects are only found in binaries.

(a) According to Abt (1961, 1965), Abt and Bidelman (1967), and Abt and Moyd (1973) *metallic-line stars* are all in binaries.

(b) *Novae, dwarf novae,* and *cataclysmic variables* are all binaries.

(c) *Wolf–Rayet stars* are frequently in binaries (Underhill 1967; Smith 1968, 1973).

(d) *Runaway stars* have escaped from binaries disrupted by the SN explosion of the companion (Blaauw 1961).

(e) The kinematics of *pulsars* ($\langle v^2 \rangle^{1/2}$ from 100 to 200 km s^{-1} in the Galaxy) suggest that they are produced in the disruption of binary systems. There are now many binary pulsars known, some of them having very eccentric orbits, suggesting the near-disruption of a binary system. There are however several binary pulsars having circular orbits, the secondary star being (or having been) a giant. Some of them have long periods (100 days); three systems with periods less than 10 hours are evaporating their low-mass companions. Several pulsars observed in globular clusters have eccentric orbits and have no doubt been formed by capture.

(f) *Compact X-ray sources* in the Galaxy involve neutron stars (or black holes) in binary systems. We discuss compact stars in binaries in Chap. 8. Here we are interested primarily in gaseous stars.

3.8.2 Structure of Stars in Binaries

The apsidal-motion test in principle offers a way of studying stellar interiors. The deformation of each of the two stars by the other introduces an additional $1/r^4$ attraction which causes a gradual precession of the line of apsides. Schwarzschild (1958) gives the relation

$$\frac{\text{orbital period}}{\text{apsidal period}} = 15 \left(k_1 \frac{R_1^5}{D^5} \frac{M_2}{M_1} + k_2 \frac{R_2^5}{D^5} \frac{M_1}{M_2} \right) \frac{1 + \frac{3}{2}e^2 + \frac{1}{8}e^4}{(1 - e^2)^5} \,,$$

where R_1, R_2, M_1, M_2 are the radii and masses of the two components, and D is the binary separation for small eccentricities e. The constants k_1, k_2 are related to the perturbations $V_{i\,\text{ext}}$ of the potential due to the deformation of star i:

$$V_{i\,\text{ext}} = -\frac{a_{0i}}{r} - \frac{a_{2i}}{r^3} P_2(\cos\theta) \,,$$

$$a_{2i} = 2k_i R_i^5 \frac{GM_{3-i}}{D^3} \,.$$

The approximate expression for k_i (Kopal 1959) is

$$k_i = \frac{16\pi}{5M_i R_i^5} \int_0^{R_i} \rho r^7 \mathrm{d}r \,.$$

This shows that the main contribution to the apsidal motion comes from a region near the surface of the star.

The method is important for theory, but can be applied to only a small number of systems, since we need to know the changes in physical quantities very accurately. For α Vir (Odell 1974), with component masses $10.9\,M_\odot$ and $6.8\,M_\odot$, standard evolutionary models for the $10.9\,M_\odot$ star cannot account for the apsidal precession ($R^5 k$ is too small). Agreement is restored for a star with a $3.5\,M_\odot$ hydrogen-exhausted core and shell burning. This

implies that mass loss has occurred. The procedure used by Monet (1980) involves comparison of two quantities, the observed

$$O = \frac{P^{13/3}(1-e^2)^5}{A\left(1+\frac{5}{4}e\right)},$$

where the orbital period P is expressed in days and the apsidal period A in years, and the calculated quantity

$$C = 0.2773 \frac{(1+16f)R_1^5 k_1 + 16f^{-1}R_2^5 k_2}{M_1^{5/3}(1+f)^{5/3}},$$

where k_1, k_2 are the apsidal motion constants defined above, and $f = M_2/M_1$ the mass ratio. Models with mass loss seem to give a value of $R^5 k$ compatible with observation.

Tides. In a binary system with eccentricity e the companion of mass M_2 produces a potential which varies in star 1 (mass M_1). The potential producing the tide oscillates with circular frequency ω and can be expanded in spherical harmonics

$$U = \sum_{lmn} U_n^{lm} \left(\frac{r}{R}\right)^n P_n^m(\cos\theta) \exp[i(\sigma_{lm}t - m\varphi)].$$

If the star's spin frequency is Ω with respect to an inertial reference system, and if the external potential rotates with the angular frequency ω, the circular frequency of the orbital motion, the excitation frequency is

$$\sigma_{lm} = l\omega - m\Omega.$$

The perturbed potential is the sum of the external potential and the internal potential due to the perturbed star. The space dependence of the external potential must make it harmonic and vanish at infinity. It thus has the form

$$\Phi = \Phi_n^{lm}(r/R)^{-n-1} P_n^m(\cos\theta) \exp[i(\sigma_{lm}t - m\varphi)].$$

In linear approximation the tide problem reduces to finding the external potential resulting from a single component of the perturbing field. In practice the problem involves finding the amplitudes Φ_n^{lm} as functions of the tidal frequency σ_{lm}. These coefficients were found by Zahn (1966, 1970, 1975) for low tidal frequencies.

Following Cowling (1941), Zahn distinguishes between static and dynamical tides. The static tide is that part of the distortion which is static when viewed by an observer comoving with the binary. In the presence of dissipation the static tide is out of phase with the line of centres and produces a torque which vanishes only when the orbit is circular and the spin and orbital periods are equal. The synchronisation and circularisation times are given by Zahn (1977, 1978) as

$$t_{\text{sync}} = \frac{1}{6q^2 k^2} \frac{I}{MR^2} \left(\frac{D}{R}\right)^6 t_F,$$

$$t_{\text{circ}} = \frac{1}{21q(1+q)k_2} \left(\frac{D}{R}\right)^8 t_F,$$

where D is the binary separation, R, M the radius and mass of the star considered, and q the ratio M_2/M with M_2 the companion mass. I is the axial moment of inertia of the star of mass M, k_2 its apsidal constant, and t_F the viscous dissipation timescale. The two expressions assume small eccentricity e. We discuss t_F in more detail below.

The dynamical tide results from resonances between the frequency of a gravity wave (g-mode) of the star and the excitation frequency σ_{lm}. The dynamical tide has an amplitude which depends on time and occurs only when the binary is eccentric or not synchronous. *Radiative dissipation* produces damping. The tidal potential excites a range of gravity waves in the star. In the absence of dissipation these waves would be reflected at the star's surface and produce standing waves in phase with the tidal potential. But in the outer layers of the star the radiative cooling time becomes comparable with the tidal period, and the waves are damped. The gravity waves are partially reflected at the surface and phase-shifted, and a net flux of mechanical energy is transported from the interior adiabatic regions to the surface. This energy transfers angular momentum from the star's spin to the orbital motion.

For the static tide the friction due to turbulent viscosity depends on the region, convective envelope, or core. The phenomenological nature of theories of convection means that we can only obtain an order of magnitude.

In convective envelopes, where the tide has a large amplitude, the frictional time t_F is found by assuming that the dissipated power L is entirely removed by convection; then we have

$$t_F \propto (MR^2/L)^{1/3}.$$

For convective cores, where the amplitude is small, the frictional time is much longer, by a scaling factor $(R/R_c)^7$, where R_c is the radius of the convective core. In this case there is another correction because the timescale of the turbulence may be longer than the tidal period. As a result stars with convective cores are synchronised and circularised by radiative damping. The timescales for these processes are

$$t_{\text{sync}}^{-1} = 5.2^{5/3} \left(\frac{GM}{R^3}\right)^{1/2} \left(\frac{MR^2}{I}\right) q^2 (1+q)^{5/6} E_2 \left(\frac{R}{D}\right)^{17/2}$$

and

$$t_{\text{circ}}^{-1} = \frac{21}{2} \left(\frac{GM}{R^3}\right)^{1/2} q(1+q)^{11/6} E_2 \left(\frac{R}{D}\right)^{21/2},$$

Table 3.5. Dynamical tide constants[a]

M_*	Radius	(R_c/R)	I/MR^2	E_2
1.6	1.152	0.1251	13.21	2.41 (E-9)
2	1.294	0.1594	12.97	1.45 (E-8)
3	1.678	0.1854	12.58	4.72 (E-8)
5	2.350	0.2143	11.98	1.53 (E-7)
7	2.919	0.2379	11.49	3.80 (E-7)
10	3.652	0.2669	10.95	1.02 (E-6)
15	4.672	0.3054	10.43	3.49 (E-6)

[a] Stellar models by M. Aizenman

where $q = M_2/M$ and E_2, the tidal coefficient, is very sensitive to the star's structure. E_2 varies as $(R_c/R)^8$, and is thus very small for convective cores. E_2 is related to the apsidal constant for the dynamical tide. In Table 3.5 we give some results of Zahn (1975), leading to characteristic timescales for the main sequence (Table 3.6).

Zahn's discussion shows that (1) for $1\,M_\odot$ stars with convective envelopes, synchronisation is achieved within the age of the Sun for $D/R < 40$, which seems to be confirmed by observation; (2) Synchronisation for stars with convective cores is achieved for $D/R \simeq 7$ to 8, close to the results of Table 3.6.

There are stars which spin much more rapidly than the orbital frequency (AR Cas, U Cep), with separations $D/R \simeq 5$ much smaller than the critical value of Table 3.6. It may be that mass transfer from the secondary to the primary has caused this in U Cep; but there is no obvious explanation for

Table 3.6. Effects of the dynamical tide[a]

M_*	Synchronisation		Circularisation	
	$(D/R)_{sync}$	P_{sync} (days)	$(D/R)_{sync}$	P_{circ} (days)
1.6	6.11	1.21	4.44	0.75
2	7.05	1.59	4.99	0.95
3	6.81	1.92	4.85	1.10
5	6.52	2.19	4.68	1.33
7	6.72	2.69	4.80	1.62
10	6.67	3.30	4.77	2.00
15	7.04	3.98	4.99	2.38

[a] The table gives the parameters describing the damping of the dynamical tide by radiative dissipation in stars with a convective core and radiative envelope. The parameters D/R, P_{synch}, P_{circ} are calculated for synchronisation and circularisation times equal to 1/4 of the main-sequence lifetime t_{nuc}. The calculations are for $M_1 = M_2$. For a different mass ratio one has to multiply $(D/R)_{sync}$ by $q^{4/17}((1+q)/2)^{5/51}$ and $(D/R)_{circ}$ by $(q/2)^{2/21}((1+q)/2)^{11/63}$ taking account of one star only.

AR Cas (a very young system?). Of course disagreement with the periods P_{sync} of Table 3.6 could simply be due to the models used (calculated in 1974).

3.8.3 Mass Transfer

Theoretical Principles. The problem is complex, and some simplifying assumptions are listed by Paczyński (1971):

(a) Stars in close binaries are regarded as spherical, even if they fill their Roche lobes.

(b) The Roche-lobe radius is taken as the critical radius for mass loss. This ignores mass loss in a wind. The stellar radius is defined as that of the photosphere. This is satisfactory for dwarfs and giants, but not for stars with extended envelopes (e.g. supergiants).

(c) We assume circular orbits in the present problem. The study of elliptical orbits involves many difficulties and is not considered here.

(d) The star's radius is assumed less than or equal to the Roche-lobe radius. In other words the mass loss keeps the star within its critical radius. This assumption fails in two (ill-understood) cases: (1) the primary has a deep convective zone at the beginning of mass loss; (2) both Roche lobes are filled, the stars being in expansion.

(e) The star is in hydrostatic equilibrium. This assumption fails in a small region near the surface, and also for primaries with a deep outer convective layer; it fails for secondaries with very high accretion rates and for contact systems where both stars expand beyond the Roche lobes.

(f) The total mass M and angular momentum are conserved during the evolution. This assumption ignores the possibility of mass loss from the system and of electromagnetic braking. For cataclysmic variables the orbit evolves under gravitational radiation losses (Chap. 8). In some cases mass loss from the system is essential to the evolution.

Modes of Evolution. Many calculations of binary evolution have been made. Three initial parameters, the masses M_1, M_2, and the separation D of the stars, define the system. During the evolution the mass loss rate is another parameter. A complete discussion of all possible binary evolutionary tracks is not yet available. The choice of initial parameters has been motivated by observation, in particular the desire to explain the formation of special systems (white dwarfs or neutron stars in X-ray binaries). It is impossible to describe all the possible cases. We therefore confine ourselves to presenting the physics of mass transfer and giving some examples of binary evolution.

We can identify modes of mass tranfer using the mass–radius diagram plotted by Webbink (1979). There are two basic ideas: (1) we assume $M_1 = M_2$, but assume that only the primary evolves; (2) we compare the
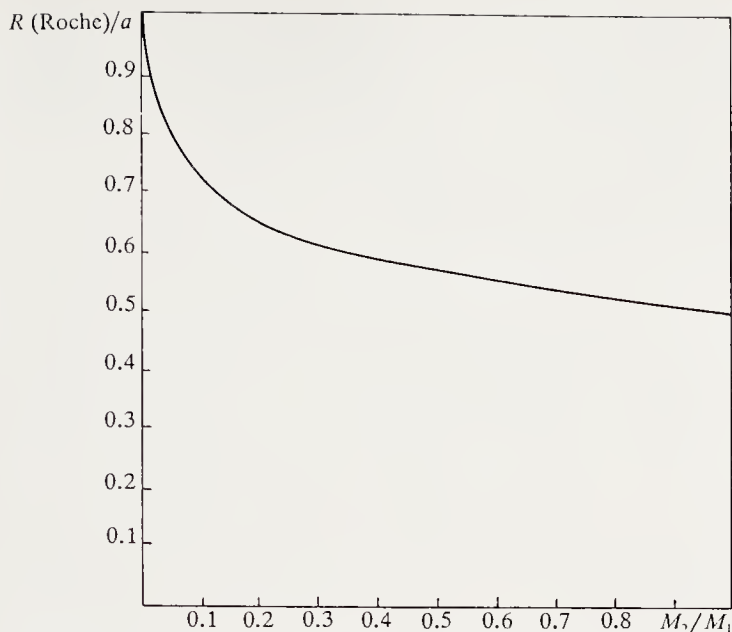
Fig. 3.31. The Roche-lobe radius as a function of mass ratio

primary's radius with that of the Roche lobe (Fig. 3.31). For $M_1 = M_2$ we have $R(\text{Roche}) = a/2$ with a the semi-major axis of the orbit. This transforms the mass–radius diagram (Fig. 3.32) into a mass–period diagram (Fig. 3.33) and gives a classification of modes of evolution (Fig. 3.34).

(a) *Mode I: contact.* These are close binaries of low angular momentum and zero age which are in physical contact, such as W UMa stars. Despite the fact that these are relatively unevolved systems (e.g. TX Cnc in Praesepe) the two components are very far from the main-sequence mass–luminosity relation; they have almost the same surface temperatures and a mass ratio $\sim 1/2$. Their presence in old galactic clusters and their high space density suggest that they are secularly stable.

High-mass contact binaries are observed. For these systems helium ignition in the primary occurs immediately on leaving the main sequence and takes it on to the "helium main sequence".

(b) *Mode II: mass loss from a radiative envelope.* This was the first type of mass transfer through the inner Lagrange point to be studied.

Once the primary radius exceeds the Roche-lobe radius the mass loss soon reaches a stationary regime. Mass flows out at the sound speed through a nozzle defined by the condition that the advection term v_s^2/l should exceed the Coriolis force,

$$\frac{v_s^2}{l} \gtrsim v_s \Omega \, ,$$

so that

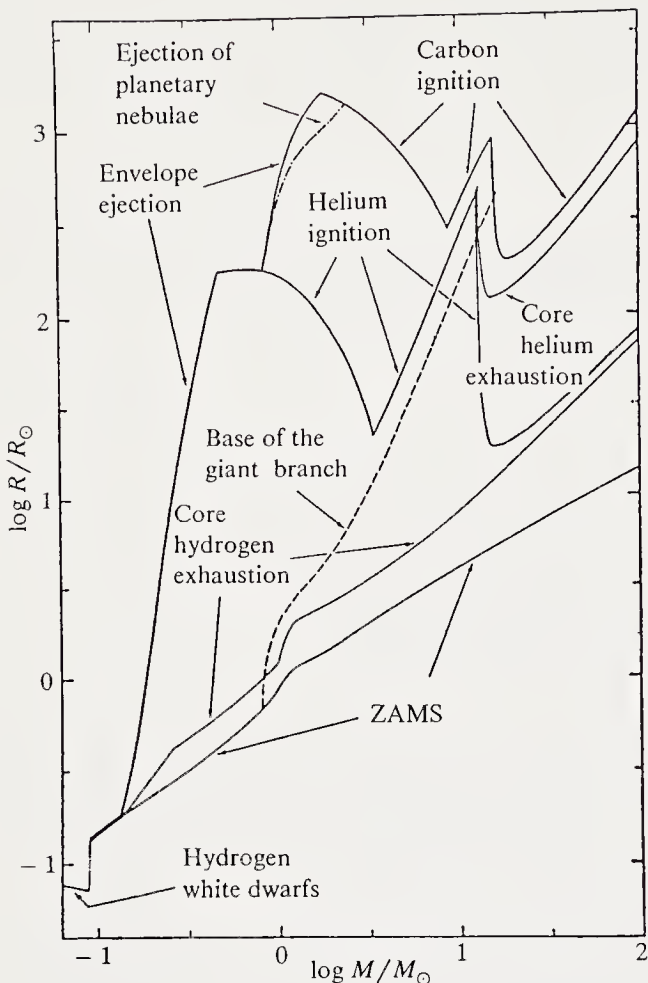$$\dot{M}_{\max} \cong \rho v_s^3 \Omega^{-2} \, .$$

Fig. 3.32. Radii of solar-composition stars at various stages of their evolution, as functions of their masses

To order of magnitude this corresponds to mass loss on a Kelvin–Helmholtz timescale: $\dot{M} \simeq -M/t_{\mathrm{KH}}$. It is actually somewhat shorter. We have

$$\frac{\tau_{\mathrm{rad}}}{\tau_{KH}} \simeq \frac{4\pi\sigma\mu^4 v_s^5}{\rho_{\mathrm{ph}} k^4} \simeq \frac{3.85 \times 10^{-6}}{\rho_{\mathrm{ph}}} \left(\frac{v_s}{10\,\mathrm{km\,s^{-1}}}\right)^5 ,$$

where μ is the mean molecular mass, ρ_{ph} is the photospheric density and τ_{rad} is the actual mass-loss timescale.

Mass loss continues at the rate $\dot{M}_{\mathrm{max}}$. Once the mass ratio is reversed the Roche-lobe radius begins to increase and the mass loss reduces simply because the star has to expand. In radiative equilibrium the star's specific entropy increases outwards in the envelope,

$$\frac{\mathrm{d}S}{\mathrm{d}r} = \frac{k}{\mu} \frac{1}{H_p} \left(1 - \frac{\Gamma_{\mathrm{rad}}}{\Gamma^*}\right) > 0 ,$$

so that as the outer layers are lost, heat has to be supplied to the new surface layers to re-establish thermal equilibrium. If the mass loss is sufficiently fast ($\tau_{\mathrm{dyn}} < \tau_{\mathrm{KH}}$), this energy deficit cannot be filled immediately. Thus the new

160
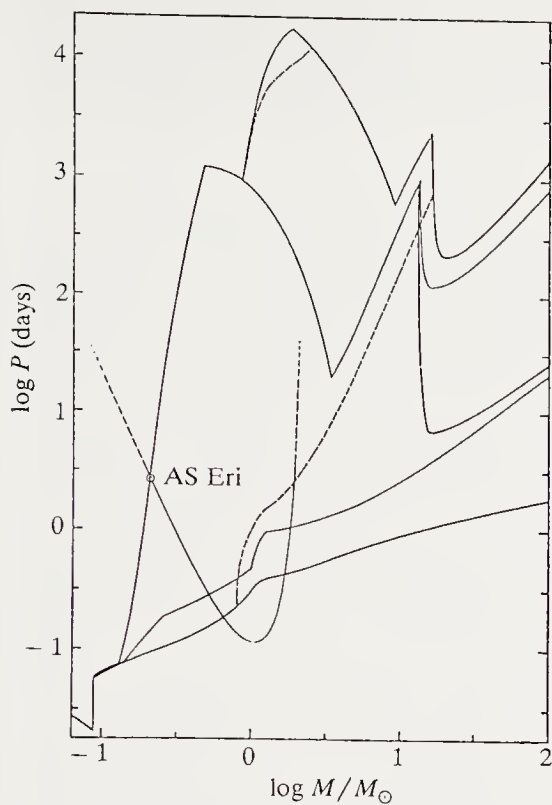
Fig. 3.33. In a binary with equal-mass components, the evolution of one star causes an instability defined by the evolutionary state (Fig. 3.32). The period corresponds to the radius given on Fig. 3.32
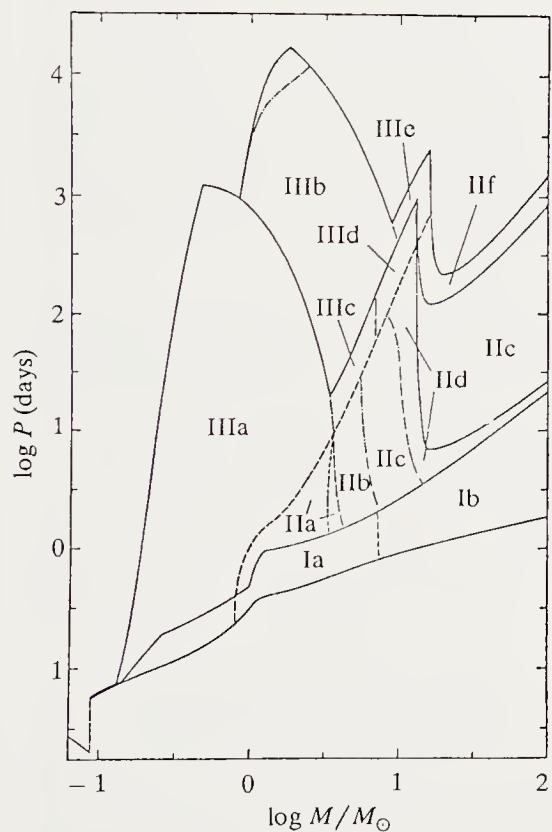

Fig. 3.34. Identification of the main modes of mass transfer in the period–mass diagram. (Symbols explained in the text)

161

outer layers are cooler and denser than an equilibrium star of the same mass: the mass-losing star becomes undersized and underluminous for its mass. Binaries following this evolution lose mass as they cross the Hertzsprung gap.

There are several sub-modes:

— *IIa* $M_{\text{primary}} < 3.5\,M_\odot$. Mass loss on the Kelvin–Helmholtz timescale reverses the mass ratio. Mass transfer then proceeds on the nuclear timescale as the degenerate core grows, up to the point where the primary's envelope is almost entirely exhausted. These systems, called *semi-detached*, can be identified without doubt as Algol-type systems. This idea is supported by studies of Algol-type binaries and statistical arguments. However, quantitative comparison of theory and observation implies that up to $50\,\%$ of the angular momentum may be lost.

In the lifetime of the Galaxy the system becomes *detached*, and consists of a helium white dwarf of mass $0.21\,M_\odot \lesssim M_{WD} \lesssim 0.46\,M_\odot$, and a main-sequence star of a few solar masses.

— *IIb* $3.5\,M_\odot \lesssim M \lesssim 7.6\,M_\odot$. These stars have a helium core which grows until helium burning starts. The star contracts and the system becomes *detached*. The final state is probably a carbon–oxygen white dwarf.

— *IIc* $5\,M_\odot \lesssim M \lesssim 14\,M_\odot$. For such stars there is a second phase of mass transfer after core helium exhaustion, because the remnant stars of 1–$2.6\,M_\odot$ expand during helium burning. After the second mass transfer the remnant is a carbon–oxygen white dwarf of mass 1–$1.4\,M_\odot$.

— *IId – f* $14\,M_\odot \lesssim M$. The helium core left by these stars is too massive for helium burning to produce a degenerate carbon–oxygen core (the central temperature is too high). These stars expand less than those of *IIc* and there is no second stage of mass transfer. The remnant stars have masses up to $1.9\,M_\odot$ (*IId*) to $4\,M_\odot$ (*IIe–f*) and end as collapsed stars (neutron star or black hole).

(c) *Mode III: mass loss from a convective envelope.* In a convective envelope the specific entropy decreases towards the surface. In contrast to the radiative case, it is not necessary to provide energy to re-establish equilibrium. These stars tend to expand once mass loss starts.

We follow the discussion of Paczyński and Simkiewicz (1972). Near the inner Lagrange point L_1 we assume that:

— the flow is adiabatic:

$$P = K\rho^{1+1/n} \, ;$$

— the flow is slow enough to be regarded as quasi-stationary;
— the interior of the mass-losing star is close to hydrostatic equilibrium.

We can then use the Bernoulli equation

$$\tfrac{1}{2}v^2 + K(n+1)\rho^{(1/n)} + \Omega = \Omega_s \, ,$$

where v is the flow velocity and Ω_s the surface potential.

At the inner Lagrange point we define coordinates x along the line of centres and y orthogonal to the line of centres. The total mass flux is

$$F = 2\pi \int_0^{y_F} \rho v_x y \, \mathrm{d}y \,,$$

where y_F is the boundary of the outflow and v_x the low velocity along the x axis. To estimate the mass flux to order of magnitude we seek its maximum value $F_{\max}$. We thus assume that ρv is constant, so that

$$\mathrm{d}(\rho v) = \rho \, \mathrm{d}v + v \, \mathrm{d}\rho = 0 \,.$$

Differentiating the Bernoulli equation gives

$$v \, \mathrm{d}v + K \frac{n+1}{n} \rho^{-1+1/n} \mathrm{d}\rho = 0 \,,$$

so that

$$-v^2 + K \frac{n+1}{n} \rho^{1/n} = 0 \,,$$

and the Bernoulli equation becomes

$$K \frac{n+1}{n} \left(\frac{1}{2} + n \right) \rho^{1/n} = \Omega_s - \Omega \,,$$

leading to the mass flux

$$(\rho v)_{\max} = \left(K \frac{n+1}{n} \right)^{-n} \frac{(\Omega_s - \Omega)^n}{\left(n + \frac{1}{2} \right)^n} \,.$$

The flux thus depends on the potential. Expanding the potential to lowest order near Ω_0, the potential at the Roche lobe and inserting in the expression for F,

$$F_{\max} = 2\pi A^2 \frac{(\Delta \Omega_s)^{n+3/2}}{\Omega_0 h (K(n+1))^n} N(n) \,,$$

where $\Delta \Omega_s$ is the difference $\Omega_s - \Omega_0$, A the binary separation, and h the coefficient of the first-order term in y^2 near L_1. The function $N(n)$ varies slowly with polytropic index n:

$$N(n) = \frac{n^n}{(2n+1) \left(n + \frac{1}{2} \right)^{n+1/2}} \,.$$

For $n = 3/2$, $N(n) = 0.0766$. The difference $\Delta \Omega_s$ is of the order of (GM_1/R). If we further assume that the star's luminosity is of the order of $4\pi r^2 \rho v_s^3$, where v_s is the sound speed at the photosphere, we find that

$$\dot{M}_{\max} \simeq -M(\tau_{KH}\tau_{\text{hydr}})^{-1/2} ,$$

where τ_{hydr}, the hydrostatic equilibrium timescale, is the time taken for a sound wave to cross the star.

The mass loss rate from a convective envelope may be very large (up to $10^{-3}\, M_\odot\, \text{y}^{-1}$), which prompts the questions of whether the companion can accept the mass and whether some mass is lost from the system. We can distinguish the following cases:

— *Mode IIIa.* Binaries with periods below 1000 d and low masses begin mass transfer before helium ignition in the degenerate core. The remnant star is a white dwarf of 0.2 to $0.5\, M_\odot$.

— *Mode IIIb.* Binaries with longer periods and larger masses begin mass transfer after the primary reaches the AGB, leaving white dwarfs with masses between 0.5 and $1.4\ M_\odot$.

— *Mode IIIc.* Stars between 3.5 and $7\ M_\odot$ give a helium star with $0.5\, M_\odot \lesssim 1.4\, M_\odot$, leaving a carbon–oxygen white dwarf.

— *Mode IIId.* For stars above $7\, M_\odot$ the helium star is heavier, $1.4\, M_\odot$–$4\, M_\odot$, leading finally to a neutron star.

3.9 Evolution to the Main Sequence

Once the optical depth of a contracting molecular cloud exceeds 1 we can regard it as a star. Its energy comes from gravitational contraction, and the characteristic timescale, determined by the rate of heat loss, is the Kelvin–Helmholtz timescale

$$\tau_{\text{KH}} \simeq \frac{GM^2}{LR} .$$

We can study the contraction by assuming hydrostatic equilibrium (i.e neglecting accelerations) while allowing for gravitational energy release. From (3.5), written here for $\gamma = 5/3$,

$$\text{div}\, F = -\frac{3}{2}\frac{\Re\rho T}{\mu}\frac{\dot{R}}{R} .$$

This approximation is valid once the sound crossing time in the star is smaller than the contraction timescale:

$$\frac{R}{R_\odot} \ll 490\left(\frac{5\,000}{T_{\text{eff}}}\right)^{8/9}\left(\frac{M}{M_\odot}\right)^{5/9} .$$

Initially the star is fully convective and follows a near-vertical trajectory in the HR diagram, called the Hayashi track; this is almost parallel to the ascending giant branch but covered in the reverse direction.
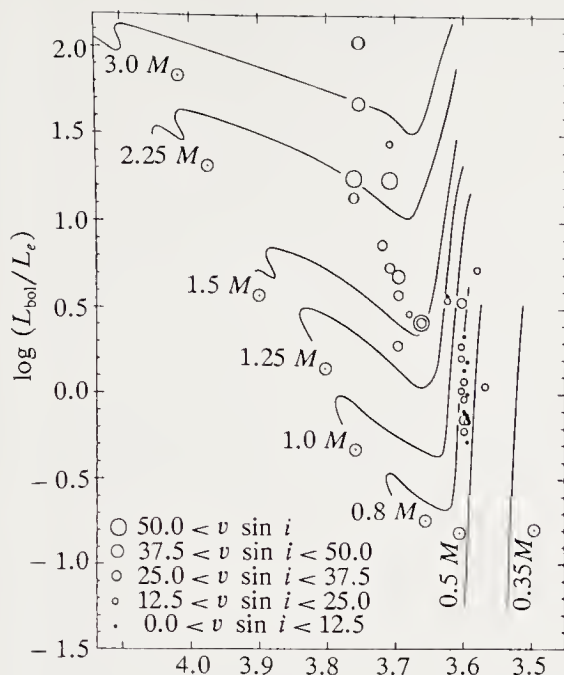
Fig. 3.35. Pre-main-sequence evolutionary tracks and the observed T Tauri stars. (From J. Bouvier et al., *Astron. Astrophys.* **165**, 110 (1986))

Once the star is near the main sequence the opacity in the central regions decreases and they become subadiabatic. A radiative core forms (at least for masses 0.5 to 3 $M_\odot$). As the contraction continues thermonuclear reactions begin, slowing the contraction until the gravitational terms become negligible. Numerical calculations give evolutionary tracks as shown in Fig. 3.35. These tracks correspond to the radii and effective temperatures of very young stars of T Tauri type.

4. Mass Loss and Stellar Winds

4.1 Introduction

We discuss here mass loss and winds from *isolated* stars. In contrast to binary systems, where the presence of the companion causes mass loss, we have here to seek *internal* causes for the mass loss in winds.

In the present incomplete state of the theory of winds, it is sensible to describe first the observational data and the different types of wind; this description is inseparable from the physics of the stellar wind, particularly its thermodynamics and the mechanism of spectral-line formation. We then discuss the successes and failures of current stellar-wind theory. The implications of mass loss for stellar evolution are given in Chap. 3.

4.2 Observational Data: General Remarks

Mass loss provides both direct and indirect data. *Direct data* are spectroscopic, showing the presence of expanding material around stars (see Chap. 1). We distinguish between different sources of information:

— Circumstellar lines (metals, CaII, MgII) blue-shifted with respect to the stellar lines;

— Infrared excesses (above the continuum corresponding to the star's effective temperature), characteristic of circumstellar envelopes;

— Radio emission (OH, SiO masers, continuum);

— P Cygni profile in UV resonance lines such as MgII, C IV, Si IV, N V, O VI or in strong subordinate lines such as H_α.

Trying to use these data to find the mass loss rate $\dot{M}$ is fraught with uncertainties such as the evaluation of the optical depth, and above all estimates of the ionisation degree of the relevant element. These give rise to large differences in the mass-loss estimates published for the same stars by different authors. The various diagnostic methods are given in Sect. 4.3.

Indirect data come from comparison of stellar-evolution theory with observation.

— The first indirect data came from white dwarfs. Study of white dwarfs in open clusters suggests (see Sect. 4.4.1) a total mass loss which may reach 85 % for masses 8 or 9 $M_\odot$.

— Difficulties in matching data for the horizontal branch and the asymptotic giant branch and the position of the RR Lyrae stars are resolved by taking account of mass loss.

— The origin of the Wolf–Rayet stars can be understood by following evolution including mass loss.

All stars show evidence for stellar winds, particularly the Sun. It is customary to reserve the term *mass loss* for cases where the timescale $M/\dot{M}$ is comparable to (or shorter than) the nuclear timescale. Here we study stellar winds: the associated mass loss is basic to the study of stellar evolution.

Some elementary theoretical considerations allow an initial classification of mass-loss processes. The momentum-conservation equation for a stationary spherically symmetric flow is

$$V\frac{dV}{dr} + \frac{GM}{r^2} + g_T + g_R + g_M = 0\,,$$

where V is the mean velocity, G the gravitational constant, M the stellar mass (the mass of the wind material is negligible, $M_{\text{wind}}/M_* \lesssim 3 \times 10^{-9}$), and r the radial distance. The term $g_T = (1/\rho)(dP/dr)$ represents the acceleration due to thermal pressure, g_R the acceleration due to radiation pressure, and g_M the acceleration of mechanical origin (hydrodynamic (HD) and magnetohydrodynamic (MHD) waves). We can then distinguish three regimes:

(1) When $g_T > g_R, g_M$ the wind has a thermal origin. This mechanism was the first proposed (Parker 1958) and regarded as the cause of the solar wind.

(2) When $g_R > g_T, g_M$ the wind is driven by radiation pressure. In hot stars radiation pressure is exerted through the UV resonance lines of abundant ions and many subordinate lines. In cool stars radiation pressure can also be exerted on dust grains.

(3) When $g_M > g_T, g_R$ the wind is produced by HD or MHD wave pressure.

Clearly this classification says nothing about the sources of the various terms (heating, temperature distribution, energy balance, wave production) which determine the mass loss. Moreover it ignores structural effects due to the presence of magnetic fields. It is however helpful in drawing distinctions between yellow giants, red giants, and red supergiants (types K–M) on the one hand, and blue supergiants (spectral types O–B) on the other. The mass loss rate is expressed as

$$\dot{M} = \frac{dM}{dt} = -4\pi r^2 \rho(r)V(r)\,.$$

4.3 Direct Data on Winds

4.3.1 The Sun

The solar environment is well known, from the chromosphere and the corona to more distant regions which have been studied since 1960 using spacecraft. Observations at all wavelengths show the heterogeneity of these outer regions. In particular closed magnetic structures are observed, extending through the lower corona and containing very hot gas. X-ray images (*Skylab* 1973) show these structures on a scale of 1/10 of the solar radius and large darker regions, *coronal holes*, where the plasma forming the solar wind appears to escape from. Clearly spherical symmetry is an extremely crude approximation for these outer layers.

Present estimates of the mass loss rate from the Sun are $\dot{M} = -(2.2 \pm 0.6) \times 10^{-14} \ M_\odot \ \mathrm{y}^{-1}$. The wind velocity at the Earth's orbit is on average about 400 km s^{-1}, but it can reach 700 km s^{-1}.

4.3.2 Red and Yellow Giants and Supergiants

Observations of the resonance lines of Ca II, and (from space) those of Mg II, show the presence of absorption components shifted towards shorter wavelengths with respect to the stellar lines (Fig. 4.1). The observed profile may be of P Cygni type. In general, however, the lines are simply asymmetric in bright red supergiants and the coolest red giants. Finding the density ρ requires a model of the expanding envelope.

We find $\dot{M}$ as follows. Using spherical transfer theory, the P Cygni profile gives the *column density* $\sigma = \int_{R_i}^{\infty} \rho(r)\mathrm{d}r$ (where R_i is the inner radius of the expanding region), the *wind velocity* (10 to 50 km s^{-1}), and the rms *turbulent velocity* (several km s^{-1}) in the wind. The mass loss rate $\dot{M} = -4\pi r^2 \rho(r) V(r) = -4\pi R_i \sigma V$ is proportional to the radius R_i. The resulting mass loss rate is uncertain by an order of magnitude since the inner radius R_i is unknown ($R_i \simeq 10^{1\pm1} R(\mathrm{star})$).

The few accurate determinations of mass loss rates come from analysing red-giant circumstellar lines (Ti II, Fe II, Sr II) seen in absorption in the spectrum of a companion with higher T_{eff}. The lines are produced by expanding material from the giant and seen on the line of sight to the companion. The distance of the expanding gas from the giant is known. For example in α^1 Her, $-\dot{M} = 1.1 \times 10^{-7} \ M_\odot \mathrm{y}^{-1}$; for α Sco A, $-\dot{M} = 7 \times 10^{-7} \ M_\odot \ \mathrm{y}^{-1}$.

We shall see in Chap. 7 that warm solar-type giants show high-temperature ions (N V, C IV, Si IV), indicating the presence of a transition zone surrounded by a corona observable through its X-ray emission. There is a clear demarcation in the HR diagram between yellow giants (spectral types earlier than K1 III), which show high-temperature ions, but no optical evidence for a wind, and red giants, where *none* of these lines are visible and all the characteristics of a cool wind are seen.
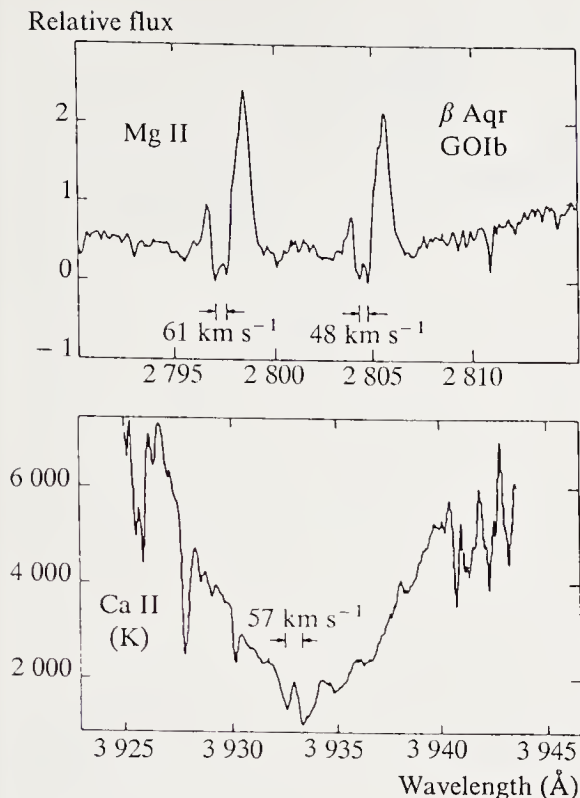
Fig. 4.1. The profile of the h and k lines of Mg II and K of Ca II (the supergiant β Aqr, Dupree 1980). There is a clear correspondence between the absorption profiles of the magnesium and calcium lines from the circumstellar envelope. The magnesium lines are asymmetric. Theoretical calculations show that for stationary mass loss the asymmetry is larger for magnesium lines than calcium lines. To find the terminal velocity requires a line-formation model. (From C. Chiosi and R. Stalio (eds.), *Effects of Mass Loss on Stellar Evolution*, Reidel 1981, p. 96. Reproduced by kind permission of Kluwer Academic Publishers)

Circumstellar envelopes are also shown up by wide infrared emission bands in the regions 11 μm and 18-20 μm from all stars in the upper right-hand corner of the HR diagram. This comes from silicate grains at temperatures of a few hundred degrees. Infrared photometry gives the contrast $(\Delta F_\nu / F_\nu)$ of the emission peak at 11 μm. We can deduce the column density of dust, and from this estimate the hydrogen column density. The resulting mass loss rate again depends on the choice of inner radius for the envelope.

The OH maser line at 18 cm has been seen in many red giants. This line is emitted only by molecules on the line of sight, and the presence of two peaks separated by about 20 to 50 km s^{-1} in radial velocity is due to emission from layers in front and behind the star. We can estimate the radial distance by the time delay of the maser emission compared with variations of the central star, and together with estimates of the optical depth this gives the mass loss rate. With a radius of 10^{16} cm and a velocity of 20 km s^{-1} one finds mass-loss rates from 5×10^{-6} $M_\odot$ y^{-1} to 3.5×10^{-5} $M_\odot$ y^{-1}.

Using solar and other data we give in Fig. 4.2 a schematic picture of the mass loss regimes for cool stars on the HR diagram.

4.3.3 Hot Stars (O, B, WR)

There are four methods, which all depend slightly on the model used to get from the spectrum to the mass-loss rate.
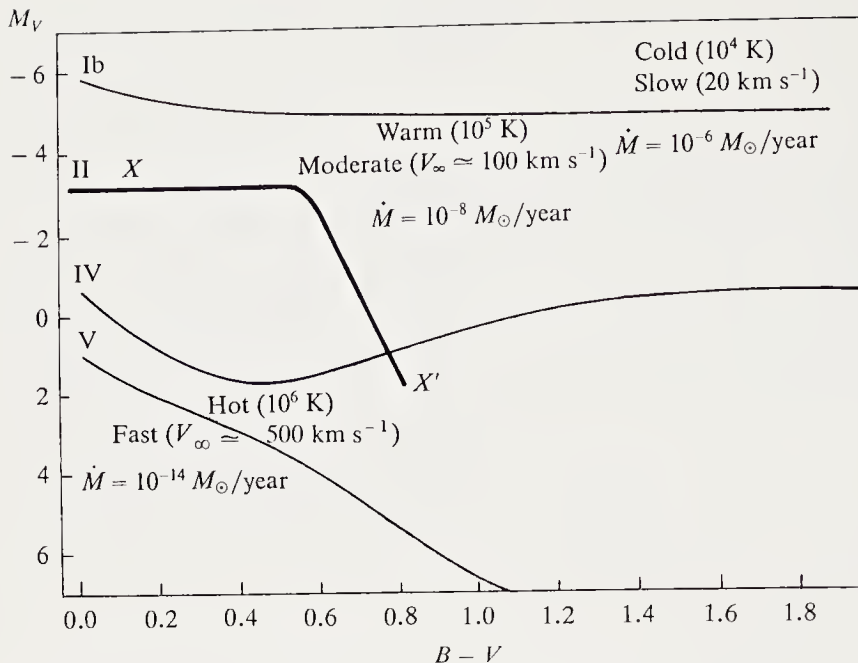
Fig. 4.2. Schematic mass loss regimes in the HR diagram for cool stars. The ordinate is the absolute visual magnitude and the abscissa the $B - V$ colour index. XX' is the Linsky–Haisch dividing line (Sect. 4.6 and Fig. 7.16, where it is called the coronal dividing line)

The UV Method. This uses the P Cygni profile of ultraviolet lines obtained by satellite observations. This is quite sensitive and can detect rather low mass loss rates (down to $10^{-10} M_\odot$ y^{-1}).

The resonance lines of the ions O VI, N V, C IV, Mg II have a broadened absorption component shifted to the blue, with a sharp boundary at a limiting velocity V_∞ of the order of 2000 to 3000 km s^{-1}, called the terminal velocity. These resonance lines are not saturated in O and B stars, but can be in A supergiants. The best diagnostic for these P Cygni lines is to synthesise them by solving the transfer equation in spherical symmetry (cf. Sect. 4.3.2). This assumes that we know a priori the temperature and velocity distributions $T(R), V(R)$ in the wind.

The Sobolev approximation simplifies this method; it is valid for flows with large velocity gradients dV/dr. Then at a given wavelength corresponding to velocity V, scattering in a resonance line occurs over a distance $\Delta r \simeq \Delta\lambda(c/\lambda)(\mathrm{d}V/\mathrm{d}r)^{-1}$, corresponding to a width $\Delta\lambda$ (natural + turbulent Doppler). We thus get the optical depth at wavelength $\lambda(1 + V/c)^{-1}$

$$\tau_r(r) = \frac{\pi e^2}{m_e c} f\lambda_0 n_i \left(\frac{\mathrm{d}V}{\mathrm{d}r}\right)^{-1},$$

which is called the Sobolev optical depth, where f is the oscillator strength, λ_0 the rest wavelength, n_i the number density of absorbing ions, and dV/dr the velocity gradient. The mass loss rate is
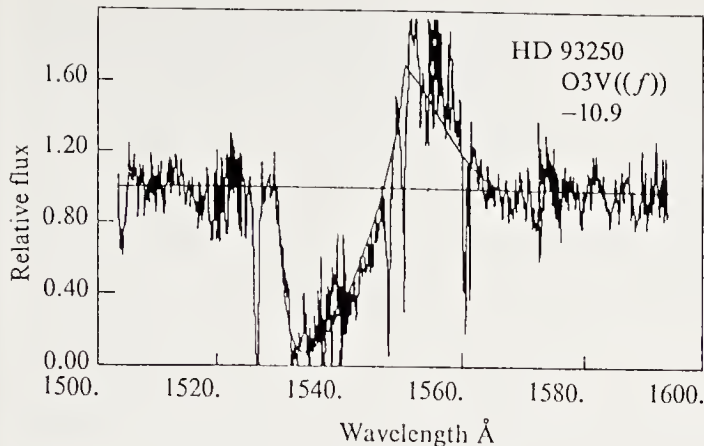
Fig. 4.3. P Cygni profile of the resonance doublet of C IV. There is good agreement with the model of Conti and Garmany. The star is HD 93250, of type O. (From C. Chiosi and R. Stalio (eds.), *Effects of Mass Loss on Stellar Evolution*, Reidel 1981, p. 7. Reproduced by kind permission of Kluwer Academic Publishers)

$$-\dot{M} = 4\pi r^2 V(r)\mu m_H \frac{n_i(r)}{g_i Z_i}\,,$$

where g_i is the relative abundance of the ion i, Z_i the abundance by number of the element with respect to hydrogen, and μ the mean molecular mass. Substituting for n_i using τ we have

$$-\dot{M} = \frac{4\pi \mu m_H}{f\lambda_0} \frac{m_e c}{\pi \mathrm{e}^2} \frac{[r^2 V(\mathrm{d}V/\mathrm{d}r)]}{g_i(r)Z_i} \tau_r(r)\,.$$

We shall return later to the determination of the function $V(r)$. For the moment we note that near $V = V_\infty/2$ the combination $r^2 V\,\mathrm{d}V/\mathrm{d}r$ is fairly insensitive to the function $V(r)$ itself. We thus evaluate $\tau(r)$ at $v = V_\infty/2$; the most delicate part of the determination of $\dot{M}$ is the calculation of g_i, since the ions such as C^{3+}, N^{4+}, Si^{4+} whose lines are studied are not always the most abundant of their element. To find n_i it is enough to fit a phenomenological profile with an estimate of τ_r to the P Cgyni profile (Fig. 4.3).

The Optical Method. This uses strong subordinate transitions such as $H\alpha$ or He II λ 4686. It is essential to allow for non-LTE line formation and find the velocity law $V(r)$ by trial and error.

The Infrared Method. This assumes that the infrared emission from the wind is due to free–free transitions by hot electrons. We measure the infrared excess above the stellar continuum.

To show how the procedure works we consider the case of a stellar wind of uniform velocity V_∞. The mass loss rate is then

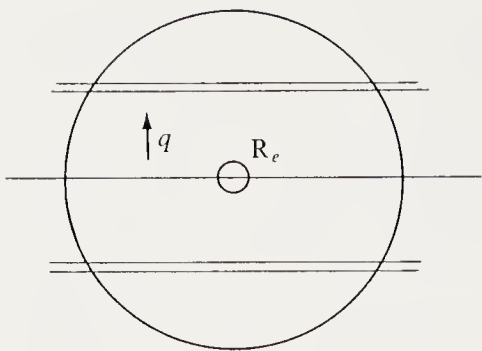$$-\dot{M} = 4\pi r^2 n\mu m_H V_\infty\,,$$

171

Fig. 4.4. Emission from a stationary spherical wind. R_e is the inner radius of the spherical layer. We calculate the flux from the cylindrical layer of radius q and integrate from R_e to infinity

where r is the distance from the centre of the star, μ the mean molecular mass, and m_H the hydrogen-atom mass. This can be written as

$$n = \frac{-\dot{M}}{4\pi\mu m_H V_\infty}\frac{1}{r^2} = \frac{A}{r^2},$$

so that the gas density decreases as the inverse square of the distance. We assume that the plasma begins to expand at a radius R_e. Consider a cylindrical layer of radius q with axis parallel to the line of sight in the wind (Fig. 4.4). If T is the plasma temperature, the intensity of the radiation from this layer is

$$I(\nu, T) = B(\nu, T)(1 - e^{-\tau(q)}),$$

where $\tau(q)$ is the optical depth along the line of sight,

$$\tau = \int_{-\infty}^{+\infty} \kappa_\nu \rho \, dl.$$

The absorption coefficient $\kappa_\nu \rho$ goes as the square of the plasma density:

$$\kappa_\nu = K(\nu, T)n^2.$$

Ignoring the contribution from the hollow sphere of radius R_e, as is legitimate for large mass loss rates, we have

$$\tau = \frac{\pi}{2}K(\nu, T)\frac{n_e}{n_i}\frac{A^2}{q^3}.$$

If D is the distance to the object, the total flux received by the observer is

$$S_\nu = \frac{B(\nu, T)}{D^2}\int_0^\infty (1 - e^{-\tau(q)})2\pi q \, dq.$$

Thus

$$S_\nu = 1.33 \times 2\pi \left(\frac{\pi}{2}\right)^{2/3}\left(\frac{n_e}{n_i}\right)^{2/3}\frac{A^{4/3}}{D^2}B(\nu, T)K^{2/3}(\nu, T).$$

172

Table 4.1. Comparison of mass loss rates for O stars ($\log(-\dot{M})$) in solar masses per year)

Star	Spectral type	$\log(-\dot{M})$			
		Radio	I.R.	Optical	U.V.
ζ Pup	O4 ef	-3.4	-5.2	-5.1	-5.2
9 Sgr	O4	-4.6			-5.6
HD 14947	O4 f		-5.3	-5.1	
Cyg OB2 *9	O5 f	-3.9		-5.0	
Cyg OB2 *5	O7 f	-4.7		-4.7	
ξ Ori	O9.5 I	-5.6	-5.7	-5.5	-5.6

Expressing the flux S_ν in Jansky (10^{-26} W m^{-2} Hz^{-1}) and the frequency ν in Hertz, the mass loss rate is

$$-\dot{M} = 0.095 \frac{\mu V_\infty}{Z_i(n_e/n_i)^{1/2}} \frac{S_\nu^{3/4} D^{3/2}}{g^{1/2}\nu^{1/2}} M_\odot/\text{year},$$

where g is the Gaunt factor and Z_i the mean ionic charge.

The shape of observed spectrum corresponds to free–free emission. The accuracy in the determination of $\dot{M}$ (within a factor of 2) is limited by the distance estimate.

The Radio Method. The free–free radio emission comes from a region where the terminal velocity V_∞ has been reached. This velocity is known from UV observations for almost all stars where radio emission has been detected. As for the infrared method, the deduced mass loss rate is then independent of the velocity law, but still depends on distance.

Table 4.1 compares estimates of $\dot{M}$ by these different methods. Stars of the same luminosity can differ in mass loss rate by a factor of 10 (Fig. 4.5) or even 100 (see Table 4.1).

4.3.4 Interpolation

In evolution theory it is useful to have approximate parametric representations of the mass loss rate as $\dot{M}(L, M, R)$, where it is assumed that no other parameters enter. For *giants and supergiants* of types K and M the mass loss rate appears to be well represented by taking its mechanical luminosity as a constant fraction of the stellar luminosity:

$$\frac{GM\dot{M}}{R} = -\text{const} \times L.$$

In solar units (Reimers 1977)

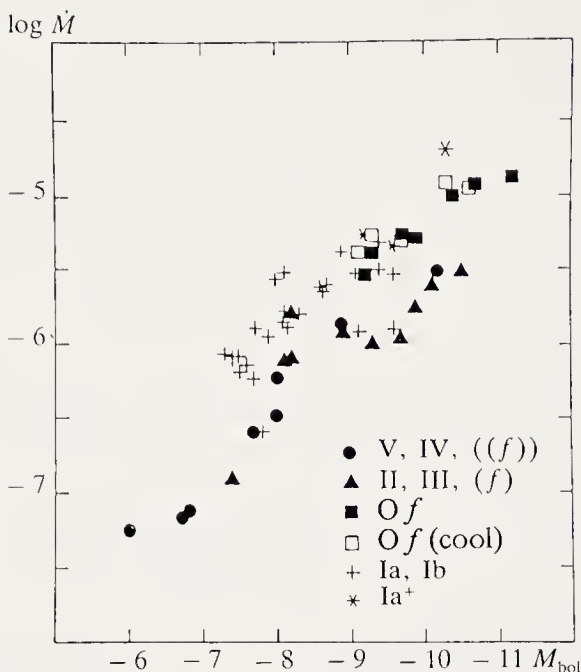$$\dot{M}_* = -4 \times 10^{-13} \eta_R \frac{L_*}{g_* R_*},$$

Fig. 4.5. Mass loss rate as a function of bolometric magnitude for 53 O and B stars. The roman numerals I, II, III, IV, and V give the *luminosity class*. We note the larger dispersion for stars brighter than $M_{bol} = -8$. (From C. Chiosi and R. Stalio (eds.), *Effects of Mass Loss on Stellar Evolution*, Reidel 1981, p. 20. Reproduced by kind permission of Kluwer Academic Publishers)

Legend:
- V, IV, $((f))$
- ▲ II, III, (f)
- ■ Of
- □ Of (cool)
- + Ia, Ib
- ∗ Ia$^+$

where g_* is the gravity, $\dot{M}_*$ is in solar masses per year, and η_R is a numerical factor, $1/3 \lesssim \eta_R \lesssim 3$. F supergiants seem to have smaller mass loss rates, and for the Sun, $\eta_R = 1/20$.

The mass loss rates for O and B stars seem to be fitted by the relation (Lamers 1981)

$$\dot{M} = 1.74 \times 10^{-13} \frac{L_*^{1.42} R_*^{0.6}}{M_*} .$$

It increases as a larger power of the luminosity than in cool giants and supergiants.

4.4 Indirect Data on Winds

4.4.1 White Dwarfs

The presence of white dwarfs in galactic clusters can only be understood if there has been significant mass loss during evolution. A white dwarf cannot have a mass bigger than about $1.4\,M_\odot$, while the stars at the tip of the cluster main sequence have masses of several $M_\odot$.

There are various ways of estimating the progenitor masses for the white dwarfs in open clusters. Table 4.2 gives the clusters for which the mass loss can be evaluated.

Table 4.2. Mass of white-dwarf progenitors

Cluster	n_{WD} [a]	$M_{WD\,min}$ [b]	$M_{WD\,max}$ [b]
Hyades	13	2.7	4.5
Pleiades	1	5.25	8
Sirius group	4	4	7
61 Cyg	3	1	(.)
γ Leo	1		
NGC 2168	4	5	7
NGC 2287	8 ± 5	4	7
NGC 2422	1	6	—
NGC 2516	3	6	11
NGC 6633	10 ± 5	$\geqslant 4$	5

[a] Number of white dwarfs in the cluster.
[b] Minimum and maximum masses of the parent star.

The method is as follows:

— From the *knee* in the cluster HR diagram we can find the age t_A of the cluster.

— From the mass and effective temperature of a white dwarf we can find the time t_{WD} since it became a white dwarf.

— The difference gives the cluster age at which the white-dwarf progenitor left the main sequence, t'_A,

$$t'_A = t_A - t_{WD}.$$

— The cluster epoch t'_A gives the mass of stars then leaving the main sequence to evolve into white dwarfs.

The main difficulty in the method is finding which part of the cooling age t_{WD} is the time spent passing to the giant branch (RGB), the horizontal branch, the asymptotic giant branch (AGB) and the cooling to the present state. In young clusters the latter may be only a small part of the evolution time. The results of Weidemann and Koester (1983) are summarised in Fig. 4.6. They suggest that the maximum mass of the progenitor is

$$M_{WDmax} = 8 \pm \tfrac{3}{2} M_\odot.$$

Nuclear Transformations. An independent method was used by Hills and Dale (1973). During its evolution a star burns a certain fraction f_H of its hydrogen. The mass M_c of the hydrogen-depleted core is the maximum mass of a white dwarf $M_f = M_c$ formed from a star of initial mass M_i. The assumption is that mass $M(1 - f_H)$ was lost during the formation of the white dwarf. If Q is the energy yield from unit mass of hydrogen, the integral $\int L \, dt = X M f_H Q$ gives the fraction f_H of hydrogen consumed. The time interval dt is proportional to the number of stars dN in the luminosity
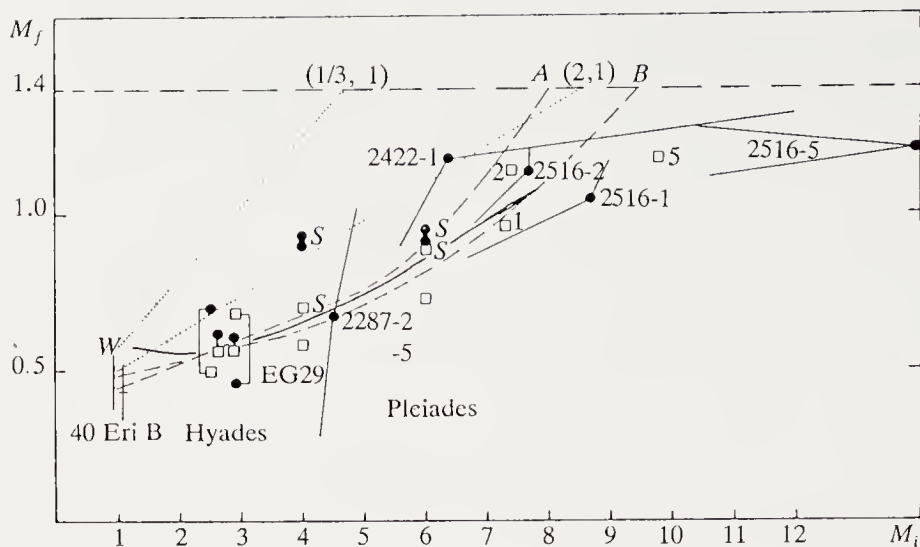
Fig. 4.6. The relation between initial and final masses for white dwarfs (from D. Koester and V. Weidemann). The *dotted* and *dashed curves* correspond to different theoretical models for the mass loss rate. For stars denoted by a *square* masses are deduced from the surface gravity, for *heavy dots* the final masses are deduced from the radii; the error bars for the stars of NGC 2516, 2422, and 2287 are given their extreme values. The *solid curve* is that of Hills and Dale (1973)

interval $d \log L$. Star counts per magnitude interval near the knee give an estimate of the element dt and thus $\int L dt$. The result is plotted in Fig. 4.6 and lies within the extremes of the curves proposed by Weidemann and Koester (1983).

4.4.2 Supernovae

The notion of a critical mass was first introduced for supernovae. Here it is defined as the lower limit above which all stars undergo supernovae at sufficiently late stages of evolution. For stars evolving *without mass loss*, every star of mass above about $2.5\,M_\odot$ has a core which eventually enters the region of thermal instability of the $^{12}\text{C}-^{12}\text{C}$ reaction in strongly degenerate conditions. We can then estimate the supernova frequency in our Galaxy as a function of critical mass (Table 4.3), using the star formation rate and initial–mass function.

Table 4.3. Frequency of SN II supernovae

Lower mass limit $(M_\odot)$	Frequency (y^{-1})
2.5	0.16
2.9	0.10
5.1	0.04
7.4	0.02
9.1	0.01

Two observational points are crucial: (1) the observed frequency of SN II (Sect. 8.5) in our Galaxy is less than 0.02 per year; (2) the explosion of a medium-mass star injects about $1\,M_\odot$ of elements near the iron peak into the interstellar medium. This would require the presence in the Galaxy of an enormous amount of iron, which is not observed. It follows that the critical mass is about $8\,M_\odot$.

4.5 Generation of Mass Loss

We could view the problem of mass loss as a problem of the outer layer of stars, or of the *stellar wind*. We could also see it as a problem of internal structure: what is the origin of the energy which drives the mass loss. In fact these two questions are related, and it is not possible to separate them completely. The mechanism of mass loss is ill understood. The mass loss rate can differ greatly from one star to another, and it can vary in time also. As well as the global parameters mass, luminosity, radius, and chemical composition, there is at least one other (probably the stellar rotation, via the dynamo effect) which controls the mass loss rate.

4.5.1 The Sonic Point

We can illustrate one of the main characteristics of stellar winds by considering the simple case of a thermal wind with $T = \text{constant}$. The steady-state equations of motion and mass conservation are

$$V \frac{dV}{dr} = -\frac{GM}{r^2} - \frac{1}{\rho}\frac{dP}{dr} \tag{4.1}$$

and

$$-\dot{M} = 4\pi r^2 \rho V \,, \tag{4.2}$$

where V is the flow velocity. These imply

$$\frac{dV}{dr} = \frac{\dfrac{2}{r}\dfrac{\Re T}{\mu} - \dfrac{GM}{r^2}}{V - \dfrac{1}{V}\dfrac{\Re T}{\mu}} \,. \tag{4.3}$$

This differential equation has a singular point, defined by

$$\begin{cases} r_s = \dfrac{1}{2}\dfrac{GM\mu}{\Re T} \\[2mm] V_s^2 = \dfrac{\Re T}{\mu} \end{cases} \,. \tag{4.4}$$

It is called the *sonic point* because there the flow speed is equal to the isothermal sound speed of the gas. Solutions of (4.1) and (4.2) with velocities finite at infinity and tending to zero at the stellar surface cross the sonic point. Before the sonic point ($r < r_s$) there is hydrodynamic causality between the star's surface and the flow, since $V < V_s$; by contrast, above this point perturbations do not affect the wind, since $V > V_s$. For the Sun, with coronal temperature T, $r_s = 5.4(10^6/T)R_\odot$.

4.5.2 Energy Constraints

Even though stellar winds are definitely not spherically symmetric, the study of the general steady-state, spherically symmetric case illustrates the basic physics of stellar winds with heat and momentum sources.

We list the hydrodynamical equations again:

— mass conservation:

$$-\dot{M}_* = 4\pi r^2 \rho V = \text{const} \,;$$

— momentum conservation:

$$V\frac{dV}{dr} = -\frac{1}{\rho}\frac{dp}{dr} - \frac{GM_*}{r^2} + D \,; \tag{4.5}$$

— energy conservation:

$$\frac{1}{\gamma-1}PV\frac{d}{dr}\ln\left(\frac{P}{\rho^\gamma}\right) = -\frac{1}{r^2}\frac{d}{dr}(qr^2) + Q \,; \tag{4.6}$$

where ρ, V, P, q are the density, velocity, thermal pressure, and heat flux respectively, and D and Q are the volume injection rates of momentum and energy. The physical processes responsible for D and Q may bring in other equations to close the system.

We wish to use observations to estimate the importance of effects producing non-zero D, Q (e.g. heat transport, dissipation, momentum transfer, etc.). We consider the energy balance. The energy flux of the wind includes advection of kinetic energy, enthalpy and gravitational potential energy, and heat transferred by conduction. Integrating (4.5) and (4.6) we get

$$F = (-\dot{M})\left(\frac{1}{2}V^2 + \frac{\gamma}{\gamma-1}\frac{P}{\rho} - \frac{GM_*}{r}\right) + 4\pi qr^2 \,, \tag{4.7}$$

which we can write as

$$F = F_0 + F_A \,, \tag{4.8}$$

where F_0 is the energy flux of the wind at a reference level r_0, here taken as the base of the wind, and F_A the contributions of the momentum and heat sources,

178

$$F_A = \int_{r_0}^{r} dr' ((-\dot{M})D + 4\pi r^2 Q). \tag{4.9}$$

These two contributions differ in that the added momentum goes directly into the kinetic energy (4.5), whereas the added heat modifies the internal energy and entropy of the gas (4.6).

In comparing the different specific energies of the wind it is convenient to represent them as terms in v^2. Thus v_t, where

$$v_t^2 = (p/\rho), \tag{4.10}$$

is the velocity associated with the internal energy, v_g, where

$$v_g^2 = \frac{2GM_*}{r} \tag{4.11}$$

is the velocity associated with the potential energy, and v_q is the velocity associated with the thermal conductivity,

$$v_q^2 = 8\pi q r^2 / (-\dot{M}_*). \tag{4.12}$$

We then have

$$E = \frac{1}{2}V^2 + \frac{\gamma}{\gamma - 1}v_t^2 - \frac{1}{2}v_g^2 + \frac{1}{2}v_q^2. \tag{4.13}$$

The minus sign in front of v_g^2 arises because we have to do work to lift the wind material in the star's gravitational potential.

For winds which are strongly gravitationally bound near the stellar surface and whose energy flux is largely kinetic at infinity, we can simplify (4.13) by writing (with $V_0 \simeq v_t \simeq 0$)

$$E_0 = \tfrac{1}{2}v_{g0}^2 + \tfrac{1}{2}v_{q0}^2, \tag{4.14}$$

$$E_\infty = \tfrac{1}{2}V_\infty^2. \tag{4.15}$$

The difference $E_\infty - E_0$ is the energy per unit mass which has to be injected into the stellar atmosphere above its base to drive the wind. Using (4.9, 12, 14, 15) we write

$$\begin{aligned} F_A(r_\infty) &= \int_{r_0}^{r_\infty} dr' [(-\dot{M})D + 4\pi r^2 Q] \\ &\simeq \frac{1}{2}(-\dot{M})(V_\infty^2 + v_{g0}^2) - 4\pi q r_0^2. \end{aligned} \tag{4.16}$$

If we include the conductive flux of energy driving the wind we can say that the driving flux F_{d0} at the base of the atmosphere and dissipated within it is

$$F_{d0} \geq F_{A\infty} + 4\pi q_0 r_0^2 \simeq \tfrac{1}{2}(-\dot{M})(V_\infty^2 + v_{g0}^2). \tag{4.17}$$

The $\geq$ sign expresses the fact that we have to allow for radiation losses in the wind as well as the energy required to drive it. Using (4.17) and assuming $r_0 = R_*$ we get in solar units

$$\frac{F_{d0}}{4\pi r_0^2} \gtrsim 3.3 \times 10^3 \left(\frac{-\dot{M}}{10^{-7}M_\odot} \right) \left(\frac{M}{M_\odot} \right)$$

$$\times \left(\frac{400 R_\odot}{R_*} \right)^3 \left(1 + \frac{V_\infty^2}{v_{g0}^2} \right) \mathrm{erg\, cm^{-2}\, s^{-1}} . \qquad (4.18)$$

Using (4.17) and (4.18) we can classify the various types of winds, under the (valid) assumption of negligible radiation losses in the wind.

Early-type Stars. These winds are very fast, $V_{\mathrm{inf}}^2 \gg v_{g0}^2$, and most of the driving energy is used to accelerate them to their terminal velocities.

Solar-type Stars. In stars of solar type and analogous cases, V_∞^2 is of the order of v_{g0}^2; comparable amounts of energy are used to lift the plasma off the star and to accelerate it to its asymptotic velocity. With $-\dot{M}_\odot = 2 \times 10^{-14} M_\odot\, \mathrm{y}^{-1}$ we find the average energy flux

$$F_{d0}/4\pi r_0^2 \simeq 10^5\, \mathrm{erg\, cm^{-2}}\, s^{-1} .$$

Winds in Red Supergiants. In general we have $V_\infty^2 \ll v_{g0}^2$, and almost all the driving energy is used to lift the material off the star. For a typical K5 supergiant ($M_* \simeq 16\, M_\odot, R_* \simeq 400 R_\odot$) and $-\dot{M}_* \simeq 10^{-7}\, M_\odot\, \mathrm{y}^{-1}$, we find $F_{d0} \simeq 6 \times 10^4\, \mathrm{erg\, cm^{-2}\, s^{-1}}$. For a Mira-type star ($M_* = M_\odot, R_* \simeq 400 R_\odot$) and a high mass loss rate ($-\dot{M}_* \simeq 2 \times 10^{-6}\, M_\odot\, \mathrm{y}^{-1}$) we find $F_{d0} \gtrsim 5 \times 10^4$ erg cm^{-2} s^{-1}.

The comparable values at the base of the wind are due to the very low gravitational potential compared with the Sun, and also the very large area ($\sim R_*^2$) over which the energy is transported.

4.5.3 Constraints on the Momentum

This is particularly significant for hot stars, where the wind is driven by radiation pressure. We can write the momentum equation above a reference level r_0:

$$V\frac{\mathrm{d}V}{\mathrm{d}r} = -\frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}r}(P + P_R) - \frac{GM}{r^2} , \qquad (4.19)$$

where P_R is the radiation pressure. We may rewrite this as

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(\rho r^2 V^2 + (1 - \eta)\frac{L}{4\pi c} \right) = -\frac{\rho v_{g0}^2 r_0}{2r^2} , \qquad (4.19')$$

where $\eta(r)$ is the fraction of the momentum flux L/c received by the wind during its motion from the reference level r_0 to r. Integrating (4.19) from r_0 to r_∞ (the asymptotic regime) we get

$$-\dot{M}V_\infty = \frac{\eta_\infty L}{c}\left\{1 + \frac{v_{g0}^2}{u_\infty^2}\frac{V_\infty}{2\langle V\rangle}\right\}^{-1}, \tag{4.20}$$

where $\langle V\rangle^{-1} = \int_{r_0}^{r_\infty} r_0 dr/r^2 V$. When the terminal velocity V_∞ is much larger than the escape velocity at r_0 (more precisely, when $v_{g0}^2 \ll 2V_\infty\langle V\rangle$), (4.20) reduces to the usual form

$$-\dot{M}V_\infty \cong \eta_\infty \frac{L}{c}. \tag{4.21}$$

η_∞ can be larger than unity if multiple scattering of photons is significant: it is less than one if only a fraction of the star's radiation couples to the wind. We often assume $\eta_\infty \simeq 1$. This is valid for winds from many hot stars.

Returning to (4.17), the terminal velocity is related to the energy added to the wind and to the mass loss

$$V_\infty^2 \simeq 2(F_{A_\infty} + 4\pi q r_0^2)\frac{1}{(-\dot{M})} - v_{g0}^2. \tag{4.22}$$

Integrating the momentum equation (4.5) to the sonic point r_s gives the mass loss rate, which is proportional to V_0 at the reference level.

$$-\dot{M}_* \sim v_t \left(\frac{r_s}{r_0}\right)^2$$

$$\times \exp\left[-\frac{v_{g0}^2}{2v_t^2}\left(1 - \frac{2}{v_{g0}^2}\int_{r_0}^{r_s} D dr - \frac{r_0}{r_s}\right) - \frac{1}{2} + \frac{V_0^2}{2v_t^2}\right]. \tag{4.23}$$

This shows that the mass loss rate $-\dot{M}_*$ increases when heat is injected (raising v_t) or when momentum is added (raising D) *in the subsonic region.*

We note the different behaviours of mass loss rate and terminal velocity. Injecting energy in the subsonic domain raises F_A and $-\dot{M}_*$, while the terminal velocity varies little or even decreases. In contrast, adding energy in the supersonic regime has little effect on $-\dot{M}_*$, but increases the terminal velocity V_∞. Thus for winds of low terminal velocity, $V_\infty^2 \ll v_{g0}^2$, most of the energy must be injected in the subsonic region. This is the important region for red supergiants as well as for O and B stars.

4.5.4 Driving Mechanisms

The considerations above give an idea of the energy and momentum contributions in various stellar winds.

Cool Stars. It is easy to see that a thermal wind would require large injections of heat and thus significant radiative emission, which is not observed. We therefore look to momentum injection. It is most likely that the wind is driven by Alfvén waves, which, being very little damped, can propagate far from the surface of the star. In this case Holtzer et al. (1983) give the mass loss rate as

$$-\dot{M}_* = 1.8 \times 10^{-13} \left(\frac{f_0}{10^6 B_0} \right) \left(\frac{R_*}{R_\odot} \right)^{7/2} \left(\frac{M_\odot}{M_*} \right)^{3/2} M_\odot \ \text{year}^{-1},$$

where f_0 is in erg cm^{-2}s^{-1}, B_0 in gauss, and these values are at the reference level $r_0 \simeq R_*$. If the Alfvén waves are undamped,

$$V_\infty \simeq v_{g0} \left[\frac{8}{7} \left(\frac{v_a}{V} \right)_{\text{sonic}} + \frac{5}{7} \right]^{1/2},$$

where $v_a^2 = B_0^2/8\pi\rho$ gives the Alfvén speed.

The mass loss rate is compatible with observation, but the predicted terminal velocity is very large and we have to assume that the Alfvén waves are damped. This is perfectly reasonable (friction of ions on neutrals), but the damping rate has to be very carefully adjusted to give the observed velocities. There is probably some kind of back reaction which arranges a similar terminal velocity for all stars, but this remains to be discovered.

Wind driving by radiation pressure on dust grains could be important for cool stars where condensation of refractory material can occur in the outer layers. Energy and momentum arguments show that this is a possible process but needs some kind of self-regulating mechanism to give agreement with the observed mass loss rates and terminal velocities.

Hot Stars. The idea of a radiation-driven wind is self-consistent in order of magnitude. In the Sobolev approximation the absorption in a line of frequency ν_i occurs over a band $\Delta\nu_i = \nu_i(1/c)(dV/dr)\Delta r$, in a layer of mass $4\pi r^2 \rho \Delta r$. The quotient $\sum_i (1/L)L(\nu_i)\Delta\nu_i$ is then the absorbed fraction of the radiation output L. The radiative acceleration is then the absorbed momentum divided by the mass of the layer of thickness Δr. Thus we have

$$g_{\text{R}} = \frac{L}{c} \sum_i \frac{L(\nu_i)}{L} \nu_i \frac{dV}{c} \frac{1}{4\pi r^2 \rho \, dr}.$$

In the region where gas pressure is negligible, the steady-state equation of motion is

$$V \frac{dV}{dr} = g_{\text{R}}.$$

Table 4.4. Mass loss from hot stars

Star	Spectral type	T_{eff} $(10^3$ K)	log g cm s^{-2}	log $(L/L_\odot)$	$\dot{M}_{obs}$ 10^{-6} $M_\odot$/y	$\dot{M}_{cal}$ 10^{-6} $M_\odot$/y	V_∞^{obs} km s^{-1}	V_∞^{calc} km s^{-1}
P Cyg	B ll a	18.0	2.0	5.64	20-30	29	400	395
ε Ori	B0 l a	28.5	3.25	5.91	3.1	3.3	2 010	1 950
ζ Ori	O9.5 l	30.0	3.45	5.79	2.3	1.9	2 290	2 274
9 Sgr	O4 (f) V	50.0	4.10	5.95	4.0	4.0	3 440	3 480
HD 48099	O6 e	39.0	4.00	5.40	0.63	0.64	3 500	3 540
HD 42088	O6.5 V	40.0	4.05	4.89	0.13	0.20	2 600	2 600
λ Cep	O6 ef	42.0	3.7	5.90	4.0	5.1	2 500	2 500

The dimensionless quantity $\sum_i (1/L)L(\nu_i)\nu_i = N_{eff}$ is interpreted as the effective number of strong lines. Then we have

$$\dot{M} = \frac{L}{c^2} N_{eff} \, .$$

Then from (4.21) and $q = 1$,

$$V_\infty \simeq \frac{c}{N_{eff}} \, ,$$

and with $N_{eff} = 100$ we have $V_\infty \simeq 3000$ km s^{-1}, which is indeed of the observed order. However, a detailed treatment shows that the transfer equation must be solved very carefully to take account of the velocity field, geometrical effects, and the deviation of the level populations from LTE. This has been done for a very large number of lines (250 000) (Kudritzki et al. 1986). The results are quite convincing (Table 4.4).

4.6 Heating Mechanisms

In some dwarfs, giants, and supergiants of solar-type strongly ionised species are seen (C IV λ 1550 Å; N V λ 1240 Å), indicating the presence of hot plasma (150 000 K). They also show X-ray emission. It thus appears that some late-type stars have chromospheres and coronae (see Chap. 7).

For blue stars with strong winds, only detailed studies (Kudritzki et al. 1986) show the non-LTE nature of lines from highly ionised species such as O VI. It thus appear possible to explain simultaneously the spectrum and wind of blue supergiants without requiring solar-type heating (see Sect. 4.6.1 below). The lack of a mechanical energy source for chromospheric heating would raise a real diffculty, which seems now to have been avoided.

The role of mechanical energy injection in driving cool winds (momentum injection from Alfvén waves) suggests that we look closely at acoustic waves in heating the outer layers and producing a chromosphere. It is tempt-

ing to extend to red giants and supergiants the mechanisms invoked to drive the solar wind, producing a chromosphere, a transition zone, and a large expansion zone which cools rapidly. The energy available in the convection zone is important here. But the main question is to know how the mechanical energy reaches the outer layers and is dissipated so as to drive the wind.

In the hydrogen convection zone near the surface, mechanical energy produced by the Rayleigh–Taylor instability propagates into the exterior as acoustic waves, which dissipate as heat and constitute an energy source in the immediate vicinity of the star; it also creates a complex magnetic field through a dynamo effect. MHD waves propagate in this field, and plasma instabilities convert magnetic energy into heat.

The complexity of the heating mechanism is clearly visible at the Sun's surface and demonstrates some of the dissipation mechanisms discussed here. This is also seen on an HR diagram where the various stellar-wind regimes are marked (Fig. 4.2). Below the line XX' (the Linsky–Haisch line, see Chap. 7) we detect X-ray emission and a transition zone (similar to the solar transition zone between the chromosphere and the corona). There are weak winds below and strong winds above. In the upper part of the diagram there are stars with permanent circumstellar lines. The main sequence is shown for reference. The division into stars with transition zones and stars with cool strong winds is caused by a thermal instability. Stars with strong winds do not appear to have hot coronae. The difference between stars with strong and weak winds may have a dynamical origin (the speed of mass loss in stars with strong winds prevents a corona forming because the cooling time is short; there is a region of thermal instability between $15\,000$ K and 5×10^5 K, so depending on the balance of heating and cooling the plasma settles at $15\,000$ K or $500\,000$ K). Alternatively it may be the formation of dust grains which prevents the formation of a corona.

4.6.1 The Solar Wind

The existence of very different regions in the outer layers of the Sun, such as coronal holes, which appear to be the source of the solar wind, and magnetic loops in the corona, show the importance of magnetic fields and various heating mechanisms.

Acoustic waves are produced in the convection zone by compressional effects associated with turbulence. The mechanical energy flux in these waves is of the order of 10^7 to 10^8 erg cm^{-2}s^{-1}. The waves quickly become shocks as they move outwards through the atmosphere. The mechanical energy flux decreases rapidly with height because of dissipation and refraction of the waves. Acoustic waves can heat the chromosphere but cannot reach the corona.

MHD waves, particularly Alfvén waves, penetrate further out, following field lines. In *coronal loops*, where the temperature reaches 2.5×10^6 K,

Alfvén waves are trapped. If t_M is the period of an oscillation caused by turbulence, there is a corresponding wavelength $v_A t_M = \lambda_A$. There is a resonance if λ_A is a fraction $2l/n$ of the loop length l. A resonance means strong dissipation of the electromagnetic energy of the wave and the magnetic energy of the loop. This is known as the alternating-current heating mechanism because of the role of the frequency v_A/l. In *coronal holes*, where the field lines are open and the temperatures lower, the heating is no doubt due to Alfvén waves propagating along the field lines and dissipating under ion–electron friction.

The chromosphere–corona transition zone presents an unsolved problem. It is probably heated by thermal conduction from the corona, but of an unusual type, because the mean free path of the electrons (50 km) is comparable to the thickness of the transition zone. This implies non-Maxwellian distributions for the particles, and the conductivity must be computed taking account of these "abnormal" conditions.

It is difficult to achieve a quantitative model for chromospheric and coronal heating. It is also difficult to construct a model of the solar wind which simultaneously accounts for the thermal structure and the driving mechanism. It is clear that the theory of stellar winds is still in a primitive state and at best gives some idea of the physical processes involved. It is still not possible to include mass loss in stellar-evolution theory in a fully coherent way.

5. Hydrodynamics of the Stellar Interior: Convection and Rotation

5.1 Introduction

Complex chaotic motions of the stellar interior which are important to its structure and evolution result from *local* instabilities. We shall distinguish these from *global* instabilities such as those producing pulsations in variable stars.

We shall first consider non-rotating stars where radiative energy transport becomes locally unstable, leading to the formation of a zone of convection or semi-convection. We then discuss rotating stars. We explain the circulation flows briefly, and then the associated instabilities. In particular we shall show how the combination of rotation and turbulence can produce macroscopic magnetic fields in a *turbulent dynamo*. An important application of convective instability theory is to give (at least in principle) an explanation of stellar winds (see Chap. 4).

5.2 Convection

The condition for the stability of radiative equilibrium in a compressible medium is analogous to that for conductive equilibrium known as the *Rayleigh–Bénard* stability criterion. A fluid heated from below in a vessel with a horizontal base becomes unstable once convective heat transport becomes more efficient than conduction. One then sees a stable system of convective cells (Bénard cells) which rapidly become chaotic and then turbulent as the heat flux increases.

5.2.1 Physical Preliminaries

The treatment of stellar convection is based on the study of incompressible convection in the Boussinesq approximation; this allows a far-reaching, rigorous physical and mathematical analysis, introducing the basic ideas. Work in hydrodynamics attempts to make contact with the astrophysical situation, whose treatment is still essentially phenomenological at present.

We start with the hydrodynamical equations for a vertically stratified medium:

— The continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 . \tag{5.1}$$

— The momentum conservation:

$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = -g\rho \boldsymbol{u}_z - \nabla p + \frac{\partial \tau_{ik}}{\partial x_k} , \tag{5.2}$$

where $\boldsymbol{u}_z$ is the unit vector along the vertical z-axis and τ_{ik} the viscosity tensor.

— The thermal energy equation:

$$\rho C_p \frac{\partial T}{\partial t} - \delta \frac{\partial p}{\partial t} + m_k \left(\frac{\partial h}{\partial x_k} - \frac{1}{\rho} \frac{\partial p}{\partial x_k} \right) = Q - \nabla \cdot \boldsymbol{F}_R + \Phi , \tag{5.3}$$

where h is the specific enthalpy and Q is an internal heat source (e.g. thermonuclear).

We use the following notation:

$$\boldsymbol{m} = \rho \boldsymbol{v} , \tag{5.4}$$

$$\delta = - \left(\frac{\partial \log \rho}{\partial \log T} \right)_P , \tag{5.5}$$

$$\tau_{ik} = \mu \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} - \frac{2}{3} \delta_{ik} \frac{\partial u_l}{\partial x_l} \right) , \tag{5.6}$$

where μ is the coefficient of dynamical viscosity, $\mu = \rho \nu$, where ν is the coefficient of kinematic viscosity,

$$\Phi = \tau_{ik} \frac{\partial}{\partial x_k} \nu_i \tag{5.7}$$

(with summation over repeated indices i and k) is the energy produced per unit volume by viscous dissipation, and $\nabla \cdot \boldsymbol{F}_R$ is the divergence of the radiative flux:

$$\boldsymbol{F}_R = -\frac{4}{3(\kappa + \sigma)\rho} \nabla J , \tag{5.8}$$

where J is the solution of the transfer equation

$$\nabla \cdot \boldsymbol{F}_R = -(\kappa + \sigma)(J - B) , \tag{5.9}$$

with

$$J \simeq B = \frac{ac}{4\pi} T^4$$

in the interior.

The linear problem (perturbation of the equilibrium equations) has both acoustic and gravity modes. If there is instability, the gravity modes become the convection. For the present problem (very subsonic convective motions) we can ignore the acoustic modes. The simplest idea is thus to expand the equations in powers of the Mach number v_s/c in the gas. The resulting approximation is known as the *anelastic approximation*.

All quantities are taken to have the form

$$f = \overline{f} + f'(x, y, z, t), \tag{5.10}$$

where $\overline{f}$ is the stationary solution corresponding to the average situation. For a fluid layer of thickness d, we assume that the perturbations of pressure, density, and temperature are small. The pressure fluctuations are associated with the vertical kinetic energy, and the pressure $\overline{\rho} w^2$, where w is the vertical velocity component, is of the order of the pressure fluctuations

$$\frac{p'}{\overline{p}} \simeq \frac{\overline{\rho} w^2}{\overline{p}} - \tag{5.11}$$

giving the relations ($H_P = -(\mathrm{d}r/\mathrm{d} \log P)$ is the pressure scaleheight)

$$w^2 \simeq \frac{\overline{p}}{\overline{\rho}} \frac{T'}{T} = g H_P \frac{T'}{T} \simeq c_s^2 \frac{T'}{T}. \tag{5.12}$$

We finally get the equations

$$m_i = \overline{\rho} v_i, \tag{5.13}$$

$$\frac{\partial m_k}{\partial x_k} = 0, \tag{5.14}$$

$$\frac{\partial m_i}{\partial t} + \frac{\partial}{\partial x_k} \left(\frac{m_i m_k}{\overline{\rho}} \right) = -\frac{\partial}{\partial x_i}(\overline{p} + p') - g(\overline{\rho} + \rho')\boldsymbol{u}_z + \frac{\partial \tau_{ik}}{\partial x_k}, \tag{5.15}$$

$$\overline{\rho} \overline{C}_p \frac{\partial T'}{\partial t} - \overline{\delta} \frac{\partial p'}{\partial t} - \overline{C}_p \beta m_3 + m_k \left(\frac{\partial h'}{\partial x_k} - \frac{1}{\overline{\rho}} \frac{\partial p'}{\partial x_k} \right) - \frac{g}{\overline{\rho}} \rho' m_3$$

$$= \tau_{ik} \frac{\partial}{\partial x_k} \left(\frac{m_i}{\overline{\rho}} \right) + \overline{Q} + Q' - \frac{\partial}{\partial x_k}(\overline{F}_k + F'). \tag{5.16}$$

With

$$\alpha = \left(\frac{\mathrm{d} \log T}{\mathrm{d} \log P} \right)_\rho \tag{5.17}$$

we have the relations or definitions

$$\frac{\rho'}{\overline{\rho}} = \alpha(\overline{p}, \overline{T}) \frac{p'}{\overline{p}} - \delta(\overline{p}, \overline{T}) \frac{T'}{T}, \tag{5.18}$$

$$\beta = -\frac{1}{\overline{C}_p}\left(\frac{\mathrm{d}\overline{h}}{\mathrm{d}x_3} - \frac{1}{\overline{\rho}}\frac{\mathrm{d}\overline{p}}{\mathrm{d}x_3}\right) \simeq -\left(\frac{\mathrm{d}\overline{T}}{\mathrm{d}x_3} - \frac{\overline{\delta}}{\overline{\rho}\overline{C}_p}\frac{\mathrm{d}\overline{p}}{\mathrm{d}x_3}\right), \tag{5.19}$$

$$\tau_{ik} = \overline{\mu}\left[\frac{\partial}{\partial x_k}\left(\frac{m_i}{\overline{\rho}}\right) + \frac{\partial}{\partial x_i}\left(\frac{m_k}{\overline{\rho}}\right) - \frac{2}{3}\delta_{ik}\frac{\partial}{\partial x_l}\left(\frac{m_l}{\overline{\rho}}\right)\right]. \tag{5.20}$$

5.2.2 Modal Theory

The continuity equation is linear and thus separable, and we can write the velocity field as

$$m_1 = \frac{\partial f}{\partial x_1}\frac{\partial w}{\partial z}, \quad m_2 = \frac{\partial f}{\partial x_2}\frac{\partial w}{\partial z}, \quad m_z = f(x_1, x_2)w(z)k^2, \tag{5.21}$$

and we have

$$\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + k^2 f = 0, \tag{5.22}$$

which has different solutions for various convective structures:

— *in rolls:*

$$f = \cos kx_1 \tag{5.23}$$

— *rectangular:*

$$f = \cos lx \cos \sqrt{k^2 - l^2}\, y \tag{5.24}$$

— *hexagonal:*

$$f = 2\cos\frac{\sqrt{3}}{2}kx \cos\frac{\sqrt{3}}{2}ky + \cos ky \tag{5.25}$$

— *cylindrical:*

$$f = J_0(kr), \tag{5.26}$$

where J_0 is the first-order Bessel function of the first kind. The function $w(z)$ must be found from the other equations, which are non-linear. Using a truncated Fourier series for f one gets a set of linear equations which can be solved numerically.

The Boussinesq Approximation. If we assume that the thickness d of the convective zone is small compared with the pressure scaleheight we can neglect density variations and revert to the incompressible case.

If K is the thermal diffusivity we can write the system (5.13–16) in dimensionless form by choosing suitable length, time, and temperature scales. The lengthscale is clearly d; the timescale is defined from the thermal diffu-

sivity as d^2/K; the temperature scale is the difference ΔT between the base and the top of the convective zone, corrected for the adiabatic difference, i.e. $\Delta T - (gd/C_p)$. We express the vertical velocity w and the temperature and velocity fluctuations θ, v in these units, and write $\beta_A = (g/C_p)$ for the adiabatic gradient, of dimension 1/(length). We also introduce the following dimensionless numbers:

— *Rayleigh number:*

$$Ra = \frac{g\alpha d^3}{K\nu}\left(\Delta T - \frac{gd}{C_p}\right), \tag{5.27}$$

where α is the thermal expansion coefficient of the gas: $\alpha = 1/T$ for a perfect gas.

— *Prandtl number:*

$$\sigma_P = \frac{\nu}{K}. \tag{5.28}$$

If we take *horizontal means*, the equations become

$$\frac{\partial}{\partial z}\left(\bar{p} + \overline{\rho w^2}\right) = -g\rho, \tag{5.29}$$

$$\frac{\partial \overline{T}}{\partial t} + \frac{\partial}{\partial z}\overline{w\theta} = K\frac{\partial^2 \overline{T}}{\partial z^2}, \tag{5.30}$$

$$\frac{1}{\sigma_P}\left(\frac{\partial v}{\partial t} + v\cdot\nabla v\right) = -\nabla\varpi + Ra\theta u_z + \nabla^2 v, \tag{5.31}$$

$$\frac{\partial \theta}{\partial t} + w\left(\frac{\partial \overline{T}}{\partial z} - \beta_A\right) + v\cdot\nabla\theta - \overline{v\cdot\nabla\theta} = \nabla^2\theta, \tag{5.32}$$

with

$$\varpi = \left(\frac{p - \bar{p}}{\rho}\right) - \overline{w^2}. \tag{5.33}$$

These equations hold if $d \lesssim H_P$, but not if the convective zone extends several scaleheights. They can however justify phenomenological treatments.

In the Sun, taking d as the scaleheight H_P, we estimate the Rayleigh number

$$Ra_\odot \simeq 10^{12} \text{ to } 10^{20}.$$

Similarly the Prandtl number is

$$\sigma_{P_\odot} \simeq 10^{-6} \text{ to } 10^{-9},$$

meaning that viscosity effects are negligible compared with those of thermal diffusivity, and the convection is very turbulent.

Only Rayleigh and Prandtl numbers much closer to unity can be realised in terrestrial experiments. There is little justification for extrapolation to the astrophysical case.

We have to find the ratio of total and radiative fluxes. In steady state $\partial/\partial t = 0$, and (5.30) can be integrated as

$$\overline{w\theta} - \frac{\mathrm{d}\overline{T}}{\mathrm{d}z} = N \,, \tag{5.34}$$

where N, the Nusselt number, is the ratio of total to radiative flux:

$$N = \frac{F}{F_R} = \frac{F}{K\dfrac{\Delta T}{d}} \,. \tag{5.35}$$

We thus need the relation between N, Ra, and σ_P.

We now introduce a concept similar to mixing length, to which we shall return later. We assume that we can regard a convective zone as obeying the Boussinesq approximation *locally*. We thus introduce a lengthscale which plays the role of a *local* thickness of the convective region.

The equation of motion in the usual units is

$$\overline{\rho}\left(\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\nabla\boldsymbol{v} - \overline{\boldsymbol{v}\cdot\nabla\boldsymbol{v}}\right) = -\nabla p' - g\frac{\overline{\rho}}{T}\theta \,. \tag{5.36}$$

We eliminate $\nabla p'$ by performing the operation $\nabla \times \nabla \times$, giving

$$\frac{\partial}{\partial t}\nabla^2 w - \frac{g}{T}\nabla_H^2\theta = -\left[\nabla\times\nabla\times(\boldsymbol{v}\cdot\nabla\boldsymbol{v} - \overline{\boldsymbol{v}\cdot\nabla\boldsymbol{v}})\right]_z \,, \tag{5.37}$$

where ∇_H^2 is the horizontal Laplacian. Analysing in vertical and horizontal wavenumbers k_z and k_H and linearising, we get

$$\left(k_z^2 + k_H^2\right)\frac{\partial w}{\partial t} - g k_H^2\theta = 0 \,. \tag{5.38}$$

Assuming constant temperature gradient β, for a displacement ξ we have

$$\theta(\xi) = \beta\xi \,, \tag{5.39}$$

with

$$\beta = -\frac{\mathrm{d}\overline{T}}{\mathrm{d}z} + \frac{g}{C_p} \,. \tag{5.40}$$

We identify the time derivative as an advection term $\partial_t \sim w\partial/\partial\xi$, and get

$$w^2 \simeq \frac{g\theta\xi}{[1 + (k_z^2/k_H^2)]\overline{T}} \,. \tag{5.41}$$

We set $\Phi = 1 + (k_z^2/k_H^2)$ in this section.

The heat equation describes the heat exchanges during the motion. As before we replace ∂_t by an advection term to get

$$w\frac{\partial\theta}{\partial\xi} - \beta w = -Kk^2\theta \tag{5.42}$$

with

$$k^2 = k_H^2 + k_z^2. \tag{5.43}$$

Integrating,

$$\theta \simeq \left(\beta - \frac{Kk^2\theta}{w}\right)\xi, \tag{5.44}$$

and we relate k_z to the chosen scale l by

$$l = \frac{\pi}{k_z}. \tag{5.45}$$

We assume now that the displacement ξ is $1/2$ (equivalent to being in the middle of a layer of thickness l) and find the convective flux

$$F_c = \overline{\rho C_p w\theta} = \tfrac{1}{4}\Phi^{-1/2}\eta^{-3}s^{-1}\left[(1+\eta^2 s)^{1/2} - 1\right]^3 K\rho C_p\beta, \tag{5.46}$$

where

$$s = \frac{gpl^4}{TK^2} = Ra\sigma_P, \tag{5.47}$$

$$\eta = \frac{2}{\pi^2}\Phi^{-3/2}(\Phi - 1). \tag{5.48}$$

We deduce the convective flux in two extreme cases:

— *efficient convection* ($\eta^2 s \gg 1$):

$$F_c = \tfrac{1}{4}\Phi^{-1/2}s^{1/2}K\rho C_p\beta; \tag{5.49}$$

— *inefficient convection* ($\eta^2 s \ll 1$):

$$F_c = \tfrac{1}{32}\Phi^{-1/2}\eta^3 s^2 K\rho C_p\beta. \tag{5.50}$$

Efficient convection gives a flux

$$F_c = \frac{1}{4}\rho C_p T\Phi^{-1/2}(\Delta\nabla)^{1/2}\left(\frac{\Re T}{\rho}\right)^{1/2}\left(\frac{l}{H_P}\right)^2, \tag{5.51}$$

where

$$\Delta\nabla = \left(\frac{d\log T}{d\log p}\right) - \left(\frac{d\log T}{d\log p}\right)_{ad}. \tag{5.52}$$

The corresponding velocity is

$$w = \left(\frac{1}{4} \frac{gH_P}{\Phi} l^2 \frac{\Delta\nabla}{H_P^2} \right)^{1/2} , \qquad (5.53)$$

and we can re-express the flux as

$$F_c = 2\rho\Phi \left(\frac{\mu C_p}{\Re} \right) w^3 , \qquad (5.54)$$

giving the order of magnitude of the turbulent vertical velocity (with $\Phi = 2$, $(\mu C_p/R) = 5/2$)

$$w \simeq \left(\frac{F}{10\rho} \right)^{1/3} . \qquad (5.55)$$

Dimensional analysis of (5.27), (5.28), and (5.35) gives the following result: if the heat flux does not depend on d or the viscosity, then

$$N = \mathrm{const}(Ra\sigma_P)^{1/3}$$

and the 1/3 power law appears to be well verified in some experiments.

For hard turbulence (following Liebschaber's terminology) with a Rayleigh number $Ra > 10^8$, boundary layers which develop at rigid boundaries in laboratory experiments are extremely important, causing a transition to a relation $N \propto (Ra\sigma_P)^{2/7}$. Despite the very large Rayleigh numbers ($Ra \simeq 10^{25}$) in stellar convective layers, it is not clear that the results of laboratory experiments can be extended to stars.

5.3 The Theory of Convection Zones

The stellar case is made difficult by the fact that the instability can extend over several scaleheights, so that there is a large density variation across the unstable zone. We cannot use the usual linearisation methods to look for the onset of instability. There have been considerable efforts to treat convection in stars, and there has been significant progress. However, the *phenomenological* description we shall give below remains indispensible for treating convection.

In a star the condition for dynamical instability is usually introduced as follows. If a fluid element, here called a blob, is displaced vertically without heat exchange with its surroundings, it undergoes an adiabatic transformation and its density varies as

$$\delta\rho_{\mathrm{ad}} = \left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P} \right)_{\mathrm{ad}} \frac{\mathrm{d}\log P}{\mathrm{d}r}\delta r = -\left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P} \right)_{\mathrm{ad}} \frac{\delta r}{H_P} , \qquad (5.56)$$

where we have defined the local scaleheight through

$$-\frac{1}{H_P} = \frac{\mathrm{d}\log P}{\mathrm{d}r} . \tag{5.57}$$

For a region of the star in radiative equilibrium the density variation is defined by

$$\delta\rho_{\mathrm{rad}} = -\left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{rad}} \frac{\delta r}{H_P} . \tag{5.58}$$

If after its vertical displacement the blob is less dense than its surroundings, buoyancy forces will cause it to rise further; thus the medium is unstable if

$$\left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{ad}} > \left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{rad}} . \tag{5.59}$$

Conversely the stability condition (K. Schwarzschild 1906) is

$$\left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{ad}} \leq \left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{rad}} \quad (\text{stability}) . \tag{5.60}$$

For a uniform chemical composition, with pressure only a function of temperature and density, the pressure difference between the blob and its surroundings vanishes if $v/c_s \ll 1$, and we have

$$\begin{aligned}
\delta\log P &= \frac{\partial\log P}{\partial\log\rho}(\delta\rho)_{\mathrm{ad}} + \frac{\partial\log P}{\partial\log T}(\delta T)_{\mathrm{ad}} \\
&= \frac{\partial\log P}{\partial\log\rho}(\delta\rho)_{\mathrm{rad}} + \frac{\partial\log P}{\partial\log T}(\delta T)_{\mathrm{rad}} .
\end{aligned} \tag{5.61}$$

Then

$$(\delta T)_{\mathrm{ad}} - (\delta T)_{\mathrm{rad}} = -\left((\delta\rho)_{\mathrm{ad}} - (\delta\rho)_{\mathrm{rad}}\right)\left(\frac{\partial\log P}{\partial\log\rho} \bigg/ \frac{\partial\log P}{\partial\log T}\right) . \tag{5.62}$$

We usually write the stability condition for the temperature, since $(\delta T)_{\mathrm{rad}}$ is given by the transfer equation. It has the opposite sign from the density condition: a rising blob which cools more rapidly than its surroundings becomes denser, and buoyancy is now a restoring force:

$$\left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right)_{\mathrm{ad}} \geq \left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right)_{\mathrm{rad}} \quad (\text{stability}) . \tag{5.63}$$

We usually write

$$\left(\frac{\mathrm{d}\log P}{\mathrm{d}\log\rho}\right)_{\mathrm{ad}} = \gamma_1 , \qquad \left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right)_{\mathrm{ad}} = \frac{\gamma_2 - 1}{\gamma_2} , \qquad \frac{\mathrm{d}\log\rho}{\mathrm{d}\log T} = \frac{1}{\gamma_3 - 1} ,$$

where local ionisation implies that $\gamma_1 \neq \gamma_2 \neq \gamma_3$. In a fully ionised medium where we can neglect the second virial coefficient and electron degeneracy,

the perfect-gas approximation holds, and $\gamma_1 = \gamma_2 = \gamma_3 = 5/3$ for a non-relativistic gas.

Combining the transfer equation with that of hydrostatic equilibrium,

$$\left(\frac{d \log T}{d \log P}\right)_{\text{rad}} = \frac{1}{16\pi ac} \frac{P}{P_R} \frac{\kappa L(r)}{GM(r)} , \tag{5.64}$$

where κ is the opacity (including effects of heat transport), P_R the radiation pressure, $L(r)$ the luminosity at r, and $M(r)$ the mass inside r. a is the radiation constant, related to Stefan's constant by $ac = 4\sigma$.

Stability holds if the temperature gradient is *subadiabatic* and instability occurs if the temperature gradient is *superadiabatic*. Two different effects tend to produce instability: the decrease in compressibility caused by ionisation, and increases in opacity when e.g. more atomic levels are available for absorption.

5.3.1 Elementary Treatment of Convection

The standard method of dealing with convection zones is a phenomenological treatment modelled on Prandtl's theory of turbulence, which brings in the idea of a *mixing length*. Despite its shortcomings and even some internal inconsistency, this mixing-length theory provides a satisfactory approximate description of wide convection zones, because these are close to adiabatic over most of their extent. On the other hand, convection zones are narrow and close to radiative equilibrium in stars of spectral types hotter than F2, so that mixing-length theory does not apply very well.

We assume a vertical velocity v and write the convective flux as the product of v and a certain quantity $\rho C_p \delta T$ of energy being transported by the fluid

$$F_c = \rho C_p \delta T v .$$

The temperature difference δT is expressed as the deviation of the fluid element from the local average T_*:

$$\delta T = T_{\text{Blob}} - T_* . \tag{5.65}$$

We assume that this temperature difference arises because the fluid element retains a separate temperature from that of its surroundings until the instant of mixing:

$$\delta T = \left|\left(\frac{\partial T}{\partial r}\right)_{\text{Blob}} - \left(\frac{\partial T}{\partial r}\right)_*\right| l , \tag{5.66}$$

where l is the mixing length. We will see that δT is given by a first-order expansion, assumed valid whatever l is. Perhaps surprisingly, the final model of the convective zone thus obtained is indistinguishable from models using

more realistic phenomenological approaches, lending it a certain a posteriori justification.

The blob begins to move with an infinitesimal temperature difference from its surroundings and mixes completely at the end of a pathlength l. There is no theory of the mixing length itself, and one simply assumes some relation between the mixing length and the characteristic length $H_P = -\mathrm{d}r/\mathrm{d}\log P$, writing

$$l = \alpha H_P , \tag{5.67}$$

where α is a parameter of order unity. Since there is no theory of l, one uses the freedom in α to satisfy global constraints on the internal structure. Although the choice of this parameter has rather little effect on global quantities such as luminosity and radius, it is important for the details of the internal structure, particularly if we wish to achieve a precise match of the radius to observations, e.g. for the Sun.

The temperature variation in the blob is caused partly by expansion and partly by radiative exchanges with the surroundings. The latter are usually estimated by calculating the rate of heat exchange of a spherical blob of radius a whose central temperature is δT higher than the surroundings. If the timescale a^2/K, with K the thermal diffusivity,

$$K = \frac{16}{3}\sigma T^3 \frac{\mu}{\kappa \rho^2 C_p} , \tag{5.68}$$

is smaller than the timescale l/v for the blob motion, we can take the temperature distribution inside the blob to be the same as in the stationary case:

$$\delta T = (\delta T)_{\text{centre}} \frac{a}{\pi r} \sin \frac{\pi r}{a} , \tag{5.69}$$

and the rate of heat loss through radiative transfer is

$$\frac{\mathrm{d}w}{\mathrm{d}t} = -4\pi a K_R \delta T , \tag{5.70}$$

where $K_R = 16\sigma T^3/3\kappa\rho$ is the radiative conductivity. Averaging over the sphere, the heat loss per unit volume is

$$\frac{\mathrm{d}w}{\mathrm{d}t} = -\frac{3}{a^2} K_R \rho T . \tag{5.71}$$

The calculation then proceeds in stages:

(a) We define a radiative gradient $(\mathrm{d}\log T/\mathrm{d}\log P)_{\text{rad}}$, written as ∇_R, through the relation

$$F = \frac{K_R T}{H_P} \nabla_R . \tag{5.72}$$

(b) We find a velocity by integrating the linearised equation of motion

$$\rho \ddot{z} = g\rho \left(\delta \frac{d \log \rho}{d \log P} \right) \frac{1}{H_P} z \, , \tag{5.73}$$

giving

$$v = \left(g \left[\delta \frac{d \log \rho}{d \log P} \right] \frac{1}{H_P} \right)^{1/2} z \, , \tag{5.74}$$

and estimate its mean value $\langle v \rangle = (1/l) \int v \, dz$,

$$\langle v \rangle = \frac{1}{2} \left[g \left(\delta \frac{d \log \rho}{d \log P} \right) \frac{1}{H_P} \right]^{1/2} l \, , \tag{5.75}$$

where the difference in density gradients, assuming a perfect gas, is given by

$$\delta \frac{d \log \rho}{d \log P} = \left(1 - \frac{d \log \mu}{d \log P} \right) (\nabla_{\text{Blob}} - \nabla_*) \tag{5.76}$$

with

$$\nabla_* = \left(\frac{d \log T}{d \log P} \right)_* , \quad \nabla_{\text{Blob}} = \left(\frac{d \log T}{d \log P} \right)_{\text{Blob}} , \tag{5.77}$$

and we set

$$\delta = \left(1 - \frac{d \log \mu}{d \log P} \right) . \tag{5.77'}$$

(c) The convective flux is defined by

$$F_c = \rho C_p \frac{T}{H_P} (\nabla_{\text{Blob}} - \nabla_*) l v \, . \tag{5.78}$$

(d) The radiative flux is defined from the average local gradient in the star ∇_*, and we write

$$F_R = \frac{K_R T}{H_P} \nabla_* \, . \tag{5.79}$$

(e) Flux conservation requires

$$\frac{K_R T}{H_P} \nabla_R = \frac{K_R T}{H_P} \nabla_* + \rho C_p \frac{T}{H_P} (\nabla_{\text{Blob}} - \nabla_*) l \langle v \rangle \tag{5.80}$$

or

$$\nabla_* - \nabla_R + \frac{\rho C_p}{K_R} (\nabla_{\text{Blob}} - \nabla_*) l \langle v \rangle = 0 \, . \tag{5.81}$$

(f) The thermodynamics of the blob provide a second relation between ∇_* and ∇_{blob}. If ε is the energy loss rate per unit mass,

$$T\frac{dS}{dt} = -\varepsilon, \tag{5.82}$$

where S is the specific entropy of the blob. Taking $l/\langle v \rangle$ as timescale, the change of heat content is

$$\frac{dQ}{dt} = \frac{\langle v \rangle}{l}\delta Q$$
$$= \frac{\langle v \rangle}{l}\frac{l}{H_P}C_p\rho T(\nabla_{\text{Blob}} - \nabla_{\text{ad}}). \tag{5.83}$$

Using

$$\delta T = \frac{l}{H_P}T(\nabla_* - \nabla_{\text{Blob}}) \tag{5.84}$$

we get

$$\frac{\langle v \rangle}{l}C_p\rho(\nabla_{\text{Blob}} - \nabla_{\text{ad}}) = \frac{3}{a^2}K_{\text{R}}(\nabla_* - \nabla_{\text{Blob}}). \tag{5.85}$$

Eliminating $\langle v \rangle / l$ from (5.75),

$$\frac{1}{2}\left[\frac{g\delta}{H_P}(\nabla_* - \nabla_{\text{Blob}})\right]^{1/2}C_p\rho(\nabla_{\text{Blob}} - \nabla_{\text{ad}}) = \frac{3}{a^2}K_{\text{R}}(\nabla_* - \nabla_{\text{Blob}}), \tag{5.86}$$

where

$$\frac{a^2}{6K_{\text{R}}}\left(\frac{g\delta}{H_P}\right)^{1/2}C_p\rho(\nabla_{\text{Blob}} - \nabla_{\text{ad}}) = (\nabla_* - \nabla_{\text{Blob}})^{1/2}. \tag{5.87}$$

Taking $(\nabla_* - \nabla_{\text{ad}})$ as unknown and setting

$$U = \frac{3K_{\text{R}}}{a^2}\left(\frac{H_P}{g\delta}\right)^{1/2}\frac{1}{C_p\rho}, \tag{5.88}$$

we write

$$\frac{1}{2U}(\nabla_{\text{Blob}} - \nabla_{\text{ad}}) = (\nabla_* - \nabla_{\text{ad}} + \nabla_{\text{ad}} - \nabla_{\text{Blob}})^{1/2}, \tag{5.89}$$

where

$$(\nabla_{\text{Blob}} - \nabla_{\text{ad}})^2 + 4U^2(\nabla_{\text{Blob}} - \nabla_{\text{ad}}) - 4U^2(\nabla_* - \nabla_{\text{ad}}) = 0, \tag{5.90}$$

giving

$$(\nabla_{\text{Blob}} - \nabla_{\text{ad}}) = 2U\left[-U + \sqrt{U^2 + (\nabla_* - \nabla_{\text{ad}})}\right]. \tag{5.91}$$

(g) We use (5.75) and (5.76) giving $\langle v \rangle$, (5.87), and (5.88) in the expression

$$F = F_R + F_c \tag{5.92}$$

for the total flux, and obtain

$$(\nabla_* - \nabla_{\rm ad}) - (\nabla_{\rm rad} - \nabla_{\rm ad}) + \frac{3}{2}\frac{l^2}{a^2}\frac{1}{U}$$
$$\times \left(-U + \sqrt{U^2 + \nabla_* - \nabla_{\rm ad}}\right)^3 = 0. \tag{5.93}$$

The parameter U is explicitly written

$$U = \frac{16\sigma T^3 H_P}{C_p \rho^2 \kappa (a^2/l^2) l^2}\left(\frac{\mu}{\Re T \delta}\right)^{1/2}. \tag{5.94}$$

Equations (5.93) and (5.94) were found without using special assumptions about $\langle v \rangle$ or F_c; all the uncertainty is in the choice of $\langle l/a \rangle$, which is to some extent free. The choice of $l/H_P = \alpha$ fixes the properties of the convective zone.

Equation (5.93) gives the difference $(\nabla_* - \nabla_{\rm ad})$. At low densities near the stellar surface, U is large and $(\nabla_* - \nabla_{\rm ad}) \simeq (\nabla_R - \nabla_{\rm ad})$, i.e. the local average gradient is close to the radiative one. Expressing U differently, we can use the approximation $\kappa P/g \simeq \tau$,

$$U \simeq \frac{6F}{\alpha^2 (a^2/l^2)\frac{1}{2}\rho v_s^3}\frac{1 + \frac{3}{2}\tau}{\tau}\left(\frac{\gamma^3}{\delta}\right)^{1/2}, \tag{5.95}$$

where F is the total flux, τ the optical depth, v_s the sound speed, and γ the adiabatic compressibility. $\rho v_s^3/2$ has the dimensions of a flux, and for $\tau \gg 1$ we see that U is the ratio of the total flux to that given by transporting the internal energy of the gas at its sound speed. We call this flux F_s. As one moves into the convection zone F/F_s rapidly becomes small, U is small, and the local gradient is close to adiabatic:

$$(\nabla_* - \nabla_{\rm ad}) = \left(\frac{2}{3}\frac{a^2}{l^2}U\right)^{2/3}(\nabla_{\rm rad} - \nabla_{\rm ad})^{2/3}.$$

5.3.2 The Convective Regime

Use of a mixing-length theory throws up some difficulties which force corrections on the phenomenological model. The most important is for red supergiants, which have very extended low-density convective zones. These can imply supersonic velocities; but supersonic turbulence must dissipate rapidly in shock waves. There are two phenomenological assumptions we can make:

— We can limit the velocity to that of sound, and require the radiative flux to make up the total flux.

— We can add a dissipation term representing the shock waves; the (constant) total flux then contains an extra term.

5.3.3 Convective Overshooting

The Schwarzschild criterion $\nabla_{\text{ad}} = \nabla_* = \nabla_{\text{rad}}$ does not define an impermeable barrier to the convective motions. Falling blobs may cross the lower boundary of a surface convection zone. In the subadiabatic region the buoyancy force decelerates the blob. In the phenomenological description the blob stops after one mixing length l.

For stellar structure and evolution there are two important boundaries: the outer boundary of the central convection zone and the lower boundary of the surface zone. In these regions convection is efficient; the blobs are optically thick and exchange heat only at the ends of their pathlengths, and we can regard the motion as adiabatic. If we call $\delta(r)$ the deviation from adiabaticity,

$$\delta(r) = \left[\left(\frac{\mathrm{d}T}{\mathrm{d}r} \right)_* \middle/ \left(\frac{\mathrm{d}T}{\mathrm{d}r} \right)_{\text{ad}} \right] - 1 \,,$$

the temperature difference for a blob which has travelled from r_1 to r is

$$\delta T(r; r_1) = - \int_{r_1}^{r} \mathrm{d}r\, \delta(r) \left(\frac{\mathrm{d}T}{\mathrm{d}r} \right)_{\text{ad}} \,,$$

and the speed of the blob is given by

$$v^2(r; r_1) = 2 \int_{r_1}^{r} \mathrm{d}r' \frac{g(r')}{T(r')} \delta T(r'; r_1) \,,$$

where $g(r)$ is the gravity at distance r from the centre. We assume $r - r_1 \leq l$. The convective flux at level r is

$$F_{a,c} = f_a C_p(\rho, T) \rho(r) v(r; r - l) \delta T(r; r - l)$$

from rising motions occupying a fraction f_a of the surface. For the falling motions the speed and the temperature difference have opposite signs, but the flux $F_{d,c}$ of the descending motions has the same sign as $F_{a,c}$. During the penetration of the radiative zone the temperature difference inverts, and the convective flux changes sign before vanishing. The radiative flux reaches a maximum before taking the value F imposed by the total flux.

As an example we show in Fig. 5.1 results found by Shaviv and Salpeter for a perfect gas, with $l/H_P = 1/2$, and far from the convective boundary, $\delta = \delta_0 = 10^{-5}$. r_ε is the radius for which the radiative flux equals the total flux. The radiative flux reaches a maximum and returns to the total flux at the point where the velocity vanishes, defining the penetration depth (here
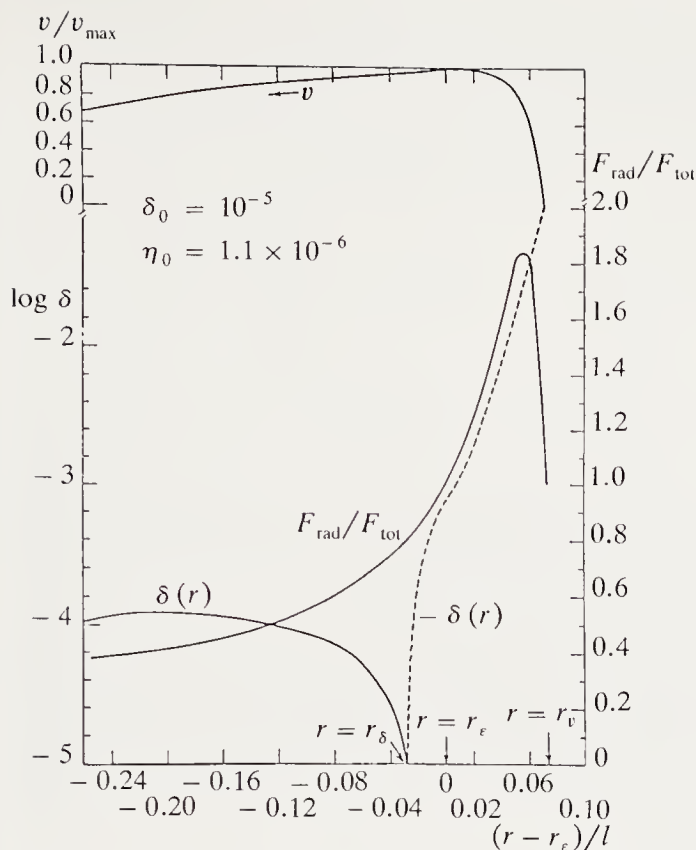
Fig. 5.1. Convective overshooting. The radiative flux as a function of depth (right-hand scale). The zone marked $r = r_\varepsilon$ on the abscissa is the point where the Schwarzschild criterion $\Delta\nabla = 0$ is satisfied. In the model describing the convection in local variables, the convective flux vanishes here. With overshooting, some heat is transported by convection to $r = r_T$, then there is a zone where the convective flux is negative and $F_{rad} > F$. The convective flux vanishes where the velocity vanishes, at $r = r_v$ (scale at *upper left*). The buoyancy factor $\delta(r) = (\nabla_*/\nabla_{ad}) - 1$, as a function of depth (*lower-left* scale) vanishes at r_δ below r_ε, so that the inertia of the blobs carries them above radius r_δ. (From G. Shaviv, E.E. Salpeter, *Astrophys. J.* **184**, 191 (1973). Reproduced by kind permission of *The Astrophysical Journal*; published by The University of Chicago Press; © 1973 The American Astronomical Society)

$(r - r_v)/l = 0.075)$. The point where $\nabla_* = \nabla_{ad}$ is distinct from the point where F_c vanishes before changing sign.

The penetration depth thus defined is a sort of average. It does not give the size of the region completely mixed by the overshooting. According to Maeder (see Sect. 3.5.1) the penetration depth derived from evolutionary models is small.

A more rigorous analysis in the plane-parallel case factorises the motion into a periodic horizontal function, with one or two modes, and an unknown vertical dependence. For example, temperature fluctuations are written as

$$T' = \sum_i f_i(n, y)\theta_i(z),$$

Table 5.1. Velocity field for the (D) model of Massaguer et al. (1984)

Depth	Temperature	Density	V (vertical)
0	0	0	–
0.2	1	0.1	–
1.20	11	1	max
2.08			0
			+
2.24			0
			–
2.5	17.5	4	

and for one mode, we can choose to represent the f_i as hexagons (5.25). We simplify by supposing there are three regions, stable, unstable, and stable, with arbitrary polytropic indices $1/(\gamma - 1)$ constrained to be respectively larger, smaller, and larger than the neutral-equilibrium value $3/2$. In the model of Massaguer et al., $n_1 = 3$, $n_2 = 1$, and $n_3 = 3$. This set of non-linear ordinary differential equations for the temperature and other fluctuations can be integrated numerically. We give here the results for the parameter choice (grid $a = 2.22$; critical Rayleigh number $Ra \simeq 10^3$) and Prandtl number $= 1$, for the two cases of ascending (A) and descending (D) motions along the axis of the hexagonal cell.

In case (D) the vertical velocity vanishes quite far from the base of the unstable zone, and direct and inverse cells follow each other. Table 5.1 gives (in reduced variables) the values of the temperature and density and the variations of the vertical velocity component as functions of depth.

For case D the penetration depth is 1.6 scaleheights.

For A cells the velocity structure requires a much smaller perturbation in the lower region at depth 1.32, corresponding to only 0.2 scaleheights, followed by very small amplitude direct and inverse cells (Fig. 5.2).

The basic result is that A cells penetrate little below the convection layer, but very little at the stable surface zone (0.2 scaleheights). D cells effectively do not penetrate the upper stable zone.

We should ask whether choice A or D is valid for the convection zone. Small-scale turbulence controls both the viscosity and thermal conductivity, giving $\sigma_P \simeq 1$; replacing the molecular viscosity by the turbulent decreases the Rayleigh number by 10 orders of magnitude and justifies calculations for $Ra = 10^3 R_{\text{critical}}$, where R_{critical} is the Rayleigh number for which the system becomes unstable.

The sign of motion along the cell axes depends on the dependence of viscosity on temperature ($\nu \sim T^{5/2}/\rho$) and agrees with observations of the solar granulation, which reveal ascending motions along the axes of the granules.

The image of the blob penetrating inside the subadiabatic region is too simple and does not correspond either to observations in the Earth's atmo-
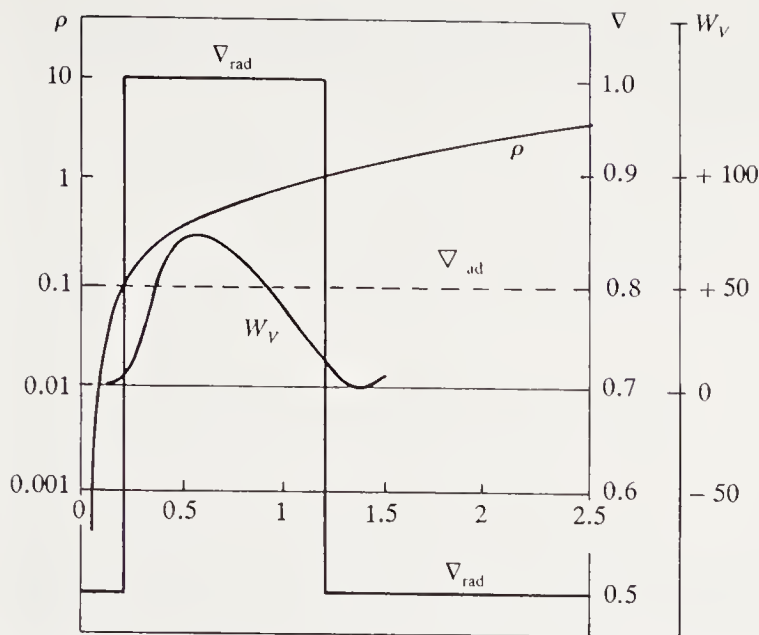
Fig. 5.2. Model of convective overshooting (from Massaguer et al.). (∇) The radiative and adiabatic gradients as functions of height. The unit of length is the thickness of the unstable zone. (ρ) Density is given in arbitrary units and varies from 0.1 to 1 in the unstable zone. W_V (in arbitrary units) is the vertical component of the velocity for a hexagonal cell with rising motion at the centre

sphere or to the results of numerical simulations of motions in a stratified medium. "Plumes" penetrate downwards to great depth (of the order of one scaleheight), opening out and sweeping up matter in their paths.

We present here a first-order analytic solution describing the penetrative motion, without giving a detailed description of the flow. We consider two cases: inward penetration from a surface convective zone, and outward penetration from a convective core.

Overshooting from a Surface Convection Zone. We aim to describe average properties at a level z. We assume that the temperature fluctuations T_1 and the vertical component w of the velocity are given by the same modal function $h(x, y)$:

$$T_1(x, y, z) = h(x, y)\delta T(z) \tag{5.96}$$

and

$$w(x, y, z) = h(x, y)W(z). \tag{5.97}$$

The function h is such that its mean value is zero but its mean square is unity over the fraction of the area where h is positive, and it is f when averaged over the whole area. The ratio of the mean cube to the mean square, $(\langle h^3 \rangle / \langle h^2 \rangle)$, measuring the asymmetry of the flow, is called c.

To a first approximation, the flow in the subadiabatic overshooting region will be regarded as adiabatic. Then neglecting the kinetic energy flux, the convective energy flux is

$$F_c = F - F_{\text{rad}} . \tag{5.98}$$

We have

$$F_{\text{rad}} = K_R \left(\frac{\mathrm{d}T}{\mathrm{d}z} \right)_{\text{ad}} , \tag{5.99}$$

where K_R is the radiative conductivity. The linear approximation gives

$$F_c = -F \frac{\mathrm{d} \ln K_R}{\mathrm{d} \ln P} \frac{z}{H_P} , \tag{5.100}$$

where z is the distance from the level of the Schwarzschild condition (z is measured positive downwards), and F is the total flux. On the other hand the convective flux can be expressed as advection of enthalpy:

$$F_c = -\rho C_P \langle w T_1 \rangle = -f C_P \rho W \delta T . \tag{5.101}$$

To obtain the vertical velocity as a function of z we write the linearised vertical deceleration:

$$\frac{1}{2} \frac{\mathrm{d} w^2}{\mathrm{d}z} = g \left(\frac{\partial \ln \rho}{\partial \ln T} \right)_P \frac{T_1}{T} . \tag{5.102}$$

Multiplying both sides by h and taking the average, we obtain

$$\frac{c}{2} \frac{\mathrm{d} W^2}{\mathrm{d}z} = -g Q \frac{\delta T}{T} , \tag{5.103}$$

where $Q = (\partial \ln \rho / \partial \ln T)_P$. Eliminating δT, integration of the last equation gives the relation between the penetration depth and the velocity W_0 at which the flow crosses the level $z = 0$:

$$\frac{L_P}{H_P} = W_0^{3/2} \left[\frac{3}{2} g Q K \chi_P \nabla_{\text{ad}} \right]^{-1/2} , \tag{5.104}$$

where χ_P is $(\partial \ln K_R / \partial \ln P)_{\text{ad}}$ and K the thermal diffusivity. The velocity W decreases from $z = 0$ and vanishes at $z = L_P$:

$$W^3 = W_0^3 [1 - (z/L_P)^2] . \tag{5.105}$$

The velocity scale $(\mathrm{d}z/\mathrm{d}\ln w)$ vanishes at the boundary of the convection zone. The boundary layer between the regions of overshooting and radiative equilibrium can be defined as the locus where the radiative energy transfer rate balances heat advection.

An expression for the maximum velocity can be found by assuming that it is the result of inertial motion from the level $z = -(H_P/\chi_P)$ to the level $z = 0$:

$$F = \frac{2cf\chi_P}{3} \frac{\rho W_0^3}{Q\nabla_{\mathrm{ad}}} .\tag{5.106}$$

The definition of the boundary layer gives a thickness L_t of the order of

$$L_t = \left(\frac{K^2 H_P}{g} \right)^{1/4} ,\tag{5.107}$$

which is of the order of a kilometer, with velocities of the order of one me-ter per second. Thus the boundary of the convection zone appears sharply defined, although it probably undulates somewhat under the velocity dis-persion of the impinging downdrafts, as illustrated in the boundary layer in our own atmosphere, where it is delineated by clouds.

Penetration from a Convective Core. In a quasi-adiabatic region, Rox-burgh (1978) has shown that the extension of the convection zone can be derived from the specific entropy equation. He obtains the condition

$$\int_0^r (L_{\mathrm{rad}} - L)\frac{1}{T^2}\frac{\mathrm{d}T}{\mathrm{d}r}\mathrm{d}r = 0 ,\tag{5.108}$$

where L_{rad} is the radiative luminosity (radiative transfer with the local value of the temperature gradient) and L the actual luminosity as a function of radius r. In this expression the fluctuations of the entropy flow, the kinetic energy flux, and the work done by the pressure fluctuations have all been neglected. These approximations are valid in the convective core, as shown by Zahn (1991), because of the very small departure from adiabaticity, even if the filling factor is very small. This defines a boundary which is different from the classical boundary defined by the Schwarzschild criterion.

5.3.4 Semi-convection

Another kind of instability occurs when there is a gradient of chemical composition and the radiative gradient increases outwards. This happens when the main opacity is scattering, with cross-section $\sigma = 0.19(1 + X)$. The important property is that σ increases with the hydrogen content. Near the convective core of an intermediate-mass or massive star, the defining condition $\nabla_{\mathrm{rad}} - \nabla_{\mathrm{ad}} = 0$ for the convective boundary at composition X_i fails for composition $X_e > X_i$ just outside the core. We thus write $\nabla_{\mathrm{rad}\,e} > \nabla_{\mathrm{rad}\,i}$, and the zone which was supposed to be stable under the condition $(\Delta\nabla)_i = 0$ is actually unstable.

We thus write the condition for marginal stability

$$\left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{rad}} = \left(\frac{\mathrm{d}\log\rho}{\mathrm{d}\log P}\right)_{\mathrm{ad}},$$

where we take account of $\nabla\mu$ on the left-hand side, but not on the right, because we assume that blobs retain the same composition during their motions. The marginal-equilibrium condition defines a zone of radially varying composition allowing a match to the outer region whose composition has not been changed by thermonuclear reactions.

This assumption allows one to calculate stellar models and follow their evolution. But we must still explain how the change of composition comes about in this zone. Very slow convection appears to have the right properties to produce the required mixing, hence the name *semi-convection zone*. In fact we can imagine that as the central convection zone evolves it creates a composition discontinuity at its boundary. Once this is sufficiently large, viscosity cannot prevent overturning motions and the composition change propagates as a wave of chemical discontinuity from the convective core to the region where the gas still has the original composition.

5.4 Circulation and Rotation

5.4.1 Von Zeipel's Theorem

Von Zeipel's theorem is fundamental to all discussions of the circulation induced in a star by its rotation. We consider a star entirely in radiative equilibrium. The flux and the energy production rate are thus related by

$$\nabla \cdot F = \rho\varepsilon_{\mathrm{Nucl}} \tag{5.109}$$

at each point, and the flux is given by

$$F = -\frac{4acT^3}{3\kappa\rho}\nabla T. \tag{5.110}$$

We also take P to be a function of density and temperature.

We now assume that the star rotates with angular velocity $\Omega = \Omega(\varpi)$, where ϖ is the distance from the rotation axis. Then we may write the hydrostatic-equilibrium equation as

$$\frac{1}{\rho}\nabla P = -\nabla\Phi, \tag{5.111}$$

where

$$\Phi(\varpi, z) = V(\varpi, z) - \int^{\varpi} \Omega^2(\varpi')\varpi'\,\mathrm{d}\varpi', \tag{5.112}$$

and the gravitational potential V is given by Poisson's equation

$$\nabla^2 V = 4\pi G\rho, \tag{5.113}$$

where G is the gravitational constant.

From (5.111) we have $P = $ constant on equipotentials, so that $P = P(\Phi)$ and (5.111) becomes

$$\frac{1}{\rho} = \frac{\mathrm{d}\Phi}{\mathrm{d}P}. \tag{5.114}$$

Thus the density is constant on equipotentials. Since the pressure is a function of temperature and density we can write $T = T(P, \rho)$, and the temperature is also constant on equipotentials. Thus the fluid satisfies $P = P(\rho)$, and is called *barotropic*.

Returning to (5.109) we easily derive

$$F = -\frac{4ac}{3}\frac{T^3}{\kappa\rho}\frac{\mathrm{d}T}{\mathrm{d}P}\nabla\Phi, \tag{5.115}$$

where

$$F = f(\Phi)\nabla\Phi. \tag{5.116}$$

We note that *the flux is proportional to the gravity on an equipotential surface in a barotropic fluid.* This is important for the study of the spectral properties of rotating stars.

Substituting for the flux divergence in (5.109) gives

$$\frac{\mathrm{d}f}{\mathrm{d}\Phi}\left(\frac{\mathrm{d}\Phi}{\mathrm{d}n}\right)^2 + f(\Phi)\nabla^2\Phi = \rho\varepsilon_{\mathrm{Nucl}}. \tag{5.117}$$

Now $(\mathrm{grad}\,\Phi)^2$ is equal to the square of the gravity, and we have also

$$\nabla^2\Phi = 4\pi G\rho - \frac{1}{\varpi}\frac{\mathrm{d}}{\mathrm{d}\varpi}(\Omega^2\varpi^2), \tag{5.118}$$

and so we can write (5.109) in the form

$$f'(\Phi)g^2 + f(\Phi)\left[4\pi G\rho - \frac{1}{\varpi}\frac{\mathrm{d}}{\mathrm{d}\varpi}(\Omega^2\varpi^2)\right] = \rho\varepsilon_{\mathrm{Nucl}}. \tag{5.119}$$

The case $\Omega = $ constant is particularly simple. We then have

$$f'(\Phi)g^2 + f(\Phi)(4\pi G\rho - 2\Omega^2) = \rho\varepsilon_{\mathrm{Nucl}}. \tag{5.120}$$

g is not constant on an equipotential. Equation (5.120) can only be satisfied if

$$f'(\Phi) = 0, \tag{5.121}$$

so that

$$f(\Phi) = \text{const}\,, \tag{5.121'}$$

requiring

$$\varepsilon_{\text{Nucl}} \sim \left(1 - \frac{\Omega^2}{2\pi G\rho}\right)\,. \tag{5.122}$$

This condition is not obviously satisfied in a real star (von Zeipel 1924). Advection of thermal energy by circulation arranges that flux is conserved, but the fluid is then no longer barotropic but *baroclinic* ($P = P(\rho, T, \lambda_i)$), where the λ_i are variables depending on ϖ and z.

5.4.2 Circulation

We begin by estimating the order of magnitude of the circulation velocity in a star and discussing the physical problems arising from rotation.

We can easily estimate the meridional circulation velocity under the assumption that $\Omega^2 R/g$ is much less than 1. We separate the radiative flux into two components, one with spherical symmetry, and the other (denoted by a prime) without:

$$K_R \nabla T = (K_R \nabla T)_s + (K_R \nabla T)'\,. \tag{5.123}$$

The relative size of the two terms is of the order of the ratio between centrifugal force and gravity (e.g at the equator and the surface respectively):

$$\frac{(K_R \nabla T)'}{(K_R \nabla T)} \simeq \frac{\Omega^2 R}{g_s}\,. \tag{5.124}$$

Replacing the radiative flux $(-K_R \nabla T)$ by $(L/4\pi R^2)$ and the ∇ operator by $1/R$, we estimate the divergence of the radiative flux as

$$\nabla \cdot (K_R \nabla T) \simeq \frac{L}{4\pi R^3} \frac{\Omega^2 R}{g_s}\,, \tag{5.125}$$

which comes entirely from the $(K_R \nabla T)'$ component, since $\text{div}(K_R \nabla T)_s = 0$ in regions where $\varepsilon_{\text{Nucl}} = 0$. To order of magnitude, with $\frac{4}{3}\pi R^3 \bar{\rho} = M$, we get (dividing by ρ)

$$\frac{1}{\rho} \nabla \cdot (K_R \nabla T) = \frac{L}{M} \frac{\Omega^2 R}{g_s} = \frac{LR^3}{GM^2}\Omega^2\,. \tag{5.126}$$

We note that the Kelvin–Helmholtz timescale

$$\tau_{\rm KH} = \left(\frac{LR}{GM^2} \right)^{-1} \tag{5.127}$$

is also the heat-diffusion timescale through the star,

$$\tau_{\rm KH} \simeq \frac{R^2}{K}, \tag{5.128}$$

where K is the thermal diffusivity.

Then to order of magnitude we have

$$\frac{1}{\rho} \nabla \cdot (K_R \nabla T) = K \Omega^2. \tag{5.129}$$

The energy equation (Sect. 3.1.2) can be written

$$\rho T \left(\frac{\partial S}{\partial t} + \boldsymbol{u} \cdot \nabla S \right) = \nabla \cdot (K_R \nabla T) + \rho \varepsilon_{\rm Nucl}, \tag{5.130}$$

where S is the entropy. The product $\boldsymbol{u} \cdot \nabla S$ describes the advection of entropy from heat sources to heat sinks.

The mean entropy gradient can also be found to order of magnitude from the quantity $\Delta \nabla$,

$$\Delta \nabla = (\nabla_{\rm ad} - \nabla_*), \tag{5.131}$$

as

$$\nabla S \simeq= \frac{\Re (\nabla_{\rm ad} - \nabla_*)}{(\Re T / g \mu)}. \tag{5.132}$$

We thus get an estimate of the Eddington–Sweet circulation velocity

$$v_{\rm ES} \simeq \frac{LR^2}{GM^2} \frac{\Omega^2 R}{g_s} \frac{\bar{\rho}}{\rho} (\Delta \nabla)^{-1} \tag{5.133}$$

and a characteristic circulation timescale

$$t_{\rm ES} \simeq t_{\rm KH} \left(\frac{\Omega^2 R}{g_s} \right)^{-1}. \tag{5.134}$$

For fairly high (but easily attained) rotation speeds we have $t_{\rm ES} \lesssim t_{\rm Nucl}$, and we can expect rotation to have a significant effect on the internal structure. However, we note (see Mestel 1953 for details) that the circulations transport material from the star's core with higher mean molecular mass; the resulting composition gradient inhibits the circulation and significantly changes the morphology of the flowlines.

Reverting to the discussion of orders of magnitude, the thermal energy equation

$$C_v \rho \frac{dT}{dt} = \frac{p}{\rho} \frac{d\rho}{dt} + (\rho \varepsilon_{\text{Nucl}} - \nabla \cdot \boldsymbol{F}) \qquad (5.135)$$

can be written as

$$\rho G(\Phi)(\boldsymbol{v} \cdot \nabla \Phi) \equiv \rho C_v \left[\frac{dT}{d\Phi} - (\gamma - 1) \frac{T}{\rho} \frac{d\rho}{d\Phi} \right]$$
$$\times (\boldsymbol{v} \cdot \nabla \Phi) = (\rho \varepsilon_{\text{Nucl}} - \nabla \cdot \boldsymbol{F}) \qquad (5.136)$$

by means of the relation $d/dt = \boldsymbol{v} \cdot \nabla$. Now distinguishing between average quantities on the level surfaces and those whose means vanish on these surfaces,

$$\nabla \cdot \boldsymbol{F} = \langle \nabla \cdot \boldsymbol{F} \rangle_\Phi + (\nabla \cdot \boldsymbol{F})' \qquad (5.137)$$

with a potential

$$U = -\int^{\varpi} \Omega^2(\varpi') \varpi' d\varpi' , \qquad (5.138)$$

we have

$$(\nabla \cdot \boldsymbol{F})_s = \frac{d}{d\Phi} f(\Phi) \langle (\nabla \Phi)^2 \rangle + f(\Phi) \left[(\nabla^2 U)_\Phi - 4\pi G \rho \right] \qquad (5.139)$$

and

$$(\nabla \cdot \boldsymbol{F})' = \frac{d}{d\Phi} f(\Phi) \left[(\nabla \Phi)^2 \right]' + f(\Phi)(\nabla^2 U)' . \qquad (5.140)$$

The mass flux across a surface $\Phi =$ constant must vanish, so

$$\int_{\Phi=\text{const}} \rho(\Phi) \left(\boldsymbol{v} \cdot \frac{\nabla \Phi}{|\nabla \Phi|} \right) dS = \int_{\Phi=\text{const}} \frac{\rho \varepsilon_{\text{Nucl}} - \nabla \cdot \boldsymbol{F}}{G(\Phi)|\nabla \Phi|} dS = 0 . \quad (5.141)$$

To first order the mean quantities obey

$$\rho \varepsilon_{\text{Nucl}} - (\nabla \cdot \boldsymbol{F})_\Phi = 0 . \qquad (5.142)$$

The vertical component v_r is then given by (5.136). Assuming $\mu = $ constant,

$$g\rho(\nabla_{\text{ad}} - \nabla_*)v_r = -\frac{F}{g} \left[(\nabla^2 U)' \right] + \left[(\nabla \Phi)^2 \right]' \left[\frac{\rho}{g^2} \left(\frac{L}{M} - \varepsilon_{\text{Nucl}} \right) \right] . \qquad (5.143)$$

The variation of gravity caused by rotation is given by the variation of the potential, V being defined by (5.113). The relative variation is of order $(\nabla U / \nabla V)$, and we thus have

$$((\nabla \Phi)^2)' \simeq g^2 \frac{\nabla U}{g} . \qquad (5.144)$$

The velocity is then given by

$$v_r = \frac{1}{(\nabla_{\mathrm{ad}} - \nabla_*)} \left[\frac{L}{M} \frac{|\nabla U|}{g^2} - \frac{(\nabla^2 U)'}{4\pi r^2 \rho} \frac{L}{g^2} \right] . \tag{5.145}$$

The second term is proportional to $1/\rho$ and becomes infinite at the surface. It vanishes for the case $\Omega = $ constant, but there is still a $1/\rho$ term of second order in $\varepsilon = (\Omega^2 R^3 / GM)$, the ratio of centrifugal and gravitational forces at the surface.

5.4.3 Validity of the Assumptions

The description above ignored several potentially important effects:

(a) The velocity becomes infinite where $\Delta\nabla$ vanishes, i.e. at the frontier of the radiative and convective regions, and does not permit a description of how the circulation penetrates the convection zone.

(b) The radial velocity component becomes infinite at the surface $\rho = 0$, except in the case $\Omega^2 = $ constant (at least to first order), but the v_θ component,

$$v_\theta = -\frac{1}{\rho r \sin\theta} \frac{\partial}{\partial r} \left(\rho r^2 \int_0^\theta v_r \sin\theta \mathrm{d}\theta \right) , \tag{5.146}$$

always becomes infinite at the surface, since $\partial\rho/\partial r \neq 0$.

(c) The vector product $\boldsymbol{v} \times \boldsymbol{\Omega}$ is non-zero. It induces differential rotation which is indeterminate in the absence of viscosity, for the Coriolis force then implies an infinite azimuthal velocity.

(d) The velocity field induced by rotation has a wide variety of instabilities which have to be considered in any serious study of the internal structure. They can induce turbulence which in turn produces a "turbulent viscosity" and mixing phenomena as described below. There is a back-reaction (rotation–circulation–instabilities–circulation) that we have to consider.

(e) We have ignored magnetic fields. In reality the magnetic force adds to the centrifugal force,

$$\nabla U = \varpi \Omega^2 + \left[\frac{(\nabla \times \boldsymbol{B}) \times \boldsymbol{B}}{4\pi\rho} \right]_{\mathrm{poloidal}} , \tag{5.147}$$

and the azimuthal component adds to the viscous force. Magnetic effects are comparable to viscosity once

$$B^2 \simeq 4\pi\rho\nu\Omega \tag{5.148}$$

and for the Sun we see that a field of 2×10^{-2} G is equivalent to the (weak) viscosity of the gas. For fields of a few gauss we can expect noticeable effects (see Sect. 5.4.9).

Difficulties (a) and (b) are related to the absence of viscosity. The solution of (c) is simultaneously the solution of (a) and (b). Introducing viscosity raises the order of the system of differential equations, and we can find a non-singular solution characteristic of a *boundary layer*.

As an example we consider a star of mass $3\,M_\odot$ with a convective core and a radiative envelope. The microscopic viscosity is dominated by radiative viscosity,

$$\mu_R = \frac{4aT^4}{15c\kappa\rho}, \tag{5.149}$$

and we write the viscosity as $\mu = 10^q \mu_R$ to take empirical account of turbulent viscosity with a Reynolds number $Re = 10^q \geq 1, q \geq 0$. If $\kappa \equiv \sigma$ (electron scattering) the thicknesses δ_c, δ_S of the boundary layers at the surface of the convective core and the surface of the stars are given by (Tassoul and Tassoul 1982):

$$\frac{\delta_c}{R_c} = k_{\mathrm{CBL}} \left(\frac{L_*^2 R_*^2}{M_*^4} \right)^{1/7} 10^{(q/7)}, \tag{5.150}$$

$$\frac{\delta_S}{R_*} = k_{\mathrm{SBL}} \left(\frac{L_*^2 R_*^4}{M_*^4} \right)^{1/10} 10^{(q/10)} \tag{5.151}$$

with $k_{\mathrm{CBL}} = 3.17 \times 10^{-5}$, $k_{\mathrm{SBL}} = 6.18 \times 10^{-4}$. The boundary layer is therefore always very thin, even for large values of the turbulent viscosity. In most of the star the circulation is given by the non-viscous solutions.

In contrast the problem of differential rotation remains. In general solutions of the system of partial-differential equations describing the hydrodynamics of the stellar interior depend on both time and initial conditions. A stationary solution is a special case: the parity (in Ω^2) of the circulation velocity imposes an expansion of the form

$$\Omega = \omega_0 [\varepsilon^{1/2} + \omega_1 \varepsilon^{3/2} + \dots], \tag{5.152}$$

$$\boldsymbol{u} = \varepsilon \boldsymbol{u}_1 + \varepsilon^2 \boldsymbol{u}_2 + \dots, \tag{5.153}$$

where $\omega_0 = (GM/R^3)^{1/2}$ and $\varepsilon = \Omega^2 R^3/GM$. Then ω_1 describes differential rotation, and we have (with $\mu = \cos\theta$) (μ_R = radiative viscosity)

$$\frac{1}{r^4} \frac{\partial}{\partial r} \left(\mu_R r^4 \frac{\partial \omega_1}{\partial r} \right) + \frac{\mu}{r^2} \frac{1}{1-\mu^2} \frac{\partial}{\partial \mu} \left[(1-\mu^2) \frac{\partial \omega_1}{\partial \mu} \right]$$

$$= 2\rho \left(\frac{1}{r} u_r - \frac{\mu}{1-\mu^2} u_\theta \right), \tag{5.154}$$

where u_r, u_θ are the radial and poloidal components of the meridional circulation. For radiative viscosity, or phenomenological turbulent viscosity

$\mu_T = 10^q \mu_R$, (5.154) gives a characteristic timescale for the viscosity, which, for radiative viscosity, is of the order of t_{visc}:

$$t_{\text{visc}} \simeq 10^{-q} \frac{Mc^2}{L} = 1.6 \times 10^{13} 10^{-q} \frac{M_*}{L_*} \; (\text{years}).$$ (5.155)

t_{visc} is only comparable to the nuclear timescale on the main sequence (see Sect. 3.4.2):

$$t_{\text{Nucl}} = 11.2 \times 10^9 \frac{M_*}{L_*} \; (\text{years})$$ (5.156)

if $q \simeq 3$ to 4. To order of magnitude from (5.154)

$$\omega_1 \simeq \frac{ru}{\nu}$$ (5.157)

or, for radiative or turbulent viscosity and $\kappa \equiv \sigma$ (electron scattering),

$$\omega_1 \omega_0 \varepsilon \simeq \frac{45\pi}{4} \frac{\Omega^2 R^4 c^2}{G^2 M^2} = 4.3 \times 10^{13} \Omega^2 \frac{R_*^4}{M_*^2} 10^{-q},$$ (5.158)

$$\omega_1 \simeq \frac{Rc^2}{GM} = 5 \times 10^5 10^{-q} \frac{R_*}{M_*}.$$ (5.158')

The assumption of strong turbulent viscosity is the more reasonable the smaller the Prandtl number σ_P is (i.e. the less that radiative equilibrium is modified). Still with $\kappa \equiv \sigma$ we have

$$(\sigma_P)_{\text{turb}} = \frac{C_v T}{5c^2} \times 10^q \simeq 3 \times 10^{-14} \cdot 10^q T,$$ (5.159)

and except in central regions $\sigma_{P\,\text{turb}} \ll 1$.

The study of time-dependent solutions removes the difficulty with the value of the viscosity, whether for the convergence of the series (5.152) or for the Prandtl number. The basis is still an expansion of the form (5.152), but this time introducing time-dependent terms ($\mu = \cos\theta$):

$$\Omega = \Omega_0 \left\{ \theta(r,t) + \varepsilon \left[\theta_1(r,t) \frac{dP_1(\mu)}{d\mu} + \theta_3(r,t) \frac{dP_3(\mu)}{d\mu} \right] \right\},$$ (5.160)

where Ω is the average angular velocity.

For full rigour we must write

$$\Omega = \Omega_0 \{ \omega(r,\mu,t) + \varepsilon \omega_1(r,\mu,t) + \varepsilon^2 \omega_2(r,\mu,t) + \dots \}.$$ (5.161)

In fact for small ε there is separation between the evolution equation for $\omega(r,\mu,t)$ and the equations giving the circulation velocity and the differential rotation. Assuming that the boundary conditions retain spherical

symmetry (zero-order approximation), $\omega(r,\mu,t)$ reduces to $\theta(r,t)$ and satisfies the diffusion equation

$$\frac{\partial\theta}{\partial t} = \frac{\eta}{\rho}\left[\frac{\partial^2\theta}{\partial r^2} + \left(\frac{4}{r} + \frac{1}{\eta}\frac{d\eta}{dr}\right)\frac{\partial\theta}{\partial r}\right], \tag{5.162}$$

where η is the dynamical viscosity. Again as an example we can assume as boundary condition Skumanich's empirical relation $\Omega \sim t^{-1/2}$, giving the angular velocity at the surface R_{rad} at the surface of the radiative zone as a function of time, and write

$$\theta(R_{\mathrm{rad}},t) = [1 + (\sigma^2 - 1)(t/t_\theta)]^{-1/2}. \tag{5.163}$$

We can clearly specify a θ-dependent turbulent viscosity, and (5.162) becomes a non-linear diffusion equation.

5.4.4 The Classical Solution (Sweet 1950)

However it arises, the case $\theta(R_{\mathrm{rad}},t) = \mathrm{const}$ is typical. Ignoring molecular weight gradients gives Sweet's (1950) solution for an inviscid fluid. The velocity components of the meridional circulation are valid except in boundary layers.

To first order in ε we then find for the perturbation $P_2(\cos\theta)$ of the potential

$$V_{1,2}P_2(\cos\theta), \tag{5.164}$$

$$V_{1,2} = h(r) - \tfrac{1}{3}\omega_0^2 r^2, \tag{5.165}$$

$$h'' + \frac{2}{r}h' - \frac{6}{r^2}h + 4\pi G\frac{\rho\rho'}{p'}h = 0, \tag{5.166}$$

and we deduce the corresponding pressure, density, and temperature perturbations

$$p_{1,2} = -\rho h, \tag{5.167}$$

$$\rho_{1,2} = -\frac{\rho\rho'}{p'}h, \tag{5.168}$$

$$T_{1,2} = -T\left(\frac{\rho}{p} - \frac{\rho'}{p'}\right)h. \tag{5.169}$$

The radial component u can be written ($\mu = \cos\theta$)

$$u(r,\mu) = u(r)P_2(\mu), \tag{5.170}$$

$$u(r) = \frac{2Lr^4}{G^2M^3}\frac{n+1}{n-\frac{3}{2}}\left[h' + \left(\frac{2}{r} - \frac{M'}{M}\right)h\right], \tag{5.171}$$

214
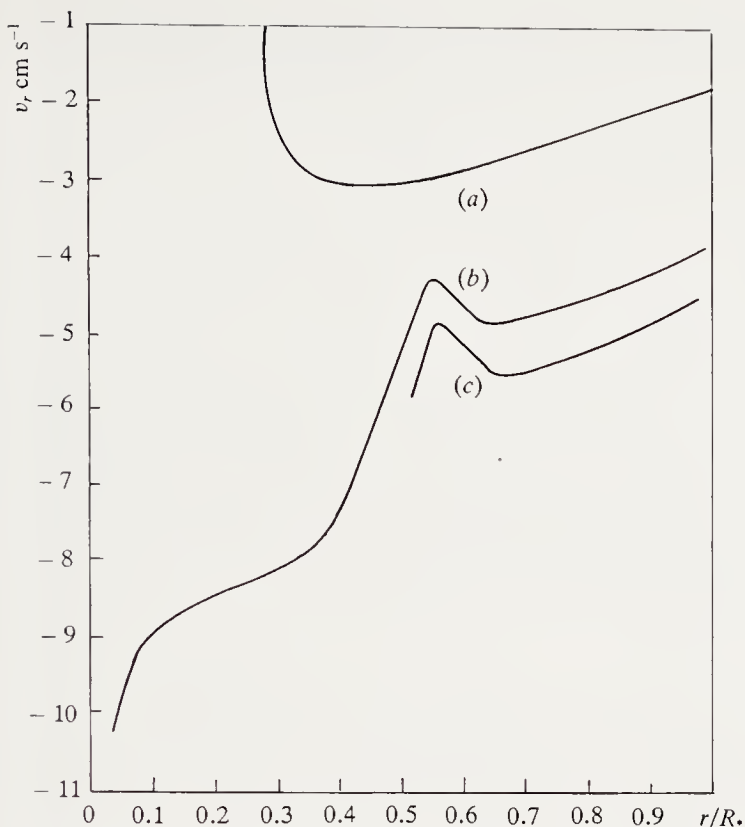
Fig. 5.3. Radial component of meridional circulation in the Sun. (a) Standard solution for a Cowling model (central point energy source surrounded by a convection zone). (b) Solution for the solar interior at $t = 4.6 \times 10^9$ y, assuming initially rigid rotation at the same angular velocity as today. (c) Solution for the solar interior assuming a decrease of Ω following the Skumanich law and initially rigid rotation at four times the present surface velocity

where n is the local polytropic index and M the mass. The poloidal component is then given by

$$v_\theta(r) = \frac{1}{6} \frac{1}{\rho r^2} \frac{\mathrm{d}}{\mathrm{d}r}(\rho r^2 u) \tag{5.172}$$

with the velocity

$$v(r, \mu) = r v_\theta(r)(1 - \mu^2)\frac{\mathrm{d}P_2}{\mathrm{d}\mu} . \tag{5.173}$$

Sweet's solution ignores the term of second order in ε which however gives another divergence through its contribution to $\bar{\rho}/\rho$ (Fig. 5.3).

A basic theorem (Gratton 1945; Öpik 1951) for which we shall give a simple proof (Mestel 1957) shows the existence of two zones with oppositely-directed circulations. Returning to (5.136) and using (5.116) we have

$$\rho \varepsilon_{\mathrm{Nucl}} - f(\Phi)(4\pi G\rho - 2\Omega^2) - f'(\Phi)g^2 = \rho G(\Phi)\boldsymbol{u} \cdot \nabla\Phi , \tag{5.174}$$

215

where $g = \mathrm{d}\Phi/\mathrm{d}n$ is the effective gravity. Dividing by g and taking the mean, and using the fact that the mass flux vanishes across a level surface, we get

$$\left[\rho\varepsilon_{\mathrm{Nucl}} - f(\Phi)(4\pi G\rho - 2\Omega^2)\right]\langle g^{-1}\rangle = f'(\Phi)\langle g\rangle . \tag{5.175}$$

Combining with (5.174) and (5.175) gives

$$\rho G(\Phi)\boldsymbol{u} \cdot \nabla\Phi = f'(\Phi)\left(\frac{\langle g\rangle}{\langle g^{-1}\rangle} - g^2\right) . \tag{5.176}$$

In a radiatively stable zone, motion against gravity requires energy input; the component of velocity across a level surface is positive if the right-hand side of (5.176) is positive. If $f'(\Phi)$ vanishes, the relation $f(\Phi) = 0$ defines a level surface which the circulation does not cross. In the outer regions, where $\varepsilon_{\mathrm{Nucl}} = 0$, this defines via (5.175) a density ρ_{inv} for the region where the circulation has an inversion:

$$2\pi G\rho_{\mathrm{inv}}(\Phi_{\mathrm{inv}}) = \Omega^2 . \tag{5.177}$$

Inside the surface $\Phi_{\mathrm{inv}}, 2\pi G\rho > \Omega^2, f'(\Phi) < 0$. At the pole, where g is a maximum, the right-hand side of (5.176) is positive and the circulation rises near the pole and falls near the equator (Fig. 5.4).

The reversal of the circulation does not occur (1) if Ω is not constant throughout the star, and (2) if the condition (5.177) on ρ_{inv} corresponds to a level surface in the boundary layer or the convection zone. From models of the outer layers we can eliminate $\Delta R/R$ between (5.151) (where we take $\delta_s = \Delta R$) and the relation for $\rho(\Delta R/R)$. As an example, for $\kappa \equiv \sigma$ we find

$$\Omega^2_{\mathrm{crit}} = 3.2 \cdot 10^{-15}(M_*^2 L_*^4 R_*^{18})^{-1/10}10^{(3/10)q} ,$$

corresponding to extremely slow rotation. Assuming $R_* \sim M_*^{3/4}$, $L_* \sim M_*^3$, and even taking $q = 6$, the condition gives $P > 3000$ d for $M_* = M_\odot$. This order of magnitude is large enough for us to conclude that every real star has an inversion of its circulation. To order of magnitude for this same model we find that the inversion is always above $0.7R_*$, corresponding to a density

Polar radius
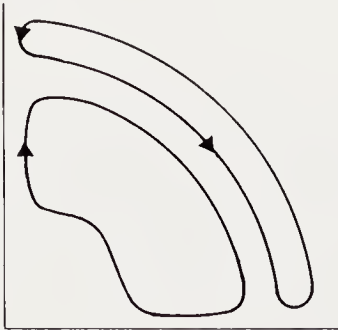
Equatorial radius

Fig. 5.4. A consequence of von Zeipel's theorem. The sense of the circulation changes on the surface defined by $2\pi G\rho_{\mathrm{inv}} = \Omega^2$

of about $0.9(M_*/R_*^3)\,\mathrm{g\,cm^{-3}}$. This is clearly different for a time-dependent model. We recall that the above discussion is for the case where radiative viscosity dominates. We note however with Tassoul (see Fig. 5.5) that in any real star, where Ω is a function of ϖ and z, there is no circulation inversion.

5.4.5 μ Currents and Ω Currents

As a star evolves, the chemical composition of its central regions changes on the nuclear timescale τ_{Nucl}. Eddington–Sweet circulation currents bring matter of higher mean molecular mass to the surface. The buoyancy forces resulting from this excess molecular mass tend to oppose the motions induced by rotation. If the excess molecular mass is sufficiently large, the circulation can be halted.

We can estimate the condition for the inhibition of Ω currents to order of magnitude. We consider a motion v_μ driven by a difference in mean molecular masses. We adopt a phenomenological description analogous to the mixing-length theory. An element of molecular mass μ_i larger than the molecular mass μ_0 of the environment is approximately in pressure equilibrium, and its temperature T_i is lower than that of the surroundings, T_0. We then assume that the advection of heat caused by the motion of the element is equal to the rate of radiative energy exchange with the surroundings. Letting τ_{rad} be the characteristic energy-exchange time, local theory gives

$$\frac{T_i \Delta\nabla}{H_P} v_\mu = (T_i - T_0)\frac{1}{\tau_{\mathrm{rad}}}$$

with

$$\Delta\nabla = (\nabla_{\mathrm{ad}} - \nabla_*).$$

The temperature difference $(T_i - T_0)$ is thus given by the relation

$$\frac{T_i - T_0}{T_0} = -\frac{\mu_i - \mu_0}{\mu_i}.$$

To relate $(\mu_i - \mu_0)$ to the stars's evolution we assume that this difference arises because Eddington–Sweet circulations require a certain time to bring the new chemical composition to level r,

$$\frac{\mu_i - \mu_0}{\mu_i} = \frac{\Delta\mu}{\mu}\frac{\tau_{\mathrm{ES}}}{\tau_{\mathrm{Nucl}}},$$

where $\Delta\mu$ is the change of chemical composition caused by thermonuclear reactions. In other words, the slower the circulation, the larger the change in composition, because more time is needed to bring the new composition to level r. Assuming a perfect gas, we thus obtain a velocity v_μ,

$$v_\mu = \frac{H_P}{\Delta\nabla\tau_{\mathrm{rad}}}\frac{\tau_{\mathrm{ES}}}{\tau_{\mathrm{Nucl}}}\frac{\Delta\mu}{\mu}.$$

The velocity of the Eddington–Sweet currents is related to the timescale τ_{ES} through the relation

$$\tau_{ES} \simeq (r/v_{ES})$$

with v_{ES} from (5.133). We rewrite it to order of magnitude:

$$v_{ES} = \frac{\Omega^2 r^3}{GM_r} \frac{\overline{\rho}(r)}{\rho(r)} (\Delta\nabla)^{-1} \frac{LR^2}{GM^2} \nabla_{ad} \ .$$

We note that in this analysis v_μ is of the order of ε^{-1}, which at first sight appears absurd, but comes from the fact that the molecular mass gradient opposing the motion is larger for slower circulations. Thus we find that circulation occurs at the surface of a convective core of mass qM_* if ε, the ratio of centrifugal force to gravity, satisfies the condition

$$\varepsilon^2 > 4 \frac{\tau_{KH}}{\tau_{Nucl}} \frac{\Delta\nabla}{\nabla^2_{ad}} \frac{\Delta\mu}{\mu} \left(\frac{\rho}{\overline{\rho}}\right)^2 \left(\frac{\rho_c}{\overline{\rho}}\right)^2 \ .$$

With $\tau_{KH}/\tau_{Nucl} \simeq 4 \times 10^{-3}(M_*/R_*)$, we find for rigid rotation that

$$\varepsilon^2 \gtrsim 0.8(M_*/R_*)\,.$$

The assumption $\varepsilon \ll 1$ is clearly violated here, and the order of magnitude we find simply shows that near the core molecular mass gradients inhibit the rotation-induced circulation.

To find the circulation exactly in the presence of μ gradients we have to look for a solution with μ expanded in powers of ε as

$$\mu = \mu_0(r,t) + \varepsilon\left[\mu_{1,0}(r,t) + \mu_{1,2}(r,t)P_2(\cos\theta)\right]\,.$$

Tassoul and Tassoul (1984) have found such a solution to order ε for a $1\,M_\odot$ star (Fig. 5.5). The growth of the μ-gradient is clearly seen to expel the flowlines from the core. To order ε this expansion does not determine a value ε_{crit} above which the flowlines penetrate the core and mix the stellar material, as to first order the topology of the flowlines is independent of ε. ε_{crit} can only be found by going to second order in ε. Then the velocity of the meridional circulation depends on the deviation from rigid rotation defined by ω_1,

$$\Omega = \Omega_0(1 + \varepsilon\omega_1 + \dots)\,,$$

where the solution for ω_1 depends on the choice of viscosity (size, anisotropy of turbulent viscosity).

5.4.6 Instabilities

The probable presence of turbulent viscosity in stars is caused by the many instabilities associated with rotation. Helioseismology (see Chap. 6) shows
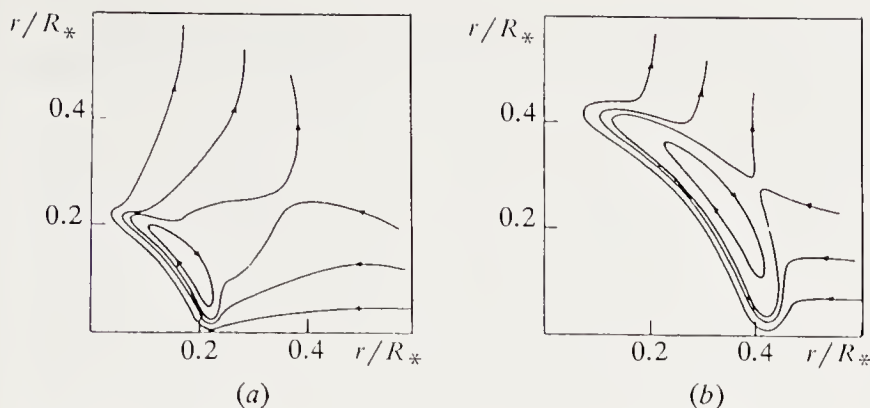
Fig. 5.5. Meridional circulation in a rotating star. The effect of composition inhomogeneity for a $1\,M_\odot$ star. Flowlines in a meridian plane. In (a), circulation at 500 million years, in (b) at 5 billion years. The change of molecular mass in the central regions expels the flowlines. (From J.L. Tassoul and M. Tassoul, *Astrophys. J.* **279**, 384 (1984). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1984 The American Astronomical Society)

that the rotation of the Sun's interior is near rigid, allowing an elementary argument which we shall consider later. The timescale needed to establish such rotation is of the order of

$$t_{\text{visc}} \simeq \frac{R^2}{\nu_{\text{total}}} \, .$$

In the Sun, with $t = t_\odot = 4.6 \times 10^9$ y, $R = R_\odot$, this gives

$$\nu_{\text{total}} \gtrsim 34\,000 \, ,$$

while the microscopic viscosity is of the order of 10 from centre to surface. This total viscosity can only result from instabilities which maintain it, although we cannot determine its precise nature (hydrodynamic turbulence? MHD turbulence? See Sect. 5.4.9 on dynamo theory). Another possibility is that angular momentum is carried away by internal waves, through interaction between the waves and the mean flow.

There are very many instabilities and it is difficult to classify them systematically. Some can be regarded as dynamical, in the sense that they exist for zero viscosity and thermal diffusivity; other instabilities can be induced by viscosity and/or thermal diffusivity. In a rotating system with an axis of symmetry there are both *axisymmetric* and *non-axisymmetric* instabilities. Finally there are both *local* and *global* instabilities. We shall consider the latter here: global instabilities are considered in Chap. 6 (variable stars).

Dynamical Instabilities. Such instabilities occur in a fluid where at each point the equilibrium is unstable with respect to an adiabatic inviscid perturbation. A local instability of this kind produces a new static equilibrium, or a new state of motion where non-linear effects rapidly become dominant

and viscous and thermal diffusivity effects can no longer be ignored. All dynamical instabilities have the same growth time, which is of the order of the rotation period.

Axisymmetric Instabilities.   A very general method, due to Fjörtoft (1946), uses an energy principle. Zahn (1974) considers the perturbation of an equilibrium state by a field $\delta r(\varpi, z)$ of axisymmetric displacements in meridian planes, with a timescale short enough to be regarded as adiabatic.

  As the fluid is in a static equilibrium state the first variation δW of the total energy W (internal + kinetic + gravitational) is zero. The second variation $\delta^2 W$ is given by the volume integral of a quadratic form

$$\delta^2 W = \int \left[\boldsymbol{\delta r} \cdot M \cdot \boldsymbol{\delta r} + \frac{1}{\gamma} \frac{P}{\rho} \left(\frac{\delta P}{P} \right)^2 \right] \rho dV \,,$$

where δP is the Eulerian pressure perturbation associated with the displacement $\delta r(\varpi, z)$ and M is a tensor representing the density and angular momentum stratification:

$$M = M_1 + M_2$$
$$= \frac{1}{C_p}(-\boldsymbol{g})\nabla S + \frac{1}{\varpi^3}\nabla(\varpi^2 \Omega)^2 \nabla \varpi \,,$$

where g is the local gravity (including centrifugal force) and S the specific entropy

$$\frac{1}{C_p}\nabla S = \frac{1}{\gamma P}\nabla P - \frac{1}{\rho}\nabla\rho = \frac{1}{\rho}\left(\nabla_{\mathrm{ad}}\rho - \nabla\rho\right) \,.$$

If the characteristic time of the perturbation is longer than the propagation time for a sound wave across the region considered, the second term in the integral is negligible compared with the first, whose sign determines the sign of $\delta^2 W$.

  If the quadratic form $\boldsymbol{\delta r} \cdot M \cdot \boldsymbol{\delta r}$ is positive definite, the stationary value of W is a minimum and the equilibrium is stable. If the quadratic form is negative definite we can find a displacement δr which decreases the total energy, and the equilibrium is unstable. The convenience of an energy principle is that the displacement δr is arbitrary except that it must vanish on the boundaries of the integration region and its timescale must be compatible with the conditions of adiabaticity and pressure equilibrium.

The Rayleigh Criterion.   For a fluid with a density distribution such that $\nabla S = 0$, the first quadratic term vanishes. Instability occurs when either of the following conditions is satisfied:

$$\frac{\partial}{\partial \varpi}(\varpi^2 \Omega)^2 < 0\,, \qquad \frac{\partial}{\partial z}(\varpi^2 \Omega) \neq 0\,.$$

The first of these is the Rayleigh criterion for instability of a homogeneous, non-gravitating, inviscid fluid. It is a sufficient condition for instability. The second condition is much stronger, since if it holds there can be no equilibrium. This is a result of the Taylor–Proudman theorem: for a slowly rotating fluid the linearised equation of motion can be written as

$$\frac{\partial v}{\partial t} + 2\Omega \wedge v = -\nabla P + \rho \nabla \Phi.$$

Taking the curl of each side we have for an incompressible fluid

$$(u_z \cdot \nabla)v = 0,$$

and the velocity of fluid elements must be independent of the coordinate along the axis of rotation.

The stability condition

$$\frac{\partial}{\partial \varpi}(\varpi^2 \Omega)^2 > 0$$

can easily be explained through the conservation of angular momentum when a fluid element is displaced from its equilibrium position. If the stability criterion is met, an element displaced outwards has a smaller angular momentum than its surroundings. The centrifugal force on it is smaller than on the neighbouring elements, decelerating its outward motion. Similarly, if the element is displaced inwards it feels a restoring force outwards which arrests its motion. In fact, as we shall see below, stability requires that the specific entropy decrease and the angular momentum increase outwards.

Barotropic Instability. Let us consider a star with rotation law $\Omega(\varpi)$ and ask if a stable density stratification can inhibit the Rayleigh instability. We assume the fluid to be *barotropic*, i.e. the equipotential surfaces are also constant pressure and density surfaces. The quadratic form

$$Q_1 = \delta r \cdot \frac{1}{C_p}(-g)\nabla S \cdot \delta r$$

is positive definite. The quadratic form

$$Q_2 = \delta r \cdot \frac{1}{\varpi^3}\frac{\partial}{\partial \varpi}(\varpi^2 \Omega)^2 \nabla \varpi \cdot \delta r$$

must be positive definite for the total form $Q_1 + Q_2$ to be positive definite (because of the symmetry of the equatorial plane it is always positive definite there). Put another way, if the condition for the Rayleigh instability is met, there are always displacements δr which cause a negative variation $\delta^2 W$ of the energy. The unstable displacements are those insensitive to the density stratification and occur on equipotentials.

Baroclinic Instabilities. This result can be extended to more general rotation laws. For a baroclinic fluid, ∇S and g are not parallel, and the angle between them is given by the equation of equilibrium

$$\frac{1}{\rho}\nabla p + \nabla V = \Omega^2 \varpi \nabla \varpi \,,$$

where V is the gravitational potential. Taking the curl of both sides we can write

$$\frac{1}{C_p}\nabla S \times (-g) + \frac{1}{\varpi^3}\nabla(\varpi^2\Omega)^2 \times \nabla\varpi = 0 \,.$$

The necessary and sufficient condition for the quadratic form $Q_1 + Q_2$ to be positive is then

$$\left[\frac{1}{C_p}\nabla S \times \frac{1}{\varpi^3}\nabla(\varpi^2\Omega)^2\right] \cdot [-g \times \nabla\varpi] > 0 \,.$$

If this condition is violated we can always find a displacement which decreases the energy, $\delta^2 W < 0$. The configuration is unstable to the *baroclinic instability*. It occurs on a surface of constant entropy.

Calling N the Brunt–Väisälä frequency,

$$N^2 = \frac{g}{H_P}(\nabla_{\rm ad} - \nabla_{\rm rad}) \,,$$

one can show (Zahn 1983) that the baroclinic instability corresponds to the condition

$$\frac{\partial \log \Omega}{\partial \log r} \gtrsim \frac{H_P}{r}\left(\frac{N}{\Omega}\right)^2$$

for a locally constant shear.

One can also show (Zahn 1983) that for a shear with a point of inflection of the form $U = U_0 \tanh(x/H)$ the baroclinic instability on a sphere corresponds to the condition

$$\frac{\partial \log \Omega}{\partial \log r} = \left(\frac{H_P}{r}\right)^2\left(\frac{N}{\Omega}\right)^2 \,.$$

Shear (Richardson) Instability. In a plane-parallel shearing flow a sufficient condition for stability is that the velocity field $v_x(z)$ has no point of inflection. This theorem was also proved by Rayleigh and its equivalent for a cylindrical flow is that the expression

$$\frac{\rm d}{{\rm d}\varpi}\left[\frac{1}{\varpi}\frac{\rm d}{{\rm d}\varpi}(\varpi^2\Omega)\right]$$

should not change sign in the domain considered. In fact viscosity can destabilise a shear flow if the Reynolds number $Re = VL/\nu$ is larger than a critical value of the order of 10^3, where V and L are the typical velocities and lengthscales of the region.

The flow can be stabilised by a stable density stratification. We describe the stability criterion as follows. A fluid element displaced vertically by l feels a restoring force of acceleration $N^2 l$, where N is the Brunt–Väisälä frequency, because of the density gradient. Its velocity changes by $\Delta v = l(dv/dz)^2$. The shear flow is stable if

$$N^2 l \gtrsim l(dv/dz)^2 \; .$$

We call

$$Ri = N^2 \left(\frac{dv}{dz}\right)^{-2}$$

the Richardson number, and one can show [1] that the stability criterion is

$$Ri > \tfrac{1}{4} \; ,$$

and for a wide variety of profiles the condition $Ri < \tfrac{1}{4}$ is a sufficient condition for instability. For a rotating star the local stability criterion is

$$N^2 \cos^2 \alpha > \tfrac{1}{4}(\varpi \, \mathrm{grad}\, \Omega)^2 \; ,$$

where α is the angle between g and $\nabla \Omega$.

Diffusive Instabilities (Viscosity, Thermal Diffusivity).   Transport processes greatly modify the behaviour of fluids and can easily destabilise an otherwise stable situation. We have already mentioned the effect of viscosity on shearing flows and the importance of the Reynolds number. We now examine the effect of thermal diffusivity.

Shear Flows (Townsend).   Heat exchange with the surroundings reduces buoyancy forces on a fluid element. If t_{ex} is the characteristic timescale for heat exchange (see Sect. 5.3 on convection), the equation giving the thermal evolution of an element simplifies to

$$\frac{d}{dt}(T_i - T_0) = -\frac{T_i - T_0}{t_{\mathrm{ex}}} + \frac{\Delta\nabla}{H_P}T\frac{dv}{dz} \; .$$

If t_{ex} is much shorter than the timescale dz/dv of the shear flow, the new temperature difference is reduced to

$$(T_i - T_0) = \frac{\Delta\nabla}{H_P}T\left(t_{\mathrm{ex}}\frac{dv}{dz}\right) \; ,$$

---

[1] Miles, J.W., *Journal of Fluid Mechanics* **10**, 496 (1961).

and we have the stability condition

$$N^2 l \left(t_{\text{ex}} \frac{\mathrm{d}v}{\mathrm{d}z} \right) \gtrsim l \left(\frac{\mathrm{d}v}{\mathrm{d}z} \right)^2 .$$

For thermal diffusivity K and typical turbulent eddy scale L in the optically thick inviscid limit, Dudis (1974) introduced a new Richardson number (Richardson–Townsend criterion)

$$Ri' = 0.667(K/VL)$$

with

$$t_{\text{ex}} = (L^2/K).$$

Zahn (1974) studied turbulent fluids in which thermal diffusivity sustains the turbulence. The conjecture is that the smallest fluid elements must satisfy the definition of the critical Reynolds number

$$Re_c = \frac{1}{\nu} l^2 \frac{\mathrm{d}v}{\mathrm{d}z} .$$

Replacing the heat exchange time by l^2/K we have to order of magnitude the instability condition

$$N^2 \left(\frac{\mathrm{d}v}{\mathrm{d}z} \right)^{-2} \left(\frac{\nu}{K} \right) Re_c \lesssim 1 ,$$

which now depends on the Prandtl number ν/K. For a rotating star this is

$$\left(\cos\theta \frac{\partial \log \Omega}{\partial \log r} \right)^2 \gtrsim \left(\frac{\nu}{K} Re_c \right) \left(\frac{N}{\Omega} \right)^2 ,$$

where θ is the colatitude. Setting $\partial \log \Omega / \partial \log r = n$ (if $n = \text{constant}$, $\Omega \sim r^n$) we find the condition on the period

$$P_{\text{crit}} \lesssim \left(\frac{K}{\nu \, Re_c} \right)^{1/2} \frac{1}{N} 2\pi n .$$

In a region where viscosity is radiative

$$P_{\text{crit}} < n \left(\frac{g_\odot}{g} \right) \times 11.6 \, \text{days} .$$

For the Sun, where molecular viscosity is larger than radiative viscosity, the corresponding critical period is smaller. For a Kramers opacity law and $Z = 0.02, X = 0.7$, with

$$\Omega^2 = \frac{1}{n^2} 5.48 \times 10^{26} \frac{Z(1+X)}{\log \Lambda} \left(\frac{g}{g_\odot} \right)^2 \mu \Delta \nabla \frac{\rho^2}{T^5} \qquad \text{and}$$

$$\nu_{\mathrm{mol}} = \frac{2m_p^{1/2}(kT)^{5/2}}{5e^4\rho\log\Lambda} \quad \text{and} \quad \Lambda = \frac{3}{2e^3}\left(\frac{m_p k^3 T^3}{\pi\rho}\right)^{1/2},$$

we find

$$P_{\mathrm{crit}} \lesssim n\left(\frac{g_\odot}{g}\right)\frac{T^{5/2}}{\rho}8.29\times10^{-12}.$$

Taking the base of the convection zone as the reference we have

$$P_{\mathrm{crit}} \lesssim n\frac{g_\odot}{g}\left(\frac{T}{1.73\times10^6}\right)^{2.5}\left(\frac{0.117}{\rho}\right)\times3.23\,\mathrm{days}.$$

The Richardson–Townsend instability condition requires very rapid rotation (100 km s^{-1} at the equator) for a star like the Sun. It is by contrast much easier to satisfy for intermediate-mass stars.

Multi-diffusive Instabilities. The most important of these is the GSF instability (Goldreich–Schubert–Fricke). Taking account of viscosity, the Rayleigh criterion can be written

$$\frac{1}{\varpi^3}\frac{\partial}{\partial\varpi}(\varpi^2\Omega)^2 < \frac{\nu}{K}N^2, \qquad \left|\varpi\frac{\partial\Omega}{\partial z}\right| > \frac{1}{K}N^2.$$

These conditions are less severe than those for inviscid fluids, particularly for slow rotators like the Sun. We can show that a molecular mass gradient stabilises the flow once

$$N_\mu^2 = \frac{g}{H_P}\nabla\mu > \frac{\nu}{K}N^2.$$

The GSF instability occurs in almost every circulation induced by rotation. Kippenhahn and Thomas (1981) have shown that phenomenological introduction of nonlinear processes associated with the GSF instability gives the order of magnitude of the diffusion process engendered. The basic idea, by analogy with μ currents, is to estimate the deviation from thermal equilibrium caused by the displacement of an unstable fluid element.

If a fluid torus of radii ϖ_0 and d expands to $\varpi_0 + l$ it must adjust its density to remain in hydrostatic equilibrium with its new surroundings, which implies a decrease of d. The temperature inside the torus is then lower than outside:

$$\frac{T_i - T_e}{T_i} = 2\varepsilon l\frac{\mathrm{d}\log(\Omega\varpi^2)}{\mathrm{d}\varpi} < 0,$$

where ε is a mean value of the ratio of centrifugal force to gravity. We assume that radiative exchange between the torus and the surroundings is compensated by advection of entropy by the motion. If τ_* is the characteristic radiative exchange time

$$\tau_* \simeq \frac{3C_p \kappa \rho^2 \xi d^2}{8acT^3} ,$$

where ξ is a geometrical factor, we find the velocity

$$v_\Omega = -\frac{H_P}{\Delta \nabla \tau_*} 2\varepsilon l \frac{d \log(\Omega \varpi^2)}{d \varpi} ,$$

where H_P is the pressure scaleheight.

The heat exchange with the torus causes a circulation of matter which mixes it with the surroundings on a timescale l/v_Ω. This occurs over a lengthscale of the order of the cross-sectional (small) diameter of the torus and defines a diffusion coefficient

$$D_\Omega = \frac{H_P}{\Delta \nabla \tau_*} 2\varepsilon d^2 \frac{d \log(\Omega \varpi^2)}{d \varpi} .$$

To order of magnitude we can define a Kippenhahn–Thomas Reynolds number $Re(\mathrm{KT})$. For radiative viscosity we have

$$Re(\mathrm{KT}) = 8\varepsilon \frac{c^2}{Rg\mu \Delta \nabla \xi} .$$

From the circulation velocity v_θ Baglin (1972) finds a Reynolds number of the order of $v_\theta H_P/\nu$ of the same order as $Re(\mathrm{KT})$, for turbulent diffusion produced by violation of the Rayleigh criterion (GSF instabilities) or classically by the shear of meridional flows. We shall return to this in connection with observational tests.

We ignore other instabilities here. However it is important to discuss how a composition gradient can stabilise the medium.

5.4.7 Stabilisation by a μ Gradient

GSF Instability. If Λ is the angle between g_eff and the equatorial plane, Γ the angle between $\mathrm{grad}(\varpi^2 \Omega)$ and this plane, K_μ the molecular diffusivity (K_μ is of the order of ν), and N_μ the Brunt-Väisälä frequency associated with $\mathrm{grad}\,\mu$,

$$N_\mu^2 = \frac{g}{H_P} \nabla \mu ,$$

the GSF instability condition becomes

$$\frac{\sin^2 \Lambda}{4 \sin \Lambda \sin(\Lambda - \Gamma)} N_\Omega^2 > \frac{\nu}{K} N^2 + \frac{\nu}{K_\mu} N_\mu^2$$

with $N_\Omega^2 = \varpi^{-3}(\partial/\partial \varpi)(\varpi^2 \Omega)^2$. Using $\nu/K \ll 1$, while $\nu/K_\mu = O(1)$, it is clear that a very small chemical composition gradient can inhibit the GSF instability.

Shear Instability. It suffices to use the Richardson criterion, replacing the density gradient by the molecular mass gradient. The instability condition becomes

$$\frac{1}{4}(\varpi \operatorname{grad}\Omega)^2 > -\cos^2\alpha\,|g|\frac{\mathrm{d}}{\mathrm{d}z}\log\mu$$

where we have assumed that the gradient of μ only depends on z, with α the angle between g and $\nabla\Omega$.

5.4.8 Turbulence and Mixing

Turbulence. We have invoked turbulence several times already: in convection zones, instabilities in radiatively stable zones, and the relation between meridional circulation and rotation. We do not of course wish to develop a theory of turbulence here, but adopt a simplified phenomenological description. By turbulence we mean random motion, describable by a number of average quantities (e.g. energy density, turbulent viscosity, turbulent diffusion), whose significance and origin we have to consider.

Two driving mechanisms have been pointed out, one associated with Rayleigh–Bénard instabilities in convection zones, and the other, in radiative regions, associated with motions driven by rotation (Rayleigh, shear and GSF instabilities). In each case if we define the Reynolds number

$$Re = \frac{LV}{\nu},$$

where L and V are characteristic lengths and velocities, we find very large values. In a convection zone, $Re \simeq 10^{11}$–10^{13} cm^2 s^{-1}; in a star like the Sun the differential rotation on an equipotential between pole and equator corresponds to $Re \simeq 10^{10}$–10^{12}, according to helioseismology data. Such very high values of the Reynolds number are in line with ideas of *fully developed turbulence*. The theory of turbulence is far from complete, and in the following we shall point out a number of elements of it which are applicable to astrophysics.

In fully developed turbulence we assume there are three scales:

(a) *Large eddy scales* which carry energy. We assume that some hydrodynamic or convective instability feeds energy ε_I into the turbulence and defines a scale $l_0 = 1/k_I$.

(b) *Dissipative scales* $\lesssim l_D$. Their dynamics results from a competition between nonlinear effects and linear dissipation caused by the molecular viscosity ν_0; kinetic energy is dissipated as heat at the rate ε_d.

(c) *Between the two scales*, if the Reynolds number is large enough there is an *inertial regime* where the nonlinear terms of the Navier–Stokes equation dominate. On this scale energy is neither produced nor dissipated but transferred to smaller and smaller sales at a rate independent of both time

and the scale itself. The rates of injection, transfer, and dissipation of energy must on average be equal, and we denote this mean value by ε.

We define the energy spectrum by considering statistically homogeneous steady turbulence. The average velocity $\langle v \rangle$ is zero, but the correlation

$$C_{ij} = \langle v_i(\boldsymbol{x}, t) v_j(\boldsymbol{x} + \boldsymbol{r}, t) \rangle$$

must be a function of the spatial separation r only. The assumption of plane symmetry and isotropy requires the tensor $C_{ij}(\boldsymbol{r})$ to be characterised by its trace $C_{ii}(\boldsymbol{r}) = C(r)$ and to depend only on the modulus $|\boldsymbol{r}| = r$. The energy spectrum $E(k)$, a function of the wavenumber k, is $4\pi k^2 \mathrm{FT} C(r)$, where $\mathrm{FT} C(r)$ is the spatial Fourier transform of $C(r)$, and usually normalised so that

$$\int_0^\infty E(k)\mathrm{d}k = \frac{1}{2}\langle v^2 \rangle \,.$$

In the regime of inertial turbulence, one finds experimentally $E(k) = M k^{-m}$ with $m = 5/3$. This is the result found through dimensional analysis by Kolmogorov in 1941. If we assume that there is a universal law, independent of the Reynolds number, such that the energy $E(k)\mathrm{d}k$ depends only on the wavenumber, the dimensions

$$E(k) \sim L^3 T^{-2} \,, \quad \varepsilon \sim L^2 T^{-3} \,,$$

give the unique combination of ε and k

$$E(k) = C_\mathrm{K} \varepsilon^{2/3} k^{-5/3} \quad \text{(Kolmogorov's law)} \,,$$

which holds for

$$Re \gg 1 \,, \quad k_I \ll k \ll k_D \,.$$

von Kármán evaluated the constant C_K as 1.4.

A characteristic feature of three-dimensional turbulence is the local growth of vorticity by the stretching of vortex tubes in an inviscid fluid (Helmholtz's theorem, conservation of vorticity).

For large Reynolds numbers a physical mechanism (e.g. rotation) can make the turbulence two dimensional. On small scales the 2-D turbulence can turn into 3-D turbulence. This occurs once inertia dominates Coriolis forces. Defining the Rossby number Ro as the ratio of inertia to Coriolis force, the critical value of Ro is, experimentally,

$$Ro = \frac{|\boldsymbol{v} \cdot \nabla \boldsymbol{v}|}{2|\boldsymbol{\Omega} \times \boldsymbol{v}|} \simeq \frac{u_c}{2\Omega L} = 0.2 \,,$$

where u_c is a characteristic convection velocity and the transition occurs for $Ro > 0.20$.

The conservation of vorticity $\nabla \times \boldsymbol{v}$ along the motion for each fluid element implies the conservation of its mean square $\langle (\nabla \times \boldsymbol{v})^2 \rangle$ or *enstrophy*.

In 2-D turbulence, from the wavenumber k_I where energy is injected, there is a cascade of enstrophy to smaller scales with a power-law energy spectrum

$$E(k) = C'(k_I^2 \varepsilon)^{2/3} k^{-3},$$

and an inverse cascade of energy to larger scales with spectrum

$$E(k) = C'' \varepsilon^{2/3} k^{-5/3}$$

up to the scale $l_E(t)$ which grows indefinitely with time: at large scales there is no global stationary regime; in real situations the growth of larger structures is limited by the size of the system.

In the enstrophy cascade domain the Rossby number is constant; in the region of 3-D turbulence the Rossby number increases at smaller scales. 2-D turbulence in a system of finite size can never satisfy conservation of enstrophy across scales exactly. In laboratory experiments one finds a power law k^{-n}, with n between $5/3$ and 3. The Rossby number grows towards smaller scales as $k^{(3-n)/2}$, until it reaches a value $Ro > 0.20$ so that 2-D turbulence turns into 3-D turbulence.

In a stratified medium we have to respect a condition analogous to the Richardson–Townsend condition: the power available to produce 3-D turbulence must exceed the power dissipated in work against buoyancy forces. The decay of 2-D turbulence into 3-D can only occur if $H_P/r < O(1)$. This excludes the presence of this kind of turbulence immediately below the boundary of the convection zone in solar-type stars.

Diffusion. In 3-D turbulence the stochastic fluid motions can transport chemical elements by a random-walk process analogous to that of microscopic diffusion. The continuity equation for the concentration c of a chemical species is

$$\frac{\partial c}{\partial t} = \nabla \cdot (cv) = 0.$$

Writing $c = \bar{c} = \boldsymbol{l} \cdot \nabla c$ where $\boldsymbol{l}$ is a random displacement of the fluid element, we have

$$\frac{\partial c}{\partial t} + \nabla \cdot (\langle c \rangle + (\boldsymbol{l} \cdot \nabla c)v) = 0.$$

Taking $\langle v \rangle = 0$, and $\langle \boldsymbol{l} \cdot \boldsymbol{v} \rangle = -D_T$, we get a diffusion equation

$$\frac{\partial c}{\partial t} = \nabla \cdot (D_T \nabla c),$$

where we have introduced a turbulent diffusion coefficient D_T. Radial turbulent diffusion occurs whenever the concentration gradient is small enough not to inhibit the generalised instabilities driving the turbulence.

Observations demand the existence of a turbulent diffusion mechanism which can change the surface composition of stars or inhibit gravitational settling in Am stars: we discuss these problems in Sect. 5.5.4.

5.4.9 The Dynamo Effect

The presence of magnetic fields at the surfaces of stars requires explanation, and we need to understand their origin and role in stellar hydrodynamics. We have already seen (Sect. 5.4.3) that even a weak magnetic field may considerably alter the circulation of matter in a rotating star.

Magnetic Field Properties. Two properties are basic to all discussions of stellar magnetic fields.

(a) If ohmic dissipation can be neglected, *the magnetic field is frozen in to the plasma.* This follows easily on calculating the variation of flux Φ across a surface bounded by a curve C moving with the fluid:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \int_S \left[\frac{\partial \boldsymbol{B}}{\partial t} - \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) \right] \cdot \boldsymbol{n} \mathrm{d}S. \tag{5.178}$$

Using the induction equation

$$\frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times (\boldsymbol{v} \times \boldsymbol{B}) - \nabla \times (\eta \nabla \times \boldsymbol{B}), \tag{5.179}$$

where $\eta = 1/\mu_0 \sigma$ is the magnetic diffusivity, with σ the conductivity, we get

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \eta \int \Delta \boldsymbol{B} \cdot \mathrm{d}S. \tag{5.180}$$

If the conductivity σ is infinite, $\eta = 0$ and the flux Φ is constant across the surface bounded by C: the magnetic field lines move with the matter.

(b) If we neglect the motion of the fluid ($v = 0$) the magnetic field obeys a diffusion equation

$$\frac{\partial B}{\partial t} = \eta \Delta B, \tag{5.181}$$

and the characteristic time

$$t_B = (L^2/\eta) \tag{5.182}$$

is also the timescale for the disappearance of the field under ohmic dissipation. The electrical conductivity of stellar matter is very high. In electromagnetic units, a plasma of ionic charge Z has

$$\sigma = \frac{e^2 T^{3/2}}{15 Z m_e} \frac{1}{\ln \Lambda},$$ (5.183)

where $\Lambda^{1/3}$ is as usual the ratio of the Debye length to the average inter-particle distance. For a hydrogen plasma inside the Sun this gives a characteristic timescale of the order of 10^{10} y, or 300 y for a sunspot. In all cases, on large scales the main effect is that magnetic fields are dragged around by matter motions. This property, stated in this simplified way, does not explain the magnetic phenomena observed at the surface of the Sun and active stars, although the assumption of a frozen-in fossil field does explain the Ap stars (p = peculiar, slowly rotating stars with strong magnetic fields).

In the solar convection zone the turnover time l/u varies from 5 minutes near the surface to a fraction of a day near the base, and turbulence is mainly due to Rayleigh–Bénard instabilities. This causes field evolution on large scales (sunspots), which is a manifestation of the turbulent dynamo effect (Steenbeck et al. 1966).

The Dynamo. The decay of magnetic fields through ohmic dissipation had already suggested the idea of a dynamo well before turbulent dynamos were envisaged. Cowling (1934) showed that a stationary axisymmetric magnetic field could not be maintained by an axisymmetric velocity field. The velocity u and magnetic field B can be decomposed into *poloidal* components U_p, B_p (in the meridian planes) and *toroidal* (azimuthal) components $U_\Phi e_\Phi, B_\Phi e_\Phi$, where e_Φ is a unit vector in the Φ direction. By symmetry the poloidal field component must vanish on a circle about the axis. This can only happen if there is an azimuthal current on this circle and an azimuthal electric force. This is impossible, since if the field B were azimuthal the product $U \times B$ must be poloidal, in contradiction with the requirement of axisymmetry.

A turbulent fluid may be axisymmetric on average. This is only a statistical symmetry, leading to the idea that through an inverse cascade from small to large scales, turbulence could produce a large-scale field. We give here a brief summary of *mean-field electrodynamics* (MFE). We shall give several astrophysical applications.

The astrophysical problem is to produce a magnetic field in a rotating body through motions which are statistically stationary. The resulting field should be essentially symmetrical about the rotation axis, while exhibiting large local deviations from this symmetry; it may be stationary, or show cyclic or period variations.

MFE expresses the velocity u and magnetic field B as the sum of slowly varying average values $\langle u \rangle, \langle B \rangle$ and fluctuating components u', B'. The turbulent motions are assumed to have a correlation time τ and length λ small compared with the variation scales t_0, l_0 of $\langle u \rangle, \langle B \rangle$; τ is the average time after which u' is no longer correlated with its initial value, and λ is a length of order the average eddy size. Equation (5.179) is then separated into its mean and fluctuating parts:

$$\frac{\partial \langle B \rangle}{\partial t} = \nabla \times (E + \langle u \rangle \times \langle B \rangle) - \nabla \times (\eta \nabla \times \langle B \rangle), \tag{5.184}$$

$$\frac{\partial B'}{\partial t} = \nabla \times (\langle u \rangle \times B' + u' \times \langle B \rangle + G) - \nabla \times (\eta \nabla \times B'), \tag{5.185}$$

where

$$E = \langle u' \times B' \rangle \tag{5.186}$$

and

$$G = u' \times B' - \langle u' \times B' \rangle. \tag{5.187}$$

The electric field E is electromotor and regenerates the field through the "α effect".

In (5.184) the electric field E can be regarded as a mean electric field resulting from the interaction of the turbulent motion and the field. We can find it by solving (5.185) for B'. We neglect $\langle u \rangle$ in (5.185), as is reasonable for a turbulent medium. To first order we can neglect G, which is higher than second order. This is justified if B' is small compared with $\langle B \rangle$. This holds for a medium with rapid fluctuations, $\tau \ll \lambda/v$, $v^2 = \langle u'^2 \rangle$: in this case the dominant term on the right-hand side of (5.185) is $\nabla \times (u'^2 \times \langle B \rangle)$. This does not apply for the Sun, where $\tau \simeq \lambda/u$. Actually one still uses the first-order approximation

$$\frac{\partial B'}{\partial t} + \nabla \times (\eta \nabla \times B') = \nabla \times (u' \times \langle B \rangle). \tag{5.188}$$

We use this equation to get an idea of the *form* of E. Using the fact that the correlation time is τ we integrate (5.188) with $u' = $ constant over this interval, with $B'(t - \tau) = 0$, to find the part B'' of B' correlated with u':

$$B'' = \int_{t-\tau}^{t} \nabla \times (u' \times \langle B \rangle) \mathrm{d}t. \tag{5.189}$$

As the right-hand side of (5.189) is linear in B_j and $\partial B_j / \partial x_k$, we write the electric field components as

$$E_i = \alpha_{ij} \langle B_j \rangle + \beta_{ijk} \frac{\partial \langle B_j \rangle}{\partial x_k}, \tag{5.190}$$

where α_{ij}, β_{ijk} depend on the local structure of the fluid velocity field. If the velocity field is isotropic,

$$\alpha_{ij} = \alpha \delta_{ij}; \quad \beta_{ijk} = \beta \varepsilon_{ijk} \tag{5.191}$$

in the usual notation, and

$$E = \alpha \langle B \rangle - \beta \nabla \times \langle B \rangle. \tag{5.192}$$

If τ is small compared with the diffusion time of the field in an eddy, $\tau_d \equiv \lambda^2/\eta$ (or, amounting to the same thing, for a magnetic-eddy Reynolds number $Rm = v\lambda/\eta \gg 1$) we can neglect the diffusion term in (5.188) and get

$$\alpha = -\tfrac{1}{3}\tau \langle \boldsymbol{u}' \cdot (\nabla \times \boldsymbol{u}') \rangle, \qquad \beta = \tfrac{1}{3}\tau v^2. \tag{5.193}$$

The quantities α, β are very poorly known. We shall see below a way of estimating them. If τ/τ_d is not small, expressions like (5.193) result, with τ replaced by τ_c:

$$\frac{1}{\tau_c} = \frac{1}{\tau} + \frac{1}{\tau_d}. \tag{5.194}$$

The quantity $\boldsymbol{u}' \cdot (\nabla \times \boldsymbol{u}')$ is called the *helicity* of the velocity field. A non-zero value of the helicity means that the vorticity $\nabla \times \boldsymbol{u}'$ is predominantly in either the direct or retrograde sense around the velocity. In general $\alpha_i \neq 0$ means that the velocity field is not plane-symmetric.

We thus get the MFE dynamo equation

$$\frac{\partial \langle \boldsymbol{B} \rangle}{\partial t} = \nabla \times (\alpha \langle \boldsymbol{B} \rangle + \langle \boldsymbol{u} \rangle \times \langle \boldsymbol{B} \rangle) - \nabla \times [(\eta + \beta)\nabla \times \langle \boldsymbol{B} \rangle]. \tag{5.195}$$

The term $\alpha \langle \boldsymbol{B} \rangle$, representing an electric field parallel to $\langle \boldsymbol{B} \rangle$, avoids the obstacle presented by Cowling's theorem for an axisymmetric dynamo.

The quantity β is a turbulent diffusivity analogous to ohmic diffusivity. It shows that magnetic fields are transported by turbulence, but does not suppress small inhomogeneities like ohmic diffusivity. Equation (5.195) gives a sufficient condition for the dynamo effect, i.e. that $\boldsymbol{B}$ should grow:

$$\frac{|\alpha| L}{\eta + \beta} > C_D, \tag{5.196}$$

where C_D is a number depending on the shape of the fluid region and L a characteristic length of the system.

The Solar Dynamo. This is a turbulent α dynamo in the convection zone, a turbulent system with non-zero helicity, typical for the dynamo effect caused by the α term. We use in (5.195) cylindrical coordinates ϖ, Φ, z with $u_\Phi = \varpi \Omega$, where $\Omega = \Omega(\varpi, z)$ in general, and with $\boldsymbol{B}_p = \nabla \times (A_\Phi \boldsymbol{e}_\Phi)$, where $A_\Phi \boldsymbol{e}_\Phi$ is a toroidal vector potential. Setting $\eta_T = \eta + \beta$ we get

$$\frac{\partial B}{\partial t} + B(\nabla \cdot \boldsymbol{u}_p) + \varpi(\boldsymbol{u}_p \cdot \nabla)(\varpi^{-1} B)$$
$$= \varpi(\boldsymbol{B}_p \cdot \nabla)\Omega + (\nabla \times (\alpha \boldsymbol{B}_p))_\Phi + \varpi^{-1}\nabla \eta_T \cdot \nabla \varpi B$$
$$+ \eta_T(\nabla^2 - \varpi^{-2})B, \tag{5.197}$$

$$\frac{\partial A}{\partial t} + \varpi^{-1}\boldsymbol{u}_p \cdot \nabla(\varpi A) = \alpha B + \eta_T(\nabla^2 - \varpi^{-2})A \tag{5.198}$$

In (5.197) the main term generating the magnetic field comes from $\nabla \Omega$ (non-uniform rotation). The α term can be compared with the term in $\nabla \Omega$. The ratio of these two terms is of the order of $v\tau/L$ and thus small. If we neglect the α term in (5.197) this is an $\alpha\omega$ *dynamo*. This is the most studied case, and holds for the Sun.

Moffat (1984) shows how to apply this model to the solar dynamo in a simple way. First, the assumption of a correlation time τ gives an estimate of the differential rotation. We consider the velocity components u, v created from 0 to t by a vertical component w. The Coriolis force implies that

$$\frac{\partial \boldsymbol{u}}{\partial t} = -2\boldsymbol{\Omega} \times \boldsymbol{u} \tag{5.199}$$

with the local Cartesian coordinates for Ω

$$\boldsymbol{\Omega} = (-\Omega \sin\theta, 0, \Omega \cos\theta). \tag{5.200}$$

With $u = v = 0$ at time $t = 0$, we have

$$v = -2w \sin\theta \cdot \Omega t + \mathrm{O}(t^3), \tag{5.201}$$

$$u = -2w \cos\theta \sin\theta (\Omega t)^2 + \mathrm{O}(t^4), \tag{5.202}$$

so that the Reynolds stress tensor has the component

$$\langle uv \rangle = 4(\Omega t)^3 \langle w^2 \rangle \cos\theta \sin^2\theta + \mathrm{O}(t^3). \tag{5.203}$$

This suggests a reasonable approximation for a statistically stationary state

$$\langle uv \rangle = 4(\Omega\tau)^3 \langle w^2 \rangle \cos\theta \sin^2\theta.$$

For the supergranulation, $\tau \simeq 3 \times 10^5$ s, $\Omega\tau \simeq 0.2$. This creates θ-dependent differential rotation Ω_1,

$$\eta_T \frac{\partial \Omega_1}{\partial \theta} = \langle uv \rangle, \tag{5.204}$$

where η_T is the turbulent viscosity, here attributed to the inertial scales (granulation), with $\eta_T \simeq 10^{12}$ cm^2 s^{-1}.

Integrating (5.204) gives

$$\Omega_1(r, \theta) = \frac{4\langle w^2 \rangle}{3\eta_T} (\Omega\tau)^3 \left(\sin^3\theta - \frac{4}{3\pi} \right), \tag{5.205}$$

where the integration constant is chosen to make $\langle \Omega_1 \rangle = 0$. This expression shows that the rotation velocity increases towards the equator, as observed, and suggests that this representation is physically reasonable. The difference between pole and equator is

$$\Omega_1 \left(R, \frac{\pi}{2} \right) - \Omega_1(R, 0) = \frac{4}{3\eta_T} \langle w^2 \rangle (\Omega\tau)^3. \tag{5.206}$$

To order of magnitude, with $\langle w^2 \rangle = 10^8, \Omega\tau = 0.2$ we have

$$\Delta\Omega \simeq 10^{-6}\,\mathrm{s}^{-1} \qquad (5.207)$$

to be compared with the observed value $\Delta\Omega = 7.9 \times 10^{-7}\mathrm{s}^{-1}$.

From this we can evaluate α. We consider a fluid element in vertical motion. In this motion the vertical component of $(\boldsymbol{\omega} + 2\boldsymbol{\Omega})/\rho$, where $\boldsymbol{\omega}$ is the vorticity, is conserved. Over a small time interval t we find the z component

$$\omega_3 = 2\Omega\cos\theta \cdot wt\frac{\mathrm{d}}{\mathrm{d}z}\log\rho$$

giving the helicity H,

$$H = \langle \boldsymbol{u} \cdot \boldsymbol{\omega} \rangle \simeq \langle w\omega_3 \rangle \simeq -(\Omega\tau)\langle w^2 \rangle\frac{\cos\theta}{H_\rho}\,, \qquad (5.208)$$

where H_ρ is the density scale height.

The α parameter is then evaluated as

$$\alpha \simeq -\frac{1}{3}H\tau \simeq \frac{1}{3}\Omega\tau^2\cos\theta\frac{\langle w^2 \rangle}{H_\rho}\,. \qquad (5.209)$$

The sufficient condition (5.196) for the $\alpha\omega$ dynamo effect is equivalent to

$$L\frac{\Omega\tau\cos\theta}{H_\rho} > 1\,, \qquad (5.210)$$

where, introducing the Rossby number $Ro = \langle v/2\Omega l \rangle = \langle 1/2\Omega\tau \rangle$,

$$\frac{L}{H_\rho} > 2Ro\,,$$

expressing the compatibility between 3-D turbulence on small scales and the dynamo effect on large scales.

Reconnection of magnetic fieldlines is caused by ohmic dissipation. It occurs continuously because of the stretching of fieldlines by turbulence. A flux tube is stretched and its length grows exponentially in time:

$$l \sim \exp\left(\frac{u_0 t}{l_0},\right)$$

where u_0 and l_0 are the characteristic velocity and lengthscale of the turbulence. In first approximation the cross-section of the flux tube decreases exponentially (as the matter is frozen to the fieldlines, the volume is conserved and ls^2 is constant, with s^2 the cross-section). Ohmic dissipation occurs when the timescale for dissipation is comparable with the growth time
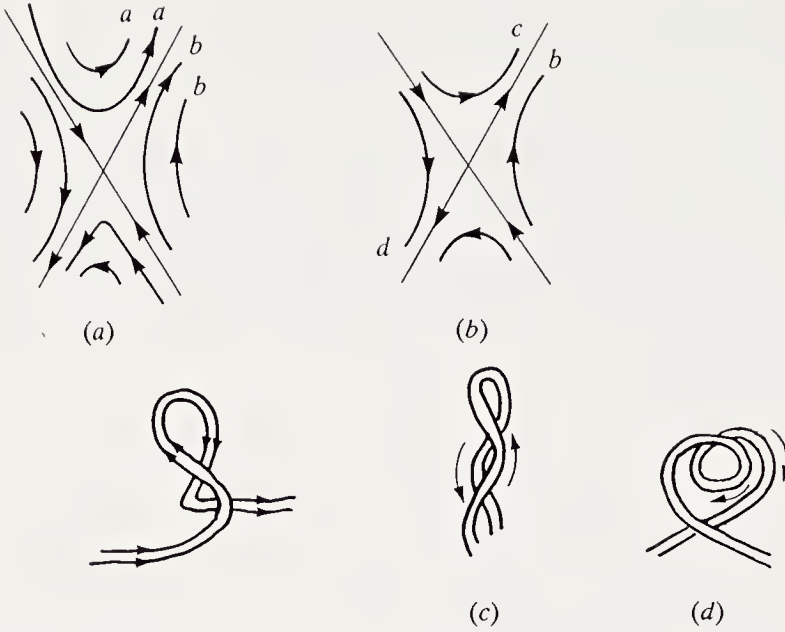
Fig. 5.6. Dissipative effects in a plasma. In (a) and (b) the evolution of the magnetic fieldlines is due to dissipative resistivity. The Coriolis force creates magnetic loops which can lead to field annihilation (c) or reinforcement (d)

$$\frac{s^2}{\eta} \simeq \frac{l_0}{u_0}, \qquad \text{or}$$

$$\frac{l_0 u_0}{\eta} \simeq \exp\left(\frac{u_0 t}{l_0}\right), \qquad t = \frac{l_0}{u_0} \log\left(\frac{l_0 u_0}{\eta}\right).$$

After a time of the order of ten times the characteristic turbulent timescale (u_0/l_0) the stretching of the flux tube is such that reconnection by ohmic dissipation takes place on the timescale u_0/l_0. Several basic cases occur (Fig. 5.6): dissipation near a neutral point (a, b), dissipation and field annihilation (c), dissipation and reinforcement of the field (d). However they occur, a turbulent dynamo is the likely explanation for variations of the solar magnetic field. The essential property is the propagation of a dynamo wave, as envisaged by Parker (1955). As an example we consider (5.195) in Cartesian coordinates. With $B_p = \nabla \times (Ae_y)$ we have two equations analogous to (5.197) and (5.198)

$$\frac{\partial A}{\partial t} + u_p \cdot \nabla A = \alpha B + \eta_T \nabla^2 A, \tag{5.211}$$

$$\frac{\partial B}{\partial t} + u_p \cdot \nabla B = B_p \cdot \nabla u + \eta_T \nabla^2 B. \tag{5.212}$$

We seek a solution of the form

$$\exp(pt + ik \cdot x)$$

and obtain the dispersion relation

$$p = s - \eta_T k^2 - i u_p \cdot k \tag{5.213}$$

with

$$s^2 = -i\alpha(k \times \nabla u)_y = 2i\gamma \tag{5.214}$$

the mean velocity and magnetic field being $u_y + u_p$ and $Be_y + B_p$ respectively. In the case $\gamma < 0, u_p = 0$,

$$p = -\eta_T k^2 \pm |\gamma|^{1/2}(1 - i) \tag{5.215}$$

we have a growing field, propagating in the direction of the wavevectors k. The amplitude increases in time if

$$-\frac{1}{2}\alpha k \frac{\partial u}{\partial z} \geq \eta_T^2 k^4 . \tag{5.216}$$

We define the dimensionless numbers $D_\alpha = \alpha R/\eta_T$, related to the production of poloidal field, and $D_\omega = \nabla\Omega R^3/\eta_T$, related to the production of toroidal field. We have

$$p = -\eta_T k^2 + \eta_T k^2 (D_\alpha D_\omega / 2k^3 R^3)^{1/2}(1 - i). \tag{5.217}$$

The dimensionless number $D_\alpha D_\omega = Dy$ is called the *dynamo number*. The dynamo effect occurs if

$$Dy > D_{\text{crit}} \tag{5.218}$$

with

$$D_{\text{crit}} = k_{\text{crit}}^3 R^3 ,$$

where k_{crit} is given by (5.216), or

$$k_{\text{crit}}^3 = R\nabla_r(\Omega\alpha/2\eta_T^2). \tag{5.216'}$$

For $k = k_{\text{crit}}$ the *dynamo wave* does not grow in time. Its period P is given by

$$(2\pi/P) = \mathrm{Im}(p) = \left(-\tfrac{1}{2}\alpha R k \nabla\Omega\right)^{1/2} . \tag{5.219}$$

Writing $\nabla\Omega = (\Delta\Omega/\Omega)(\Omega/R)$ we find, using (5.55) for w,

$$(2\pi/P) \simeq 2^{-5/6} \left[\frac{5}{27} \frac{H_P \Omega^4}{(F/10\rho)^{1/3}} \left(\frac{\Delta\Omega}{\Omega}\right)^2 \right]^{1/3} . \tag{5.220}$$

The period thus found is of the order of a year at the base of the convection zone and increases slowly, roughly as $T^{-1/6}$, towards the surface, reach-

237

ing about two years there. This disagreement with observation is discussed below.

The MFE theory successfully explains the essential characteristics of the solar cycle:

— propagation of a dynamo wave towards the equator, explaining the Maunder butterfly diagram (Fig. 7.2);

— polarity reversal every eleven years;

— polarity rules for sunspot pairs in the two hemispheres (NS in the north, SN in the south, with exchange of polarities every eleven years);

— phase difference between the poloidal and toroidal fields ($\Phi = \pi/4$).

A number of problems remain:

— The period predicted for the solar cycle is too short. This could result from an overestimate of α. Assuming with Zeldovitch et al. (1983) that the helicity is the time average

$$\alpha = \int \boldsymbol{u} \cdot (\nabla \times \boldsymbol{u}) \mathrm{d}t \,,$$

we have to take this average over an eddy turnover time τ as long as $\Omega\tau < 1$; in contrast in the lower convection zone we have to take the average over a period $\sim 1/\Omega$. In the upper part of the convection zone

$$\Omega\tau < 1, \quad \alpha \sim l\Omega \,,$$

and in the lower part

$$\Omega\tau > 1, \quad \alpha \sim w \,.$$

At the base of the convection zone we deduce

$$(2\pi/P) \simeq 2^{-5/6} \left[\frac{5}{27} \frac{F}{10\rho} \frac{1}{H_P^3} \left(\frac{\Delta\Omega}{\Omega} \right)^2 \right]^{1/3} \tag{5.221}$$

with $\Delta\Omega/\Omega = 1, P = 460$ days, with $\Delta\Omega/\Omega = 0.01, P = 27$ years.

— The depth at which the solar dynamo operates is not clearly understood. Observations suggest magnetic field production at the base of the convection zone, in particular the production of the toroidal component.

— The stability of the flux tubes under buoyancy forces requires that the rise time should be comparable to the timescale for field amplification and that for field diffusion, with both comparable to the period of activity.

238

5.5 Observations and Interpretation

The data pertaining to the hydrodynamics of the stellar interior are simultaneously meagre and fundamental. Only indirect effects of fluid motions are observed, and not these motions themselves.

5.5.1 The Solar Granulation

White-light observation of the Sun's surface shows a granular structure (see Sect. 2.4.1). The first observations, made by Janssen at the Observatoire de Meudon at the end of the 19th century, revealed sizes from 400 to 1000 km. We have long assumed that this pattern is due to rising columns of hot gas, forming bright granules surrounded by descending areas of cool gas constituting the darker intergranular regions.

In fact we can observationally distinguish several scales of convective motions: *global convection* with a scale of 100 000 km and extending over the whole convection zone; *supergranules*, of characteristic scale 30 000 km; *mesogranules*, of intermediate size; and *granules*, with scale 1000 km. Magnetic field measurements and the study of solar activity show the existence of similar scales for magnetic activity: large scale (the 22 year cycle in particular); intermediate scale, in *active regions*, with size comparable to the supergranulation; and the *filigree* at the scale of the granulation. The nature of interaction between hydrodynamical motions and magnetic fields is not well understood. On large scales the dynamo appears established in principle, but it seems difficult to attain agreement between numerical simulations and observations. The circulation on the supergranulation and granulation scales expels the field from the centre of the convection cell towards the walls: this is as observed. The fieldlines leaving the Sun sketch out the supergranulation pattern.

Models of radiative transfer in the granules give a good representation of the flux at the Sun's surface. The combined effects of entropy advection and radiative heating explain the appearance of the granulation satisfactorily.

Fourier analysis of the velocity field at the Sun's surface shows that many modes of high horizontal wavenumber appear. The appearance of the granulation is due to the combination of oscillatory motions of the surface and random motions of the highly unstable fluid in the convection zone. It is tempting to try to find the properties of the turbulence by subtracting the basic periodic motions.

A simplified method assumes that the basic periodic motion occurs near a period of 5 min. Adding Doppler images taken at 2.5 min intervals implies opposite phases, and to a first approximation we are left with the random motion. From the combined image we can calculate the two-dimensional correlation function and, by Fourier transforming, find the power spectrum $F(k)dk$ of the turbulence (Fig. 5.7). We thus find that the granulation region obeys a $k^{-5/3}$ law (Fig. 5.8) and corresponds to the inertial domain of the
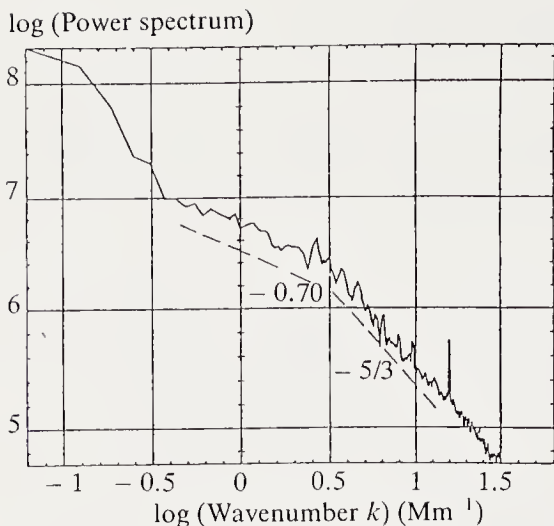
Fig. 5.7. Power spectrum of solar photospheric motions obtained by superposing two Doppler images separated by 2.5 minutes. (From J.P. Zahn, in *Solar and Stellar Physics*, ed. by E.H. Schröter and M. Schüssler, p. 64 (1987))
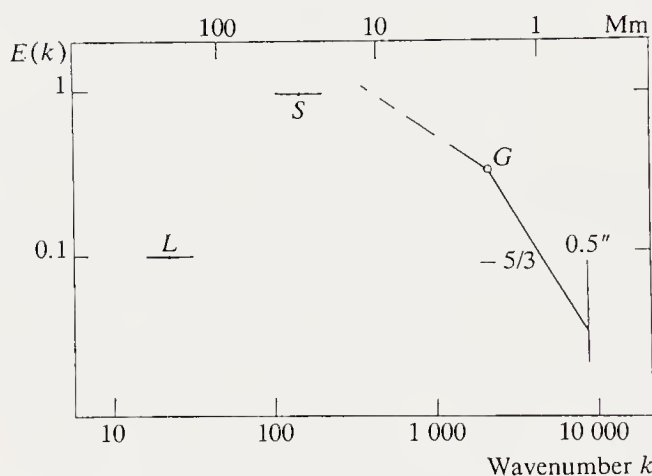
Fig. 5.8. The power spectrum of photospheric turbulence, from 200 Mm (the size of the rolls causing solar activity) to the granulation (450 km). L denotes large-scale motions, S the motions of the supergranules, G the motions of the granules. (From J.P. Zahn, in *Solar and Stellar Physics*, ed by E.H. Schröter and M. Schüssler, p. 68 (1987))

turbulence. We can thus imagine that there is a cascade towards smaller scales from the supergranulation scale. The motions of the granules under these conditions are caused by a cascade from large scales, and not by Rayleigh–Bénard instabilities.

5.5.2 Solar and Stellar Activity

A second tracer of MHD effects in the convection zone is solar and stellar activity, and in particular the associated X-ray emission (see Chap. 7). We note here the importance of the Rossby number, which basically decides whether a dynamo can work. Directly connected with the dynamo effect is the problem of stellar rotation and angular momentum loss (see below).

Stellar Rotation. The classical problem of the Sun's rotation first arose when Babinet (1861) and Fouché (1884) tried to apply angular momentum conservation to Laplace's model of the formation of the Solar System. In present terms, it is easy to see that the contraction of an interstellar cloud towards the main sequence results in a star rotating near breakup. At the equator, the relation

$$V^2 = \frac{GM}{R}$$

gives the rotation speed. For the Sun this corresponds to an equatorial velocity of about 400 km s^{-1}, about 200 times the present value.

The problem of supplying an efficient process of angular-momentum exchange was the main reason for abandoning of the Kant–Laplace cosmogony. One had to wait almost a century until, mainly through the work of Struve and Elvey (1931), the prime observational discovery emerged that the rotation velocity of stars depends on their spectral type. Summarising, main-sequence stars later than F2 spin slowly, and those earlier than F2 spin rapidly. Typical equatorial velocities are respectively 10 km s^{-1} and 50 km s^{-1} and above (Fig. 5.9).

A second observational fact, discovered by Kraft (1967) from studies of the rotation of stars belonging to clusters of known age, was that the rotation velocity of a star depends on its age t. Significant angular-momentum loss could explain the spindown of stars; if the angular momentum had to be carried off by matter leaving the stellar surface, the observed behaviour of Ω would have implied very high mass-loss rates, which were not observed
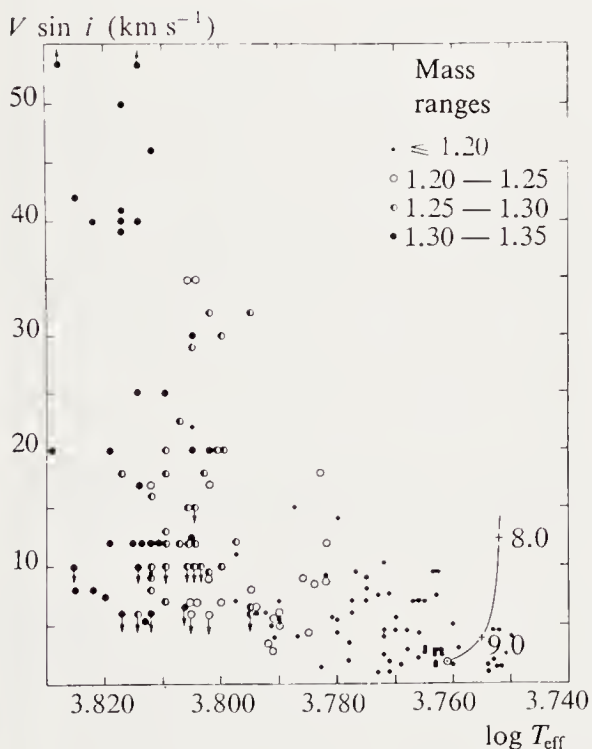
Fig. 5.9. Equatorial velocity as a function of effective temperature. We have plotted the Sun's trajectory assuming the rotation period proportional to $t^{1/2}$. We note the rapid rise of $V \sin i$ towards $\log T_{\text{eff}} = 3.780$ ($T_{\text{eff}} = 6100$ K). (From D. Soderblom, *Astrophys. J. Suppl. Ser*, **53**, 1, 1983. Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1983 The American Astronomical Society)

for solar-type stars. Angular-momentum loss at a distance d much larger than the stellar radius is required to solve the problem.

This can happen if the matter leaving the star is forced to corotate with the stellar magnetic field (Schatzman 1959). The difference between stars later and earlier than F2 lies in the nature of the convection zone and the dynamo effect. A simple idea, due to Parker, is to assume that there is a dynamo effect if the dynamo number Dy is large enough:

$$Dy = \frac{\alpha \Delta \Omega L^4}{\eta_T^2} > 1 \,,$$

where $\Delta \Omega$ is the differential rotation, L the scale of the generating system (the depth of the generating region), and η_T the magnetic diffusivity. Evaluating α and η_T,

$$\eta_T = ul\,, \qquad \alpha = \frac{\Omega l^2}{H_\rho}\,,$$

we find the dynamo number

$$Dy = Ro^{-2} = (u_c/l\Omega)^{-2} \,.$$

where u_c is a characteristic velocity of the convection. The dynamo effect occurs if the Rossby number is small enough, i.e. if the Coriolis force is stronger than inertial turbulence. This assumes that the differential rotation is determined by turbulent viscosity:

$$\Delta \Omega = \frac{l^2 \Omega}{H_\rho^2} \,.$$

For increasing masses, the rapid decrease in the depth of the convection zone near $\log(M/M_\odot) \simeq 0.05 - 0.1$ increases the Rossby number rapidly (Fig. 5.10), making dynamos ineffective at intermediate masses, whatever the angular velocity.

To estimate the distance where angular momentum is lost we follow Cowling in assuming that this happens once the Alfvén velocity given by

$$u_A^2 = \frac{B_A^2}{4\pi\rho} \,.$$

becomes equal to the stellar wind velocity. In the stellar wind the magnetic field varies as r^{-2} because of the freezing-in of the field in the plasma. If B_0 is the field at the surface of the star,

$$B_0 R^2 = B_A r_A^2 \,.$$
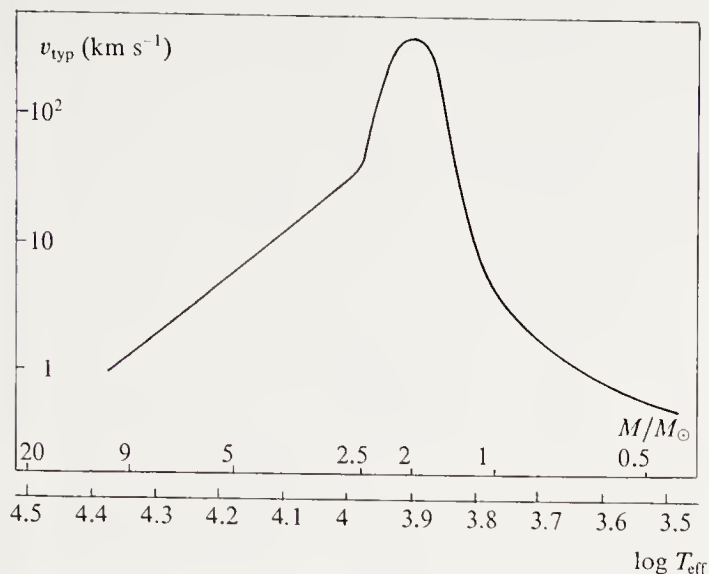
Fig. 5.10. The Rossby number as a function of mass for main-sequence stars. The typical velocity $v_{\text{typ}} = v_m R/2H_c$, where H_c is the thickness of the convection zone, v_m the maximum velocity in the zone and R the stellar radius. The effective Rossby number is $Ro = v_{\text{typ}}/v_{\text{eq}}$ where v_{eq} is the equatorial velocity. For $\Omega = $ const, v_{typ} varies roughly as the Rossby number Ro. (We have assumed Ω corresponding to $v_{\text{eq}} = 1$ km s^{-1})

As the angular-momentum loss occurs at large distances from the star $(r_A/R \gg 1)$ we can take u_A as the escape velocity,

$$u_A \simeq (2GM/R)^{1/2} .$$

We thus obtain an angular-momentum loss rate

$$\frac{\mathrm{d}}{\mathrm{d}t} I\Omega = -\frac{2}{3}(B_0 R^2)^2 \left(\frac{GM}{R}\right)^{-1/2} \Omega .$$

We now need the dependence of B_0 on Ω. Durney and Latour (1978) take the observational data as a starting point, assuming with Skumanich (1972) that the surface field B_0 decreases as Ω.

This property should no doubt be associated with non-linear effects of the turbulent dynamo, for which there is no unique theory. If we assume that magnetic field growth is limited by field expulsion through buoyancy, one of the solutions actually gives $B \sim \Omega$. We write with Durney and Latour (1978)

$$B_0 = \frac{B_n}{Ro} = B_n \left(\frac{l\Omega}{u_c}\right) ,$$

where B_n is deduced from observation. We obtain the angular momentum as a function of time:

$$\Omega = \Omega_0 \left(1 + \frac{t}{t_0}\right)^{-1/2} \qquad \text{with}$$

243
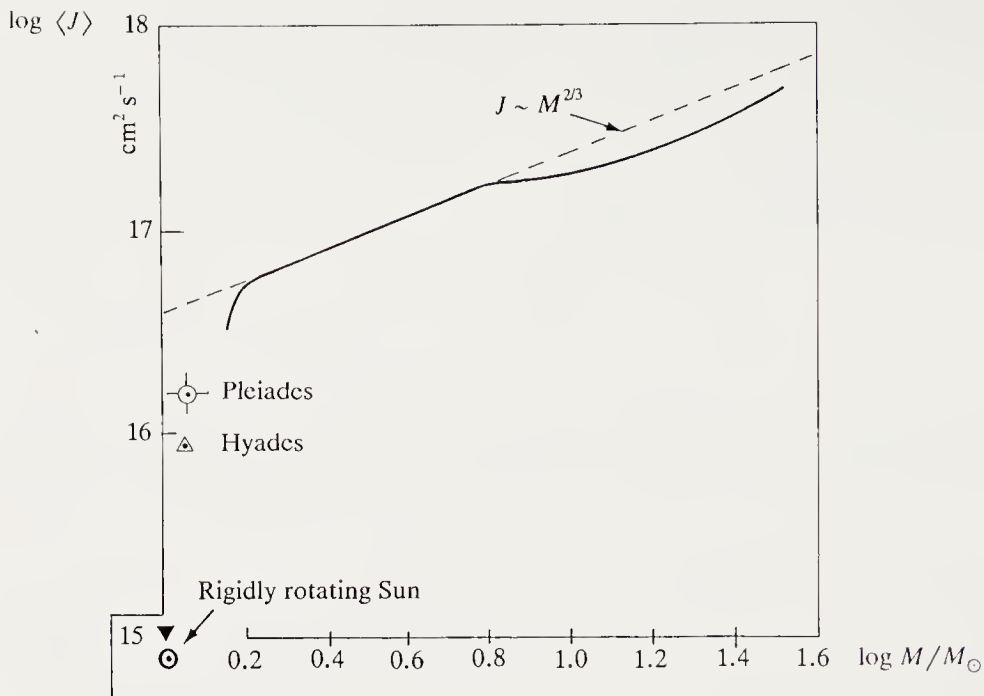
Fig. 5.11. Specific angular momentum (per unit mass) for main-sequence stars (assuming rigid rotation). (From R.P. Kraft, *Astrophys. J.* **150**, 551, 1967. Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1967 The American Astronomical Society)

$$t_0^{-1} = \frac{4}{3}\left(\frac{R}{2GM}\right)^{1/2} B_n^2 \left(\frac{l}{u_c}\right)^2 \frac{\Omega_0^2}{KMR^2}\,,$$

where u_c is a characteristic convection velocity, Ω_0 is the initial angular velocity and K is the star's moment of inertia. To have an idea of the order of magnitude we take a mass–radius relation $R \sim M^{3/4}$, t as the nuclear timescale, with $t \sim M^{-3}$, and adjust to the Sun with an angular-momentum-loss timescale of 10^{10} years. We then have

$$(t_{\text{Nucl}}/t_0) = 71400 \left(\frac{\Omega_{0\odot}}{\Omega_{0\odot\,\text{max}}}\right)^2 \left(\frac{l/u_c}{(l/u_c)_\odot}\right)^2 \left(\frac{B_n}{B_{n\odot}}\right)^2 \left(\frac{M}{M_\odot}\right)^{-31/8}.$$

Figure 5.11 gives the specific angular momentum $KR^2\Omega$ as a function of mass. The main effect is caused by the shrinking of the hydrogen convection zone as the mass increases. This causes a very rapid increase in the specific angular momentum near $M/M_\odot = 1.12$ or $\log(M/M_\odot) = 0.05$. In fact, observations indicate that the transition from slow to fast rotation occurs for a slightly higher mass. This difference from the model is probably mainly caused by the fact that the star has lost some angular momentum before reaching the main sequence. Analysis of data for cluster stars for which the age is known actually shows that the equatorial velocity for zero-age stars on the main sequence is clearly smaller than the maximum equatorial velocity.

244
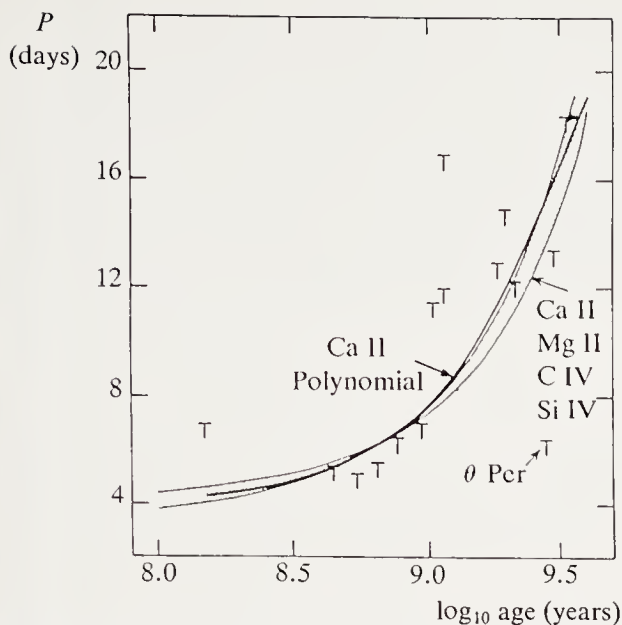
Fig. 5.12. The period as a function of age for stars of 1.1 $M_\odot$. The lithium abundance has been used to measure the age of field stars. A good representation is $P = 3.5(1 + t/0.52)^{0.75}$, where t is in billions of years and P in days. The symbols T represent upper limits for individual stars in Soderblom's list. (From Th. Simon et al., *Astrophys. J.* **293**, 551, 1985. Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1985 The American Astronomical Society)

According to Simon et al. (1985) the $(1 + t/t_0)^{-1/2}$ relation is not exactly verified, and in any case indicates that $\Omega_0 \ll \Omega_{0\ \max}$ (Fig. 5.12).

However, the situation is not as simple as might appear here. The presence in very young clusters such as α Per (50 million years) of both rapidly and slowly rotating stars shows that there is no unique initial distribution of angular momentum.

We have given here only a sketch of a rapidly developing subject.

5.5.3 Abundance of Trace Elements: Gravitational and Radiative Separation

The existence of stars with numerous abundance anomalies (Am and Ap stars) has long been a puzzle which it has been difficult to piece together. As an example we show here the element abundances in two Am and Ap stars (Fig. 5.13).

The main idea is to take account, in a stable atmosphere, of the transport processes associated with gravitation and radiation pressure (Michaud 1970). These two effects can produce enormous abundance anomalies both in excess and in deficit. Two phenomena limit the size of the anomalies: the presence of a weak stellar wind, with a very long timescale $(M/\dot{M})^{-1} \simeq 10^{14}$ y (for the Sun $(M/\dot{M})_\odot \simeq 10^{13}$ y); and the turbulent diffusion caused by rotation.
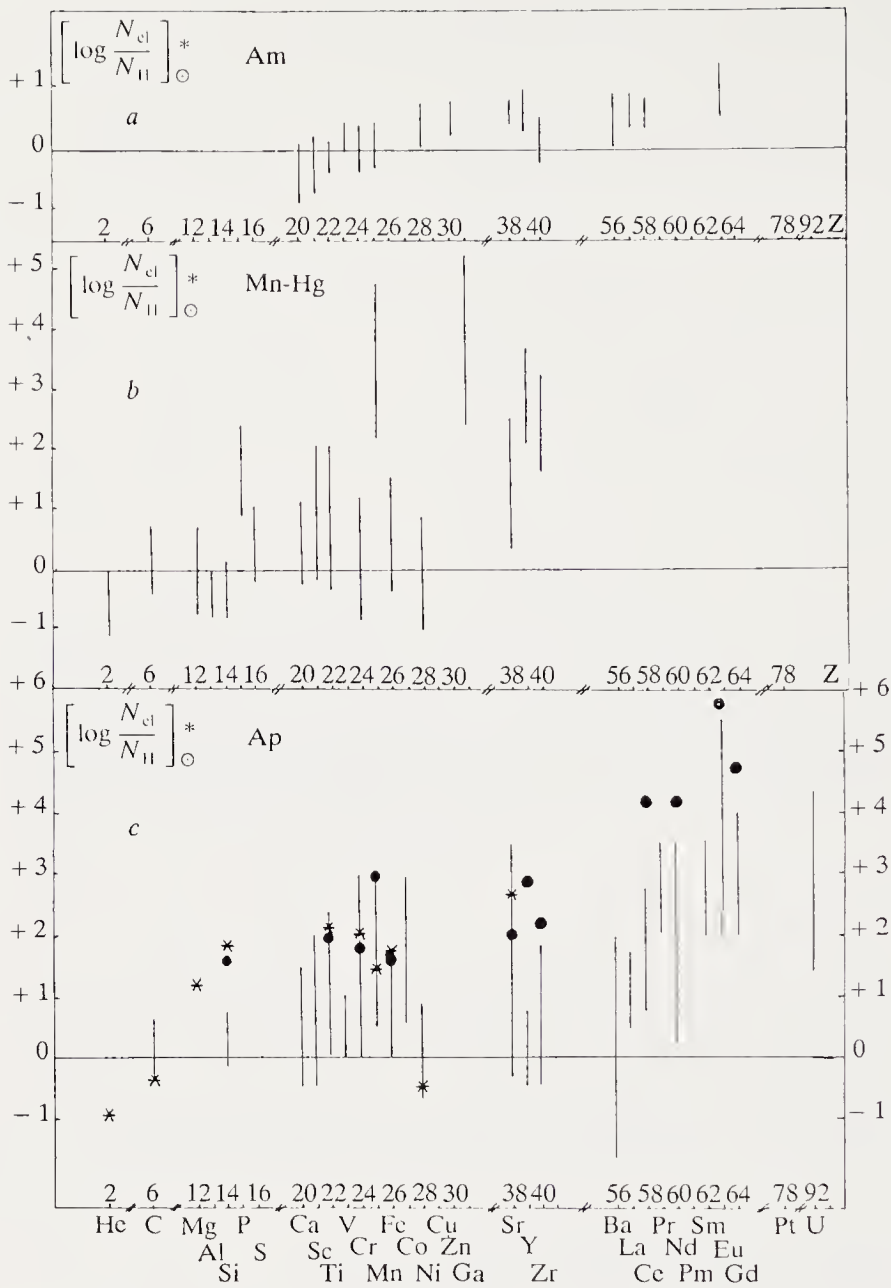
Fig. 5.13. Element abundances in an Am and an Ap star

The dependence on mass introduces on one hand the depth of the convection zone which governs the speed of the diffusion process (see below), on the other hand the turbulent mixing processes, which become important for rapid rotation. For stellar masses ranging from about 1 to $3\,M_\odot$ there is a narrow band where conditions permit abundance anomalies.

The equation governing the evolution of the surface abundance gives the velocity of a trace element (i.e. one with a very small abundance) under the combined effects of (1) microscopic diffusion down the concentration gradient, (2) settling under gravity, (3) driving by the temperature gradient, (4) radiation pressure, and possibly (5) turbulent diffusion.

246

If D_i is the microscopic diffusion coefficient for a trace element with concentration c_i, the velocity contribution due to microscopic diffusion is $-D_i \nabla \log c_i$. The gravitational contribution is easy to understand, and has two components, the weight $-gAm_H$ of atom of mass A and the electric force caused by the very small charge separation of an equilibrium plasma. In a completely ionised hydrogen and helium mixture with mass concentrations X and $1 - X$ (we neglect the heavy elements), the electric field is given by the equilibrium condition

$$eE = \frac{g(m_i - m_e)}{Z_i + 1},$$

which can be generalised for a mixture (neglecting m_e):

$$eE\left(\frac{(Z_1 + 1)X}{A_1} + \frac{(Z_2 + 1)(1 - X)}{A_2}\right) = gm_H = eE\left\langle\frac{Z + 1}{A}\right\rangle.$$

The corresponding speed is due to the friction on the ion under consideration of the ions and electrons of the plasma; the contribution is $D_i k_\rho \nabla \log \rho$, where

$$k_\rho = \left(\left\langle\frac{Z + 1}{A}\right\rangle 2A_i - Z_i - 1\right).$$

The driving by the temperature gradient comes from the fact that the cross-section depends on velocity, and thus on temperature. The resulting contribution is $D_i k_T \nabla \log T$, where

$$k_T = 2.65Z^2 + 0.805(Z^2 - Z).$$

Radiation pressure gives a speed $-D_i(m_i F/kT)$, where F is the acceleration due to radiation pressure. To find F we need the opacity of the atomic species i, Z-times ionised, and also the solution of the transfer equation. In the stellar interior we can use the quasi-isotropic approximation and write

$$F_\nu = -\frac{4}{3}\frac{1}{(\kappa_{\text{line}} + \kappa_{\text{cont}})\rho}\frac{\partial B_\nu}{\partial T}\frac{dT}{dz}.$$

The acceleration by radiation is then

$$F = \int \frac{\kappa_\nu F_\nu}{c}d\nu$$

$$= -\frac{4}{3}\int \frac{1}{c}\frac{\kappa_{\text{line}}}{(\kappa_{\text{cont}} + \kappa_{\text{line}})\rho}\frac{\partial B_\nu}{\partial T}\frac{\partial T}{\partial z}d\nu,$$

where the integral is over the whole spectrum of the atomic i, Z-times ionised. If $\kappa_{\text{line}} \gg \kappa_\nu$, the line is saturated and the acceration F decreases

as ρ^{-1} (Doppler profile) or $\rho^{-1/2}$ (damping). If there is turbulence we add a term $D_T \nabla \log c$, and obtain finally

$$v = -D_i \left(\nabla \log c - k_\rho \nabla \log \rho - k_T \nabla \log T + \frac{m_i F}{kT} \right) - D_T \nabla \log c.$$

The problem is then to solve the equation

$$\frac{\partial c}{\partial t} + \nabla \cdot (vc) = 0.$$

The diffusion coefficient D_i is to a first approximation

$$D_i = \tfrac{1}{3} l v,$$

where l is the mean free path of the trace atoms. Scattering through $\pi/2$ implies a cross-section of order

$$\sigma_i = \pi \left(\frac{Z_i Z_1 e^2}{kT} \right)^2.$$

Plasma effects give a correction term (Spitzer 1962), and we finally get for a binary mixture of ionised elements

$$D_i = \frac{3}{8\pi} \left\{ \frac{kT(m_i + m_H)}{2\pi m_i m_H} \right\}^{1/2} \left(\frac{2kT}{Z_i Z_1} \right)^2 \frac{1}{A_1(2)},$$

where $A_1(2) = \ln(1 + x^2), x = 4\lambda_D kT/Z_1 Z_2 e^2$, where λ_D is the Debye length. The full calculation of the diffusion coefficient also takes account of collisions with neutrals.

We note particularly the $T^{5/2}/\rho$ dependence of D_i. In an interior radiative zone, where $\rho \sim T^n$ with n of order 3 to 4, the microscopic diffusion coefficient decreases rapidly towards the interior, while the gradients $\nabla \log \rho$ and $\nabla \log T$ also decrease as $1/T$; thus the entrainments by gravity, temperature gradient and radiation pressure all decrease. ∇v again introduces the scaleheight, so that the characteristic time increases as $\rho/T^{1/2}$. We thus see that *in deep layers, the concentration does not vary in time.*

This does not hold for layers close to the surface, where elemental separation can be very efficient. In general, separation of elements does occur in radiatively stable regions.

A range of anomalous compositions exists. Figure 5.14 shows them on an HR diagram. The broken curves give the limits of the region where diffusion can produce observable effects. This includes the ZAMS, ZAHB, and white dwarfs of different spectral types in a $0.6\,M_\odot$ cooling sequence. The subdwarfs SdO and SdB are shown as a continuation of the horizontal branch, following the suggestion of Greenstein and Sargent (1974).
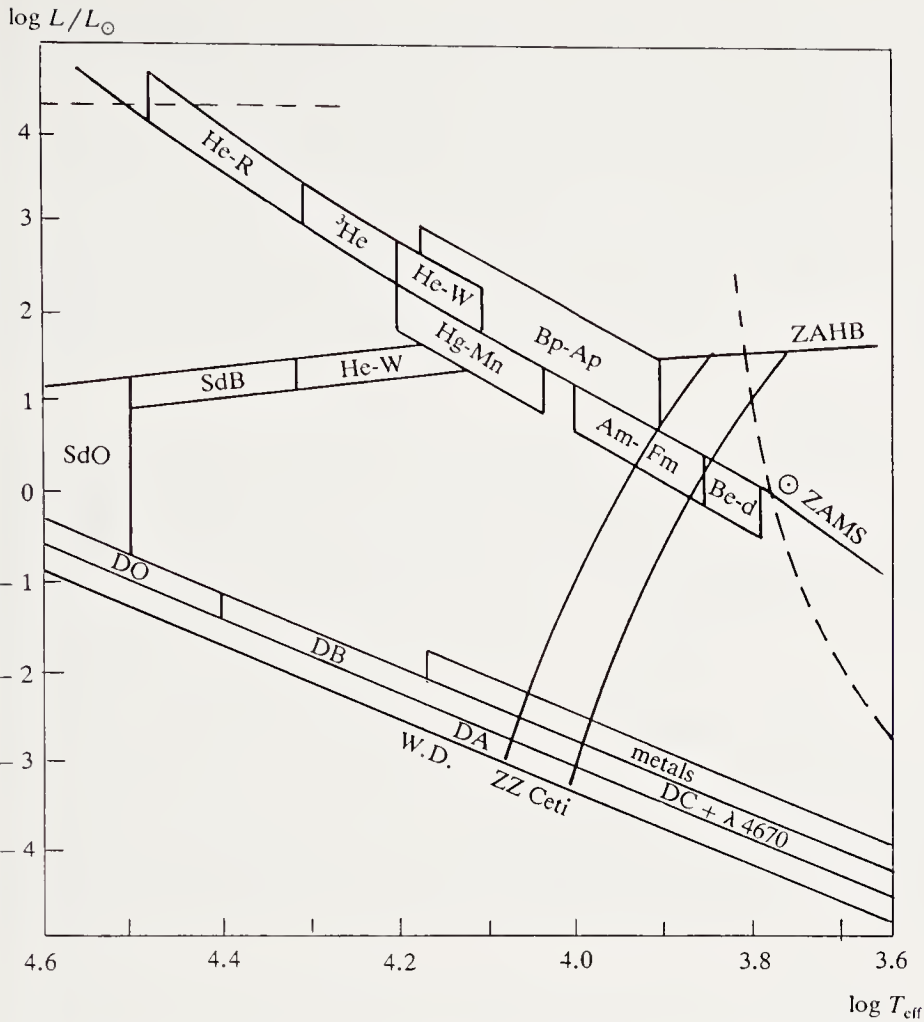
Fig. 5.14. Abundance anomalies. HR diagram for stars with peculiar chemical composition (CP stars). The figure shows the ZAMS, the ZAHB with its extension to the sub-dwarf region, and the white-dwarf sequence, in a $\log T_{\text{eff}} - \log(L/L_\odot)$ plot. The positions of various types of stars of peculiar chemical composition are shown. There is no attempt to show their ranges of luminosity. Be-d: F stars deficient in beryllium. He-W: stars underabundant in helium. ^{3}He: stars with little helium and anomalous isotopic ratio ^{3}He/^{4}He. He-R: helium-rich stars. SdB: B subdwarfs. SdO: O subdwarfs. The white-dwarf sequence is divided into a hydrogen (DA) sequence and a helium (DO, DB, DC, $\lambda4670$) sequence. "Metals" implies that the white dwarfs have metal lines in absorption. The other notations are standard. The broken line is at $M_{\text{Bol}} = -6$ and shows the upper limit of the region where diffusion can be important. Above this, mass loss cancels the effects of diffusion. The broken curve to the red side of the figure is the limit where the diffusion time at the base of the convection zone is equal to the lifetime of the star. We note that the Sun is to the right of this line: diffusion has a negligible effect on its surface composition. In the region between these two bounds diffusion can produce abundance anomalies. The instability strip is marked. It crosses the horizontal branch at the edge of the diffusion region (RR Lyrae); variable main-sequence stars (δ Scuti stars) and white dwarfs (ZZ Ceti stars) are in the diffusion region. (From S. and G. Vauclair, *Ann. Rev. Astron. Astrophys.* **20**, 37 (1982))

The Cepheid instability strip, extended to the main sequence and the white dwarfs, cuts the two sequences at the δ Scuti stars (which coincide with the Am stars) and the variable white-dwarf stars (ZZ Ceti).

We give here a brief indication of the spectral characteristics of the various stars concerned, and the reasons for their anomalies.

Am Stars. The classic definition is: the Balmer lines indicate late A spectral type, the metal lines type F, the H and K lines an early type. In other words there is an excess of metals and a deficit of calcium. The δ Scuti variables are in the same region of the HR diagram as the Am stars, but have mutually exclusive properties.

Gravitational separation begins with the separation of helium and hydrogen. As the helium sinks, the structure of the outer layers changes; the convection zone associated with helium ionization disappears and the interior radiative zone begins at a level where diffusion can occur at the base of the H–He I convection zone.

At the same time, the decrease in the helium abundance removes a possible way of driving oscillations by the kappa mechanism (see Chap. 6). There is a boundary region where there is still enough helium to drive oscillations, but where separation of metals can occur. These stars have attenuated Am features.

Rotation and turbulent diffusion retard gravitational separation. In the first phase, characterised by the establishing of a concentration gradient, the timescale changes from t_{mic} to t_{turb}:

$$t_{\mathrm{mic}} = \frac{H_{\mathrm{HCZ}}}{k_\rho D_{\mathrm{mic}}} .$$

Turbulent diffusion acting on ∇c will in the end quickly remove the trace element. A calibration of D_T based on the lithium destruction rate in the Hyades (see below) gives an estimate of the equatorial velocity required to prevent element separation in an A star. We find $V_{\mathrm{eq}} \gtrsim 50$ km s^{-1}. However this is not valid for a star in a close binary. Here the dissipative effects of the tides (see Sect. 3.8) bring the two stars towards rigid rotation, and the differential rotation producing the turbulence is of higher order in $\varepsilon = (\Omega^2 R^3/GM)$. The fact that almost all Am stars belong to short-period binaries (1 to 100 d) thus explains why the maximum observed equatorial velocity is about 100 km s^{-1}.

The excesses predicted by settling theory can be suppressed by a very weak stellar wind, of the order of $10^{-14}\ M_\odot$ y^{-1}. It is required only that the wind produce a velocity of the order of the diffusive velocity at the base of the convection zone. The wind then removes those elements pushed into the atmosphere by radiation pressure.

Magnetic Ap Stars. Magnetic fields of tens to several thousand gauss have been detected in main-sequence stars with temperatures between 8000 and

30 000 K. All these magnetic stars show abundance anomalies, which also depend strongly on the temperature. The coolest stars have an excess of rare earths (Sr, Eu). The intermediate stars between 12 000 and 15 000 K have an excess of silicon. The hottest stars are helium rich.

These anomalies vary in time. For a given star they all vary with the same period. This is explained in terms of a rotating star with a frozen-in magnetic field. The importance of the magnetic field for the anomalies is incontestable. Non-magnetic stars (i.e. with as yet unmeasurable fields) of effective temperature 12 000 K have an overabundance of mercury and manganese, while magnetic stars have an excess of silicon.

White Dwarfs. Gravitational separation of hydrogen in white dwarfs occurs rapidly, because of the strong gravitational field and the small distances involved, despite the high density. In the DB stars the outer convection zone dredges up elements from deep in the star and brings helium to the surface. The occurrence of this behaviour depends of course on the mass of the surface hydrogen layer.

5.5.4 Abundance of Trace Elements: Lithium Burning

The underabundance of lithium in Population I stars by comparison with the cosmic value ($\log_{10}(\text{Li/H}) = -9$) is closely related to lithium burning. This depends very strongly on the temperature (as about T^{20}) and is completed on a timescale short compared to the nuclear lifetime of the star, except in the outer layers ($T < 2 \times 10^6$ K). The presence in the Hyades of a lithium deficiency (lithium dip) between effective temperatures 6400 K and 6700 K (Fig. 5.15) raises new problems. The beginning of the formation of a lithium dip in younger clusters suggests that it grows with age and depends on stellar properties.

Several mechanisms have been proposed to explain the lithium deficit: (1) destruction of lithium at the base of the convection zone before the main-sequence stage; (2) destruction of lithium at the base of the convection zone because of the extension of the mixing region by overshooting; (3) lithium transport by turbulent diffusion induced by rotational instabilities into the lithium-burning region; (4) transport to the surface by meridional circulation of layers where lithium has been destroyed; (5) diffusive transport induced by internal waves; and (6) radiative and gravitational separation.

We examine briefly the effectiveness of these mechanisms.

(1) The contribution of pre-main-sequence burning to the lithium deficiency must certainly be considered, but by itself it does not explain the obvious age dependence of the lithium abundance; (2) the second idea requires the physics of the convection zone to depend strongly on the star's mass; (3 and 4) the third and fourth suggestions assume a relation between the burning rate and the rotational velocity, but the large intrinsic velocity dispersion around the lithium dip (Fig. 5.15) makes a one-to-one relation
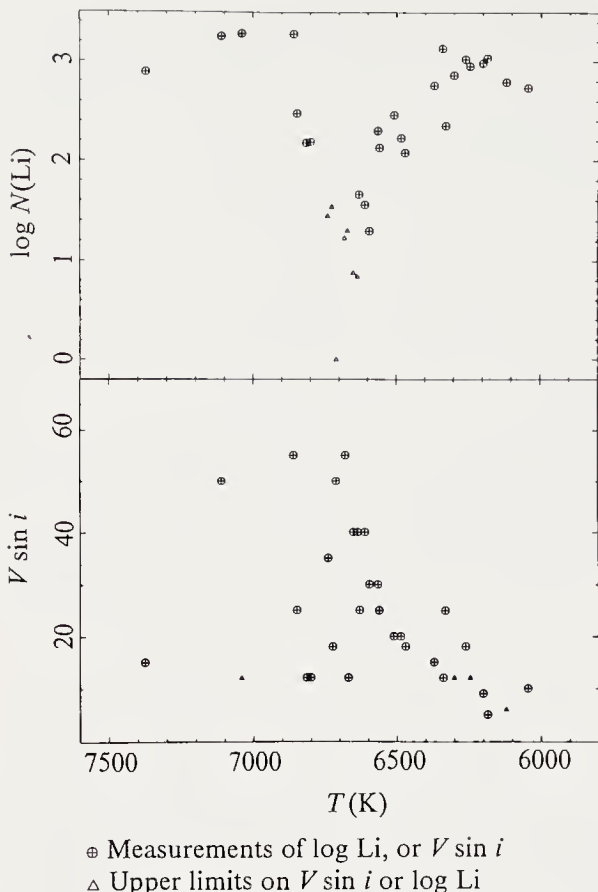
Fig. 5.15. Lithium abundance and projected equatorial velocities $(V \sin i)$ in the Hyades. *Upper part*: Li abundance for $\log N_H = 12$; *lower part*: equatorial velocities $V \sin i$ in km s^{-1}. (From A.M. Boesgaard, *P. A. S. P.* **99**, 1067, 1987. By kind permission of the Astronomical Society of the Pacific)

⊕ Measurements of log Li, or $V \sin i$
△ Upper limits on $V \sin i$ or log Li

between lithium depletion and rotation implausible; (5) gravity waves are generated by the motions in the convection zone, non-linear effects inducing a diffusive process which can carry the lithium to the burning level; and (6) radiative and gravitational settling are clearly significant.

For a diffusion coefficient D, the characteristic destruction time for lithium depends on the distance Δr from the burning region to the base of the convection zone and to order of magnitude is

$$t_{\text{burn}} = \left[\int_0^{\Delta r} \frac{dr}{D^{1/2}} \right]^{-2},$$

where it is naturally important to define the distance Δr accurately. If diffusion dominates we have to solve the diffusion equation

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \rho D \frac{\partial c}{\partial r} \right) = \rho \frac{\partial c}{\partial t} + K(r)c$$

to obtain the surface concentration of lithium as a function of time, where $K(r)$ is the rate of destruction of lithium by thermonuclear reactions.

The behaviour of the lithium abundance as a function of mass for a cluster of given age (e.g. the Hyades) seems to be clearly related to the properties of the surface convection zone, its depth increasing as the mass

decreases (Fig. 3.4). The closer the base of the convection zone to the burning region, the more efficient the transport mechanism and the higher the burning rate become.

The lithium dip discovered in the Hyades is difficult to explain. Excluding rotation one has to consider gravitational and radiative settling and check the role of meridional circulation. In the interval $6000 < T_{\mathrm{eff}} < 7000$ K the start of the lithium dip could be caused by the increasing efficiency of transport by internal waves and the enhanced abundances by the near-disappearance of the convection zone. This problem is still under discussion (1993).

5.5.5 Abundance of Trace Elements: Formation of the ^{13}C Isotope

In the central regions of main-sequence stars the carbon cycle alters the relative abundances of ^{12}C, ^{13}C, ^{14}N, ^{15}N. At the end of the main-sequence lifetime the chemical composition as a function of mass fraction (Fig. 5.16) has a number of distinct characteristics: an abundance peak of ^{13}C, followed by a rapid decrease of the two carbon species towards the interior, accompanied by a growth in ^{14}N.

Observations show that the ratio (^{12}C/^{13}C) decreases along the giant branch, from about 80 on the main sequence to about 10. The lowering of the base of the convection zone during the passage along the giant branch does not dredge up enough ^{13}C to reach (^{12}C/^{13}C) ratios below 40. We thus need a mechanism transporting the ^{13}C isotope outwards from the region
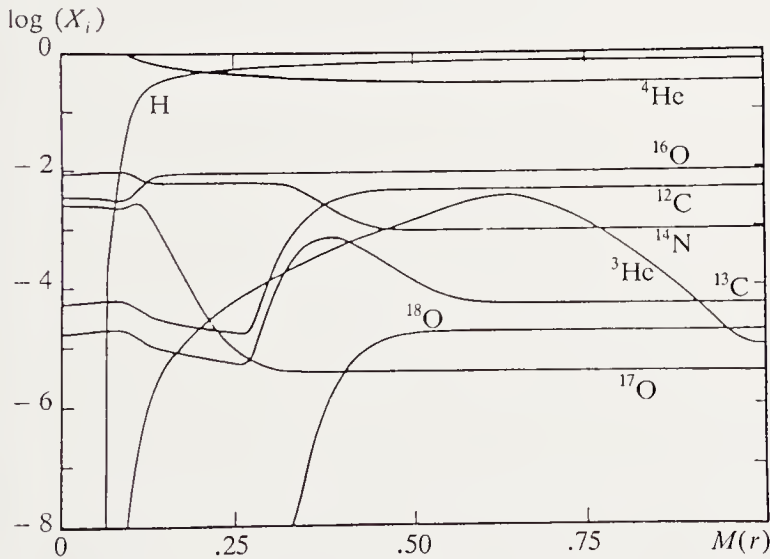
Fig. 5.16. Chemical composition inside a main-sequence star ($M = 1\,M_\odot$) as a function of mass fraction M_r/M_*. (After D.S.P. Dearborn et al., *Astrophys. J.* **203**, 455, 1976. Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1976 The American Astronomical Society)

where it is abundant, during the main-sequence phase, so that the lowering of the base of the convection zone dredges more carbon ^{13}C to the surface.

The presence of a molecular-weight gradient $\nabla\mu$ prevents turbulent diffusion from transporting ^{14}N towards the surface, so that the (C/N) ratio stays in a restricted range. In contrast, turbulent diffusion, acting in a region where $\nabla\mu$ remains negligible, brings considerable quantities of ^{13}C to the outer regions, and thus appears to offer an explanation of the observations.

6. Variable Stars

6.1 Classification of Variable Stars

There is a wide variety of variable stars, and in some sense all stars are variable, the Sun being a typical example. The proof of variability may depend on the detection threshold, for luminosity variations (the lower the threshold the more variables are found); or the resolution, whether angular (the Sun, stars studied by speckle interferometry), temporal (days, hours, minutes, or fractions of a second), or spectral (radial velocity); or the wavelength (from X-rays to radio waves). We are concerned here with *intrinsic variables*, i.e. isolated stars whose variability has an internal physical origin. Other causes of marked variation are known, such as motion (eclipsing binaries), accretion of material from another star (cataclysmic variables), rotation (pulsars), or accretion (gamma-ray bursts). Some stars have variations combining rotation with a particular type of activity (flare stars).

By limiting ourselves in this chapter to intrinsic variables we necessarily defer until later the study of the other variables: cataclysmic variables and pulsars to Chap. 8 ("The Last Stages of Stellar Evolution"). Even though their variability has an internal origin, flare stars involve a quite different mechanism and we give an introduction to their study in Chap. 7 ("Solar and Stellar Activity").

The classification of variable stars is primarily based on the analysis of their *periodicities* (periodic and irregular variables) and the *forms of their light curves*; further more refined study, particularly spectroscopic, reveals the nature of the variability. The first variable stars were found by accident (see the book by Harwit, *Cosmic Discovery*, 1981), followed by systematic searches using a blink comparator. This instrument allows comparison of photographic plates taken by the same telescope at different epochs. An optical system superposes the images of the two plates being compared via two microscopes, over the entire field. Two synchronised discs alternately occult the two light beams. The image of a variable star is different at the two epochs, and appears to the observer as scintillating, allowing immediate recognition. The modern detection method uses image processing. The *Hipparcos* satellite should thus permit the discovery of thousands of variable stars.

The general catalogue of variable stars, compiled at present in the USSR under the auspices of the International Astronomical Union, contains data

Table **6.1**. Julian date for 1 January 12 hours UT of each year

Gregorian year		0	1	2	3	4	5	6	7	8	9
1900	241	5 021	5 386	5 751	6 116	6 481	6 847	7 212	7 577	7 942	8 308
1910		8 673	9 038	9 403	9 769	*0 134	*0 499	*0 864	*1 230	*1 595	*1 960
1920	242	2 325	2 691	3 056	3 421	3 786	4 152	4 517	4 882	5 247	5 613
1930		5 978	6 343	6 708	7 074	7 439	7 804	8 169	8 535	8 900	9 265
1940		9 630	9 996	*0 361	*0 726	*1 091	*1 457	*1 822	*2 187	*2 552	*2 918
1950	243	3 283	3 648	4 013	4 379	4 744	5 109	5 474	5 840	6 205	6 570
1960		6 935	7 301	7 666	8 031	8 396	8 762	9 127	9 492	9 857	*0 223
1970	244	0 588	0 953	1 318	1 684	2 049	2 414	2 779	3 145	3 510	3 875
1980		4 240	4 606	4 971	5 336	5 701	6 067	6 432	6 797	7 162	7 528
1990		7 893	8 258	8 623	8 989	9 354	9 719	*0 084	*0 450	*0 815	*1 180
2000	245	1 545									

on all the known variable stars. They are classified here by the shape of their light curves. The catalogue contains information on all stars on which data have been published, so that all recognised variables appear. They are denoted in order of discovery by the name of the constellation preceded by one or two letters (R, S,...,Z, RR, RS,...,SS, ST,...,ZZ,...,AA,...,QZ) or, for constellations where this system of letters is exhausted, by the letter V followed by a number (the first being 335).

The dates of the events occurring in each star are given in Julian days, the prolongation of the Julian calendar beyond 1582, when the Gregorian calendar was introduced. Julian days are counted from 1 January 4713 BC. Table 6.1 gives the correspondence between the Julian day and 1 January 12 hours UT from 1900 to 2000.

The phase of light variation of periodic variables is counted from a light minimum to the following minimum.

6.1.1 Inventory

When the intrinsic variables are plotted on the Hertzsprung–Russell diagram (Fig. 6.1), we see that they are distributed in well-defined areas. The positions of these provide a natural division into four groups:

(a) A (near) vertical strip, from visual magnitude +1 to −6, containing the most regular variables. This is often referred to as the *instability strip*.
(b) To the right of the instability strip, roughly parallel to the main sequence and 9 to 10 magnitudes above, is the *red sequence*. Stars in this region show different types and degrees of irregularity.
(c) Near the main sequence there are several groups of pulsating stars, all with short periods and small oscillation amplitudes.
(d) The ZZ Ceti variables, on the white-dwarf cooling sequence.
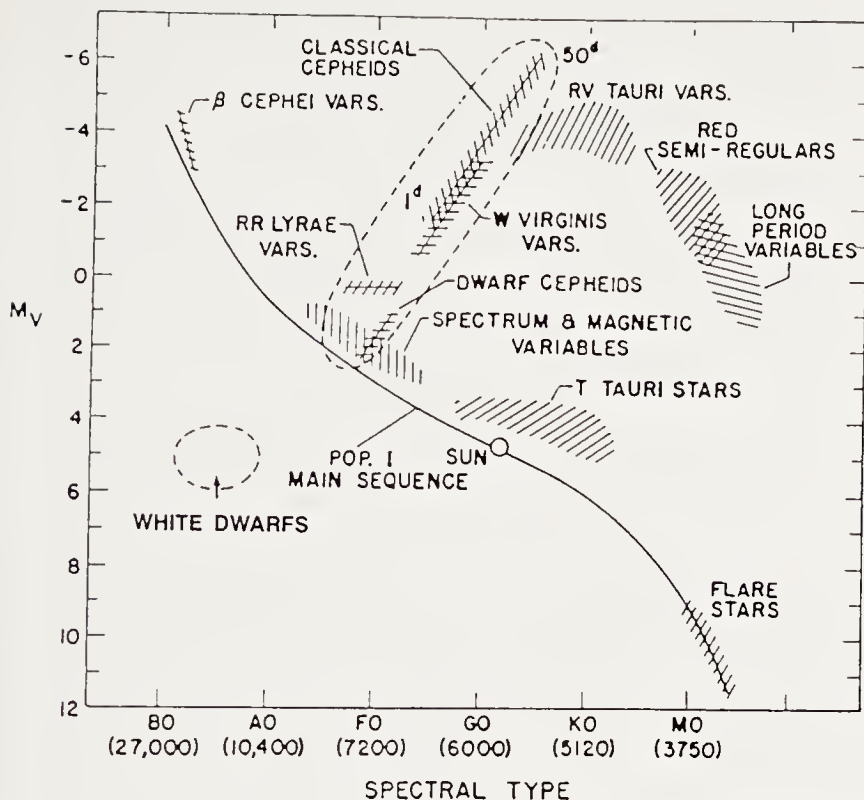
Fig. 6.1. Positions of intrinsically variable stars on the Hertzsprung–Russell diagram

6.1.2 Periodic Variables

At present and from ground-based observations only, there are around 14 000 periodic intrinsic variables. Their periods lie between a few seconds (periodic white dwarfs) and several hundred days (Mira Ceti type). The histogram of variable stars reflects both observational selection effects and intrinsic properties. As for all stars, the observed number is affected by the number of progenitor stars and the time over which the observed characteristics are present. The most numerous variables involve the presence of suitable physical conditions for variability in stars which are evolving slowly, or relatively slowly (white dwarfs, main sequence, horizontal branch, yellow and red giants).

Classification by period is not unique; there are clear differences in the light curves. A complete classification involves the luminosity class, spectral type, and stellar population (I or II).

In the group of periodic variables and the *General Catalogue of Variable Stars* (Kukarkin et al.), the Cepheids comprise about 700 stars, the W Virginis stars about a hundred, RR Lyrae stars of various types about 6000, dwarf Cepheids and δ Scuti stars around 150. There are about 5000 long-period variables; about 50 blue variables of β CMa type and about ten ZZ Ceti variable white dwarfs. Table 6.2 summarises the properties of these stars under the designation *pulsating stars* (see Sect. 6.2 below). The galactic population can be estimated from the space distribution of the various

Table 6.2. Pulsating variables

Type	Periods	Characteristic period	Population	Spectral Type	M_V
RR Lyrae	1 h 5 to 24 h	0.5 d	II	A2 to F6	0.0 to +0.5
Classical Cepheids	1 d to 50 d	5 d to 10 d	I	F6 to K2	−0.5 to −6
W Virginis	2 d to 45 d	12 d to 20 d	II	F2 to G6	0 to −3
RV Tauri	20 d to 150 d	75 d	II	G, K	≃ −3
Red semi-regular variables	100 d to 200 d	100 d	I and II	(K),M,R,N,S	−1 to −3
Long-period variables	100 d to 700 d	270 d	I and II	M, R, N, S	+1 to −2
β CMa (β Cephei)	4 h to 6 h	5 h	I	B1 to B2	−3.5 to −4.5
δ Scu and dwarf Cepheids	1 h to 3 h	2 h	I	A2 to F5	+2 to +3
ZZ Ceti (white dwarfs)	1 min to 15 min	4 min	I	DA	+10

types; it is about 15 000 for the Cepheids, 150 000 for the RR Lyrae stars, and 200 000 for red variables of Mira Ceti type.

6.1.3 Irregular or Semi-regular Variables

These stars are ill-understood theoretically, and are briefly reviewed in Sect. 6.4. We give some details of their observed properties.

This heading applies to several types of variable stars among the red giants and supergiants (Fig. 6.1).

RV Tau Stars. These can be characterised as follows:

(a) The light curve alternately shows deep and shallow minima (Fig. 6.2) which occasionally switch.

(b) The light curve has transient irregularities.
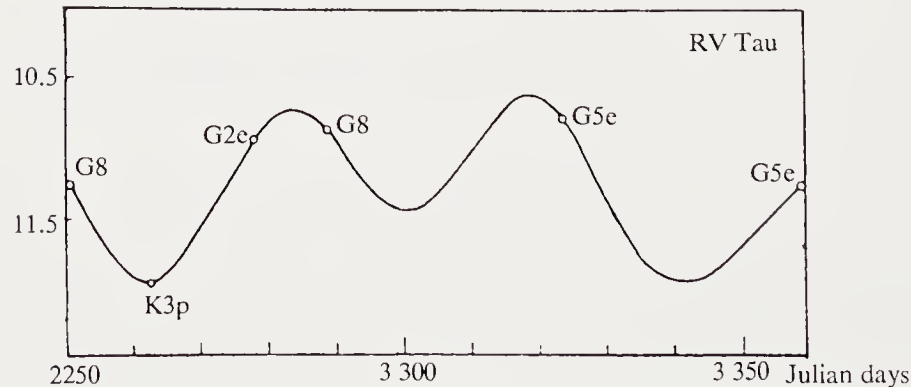
Fig. 6.2. Variations of the spectral type of RV Tau. (After J.C. Pecker and E. Schatzman, in *Astrophysique Générale*, Masson 1963, Fig. 220, p. 395)
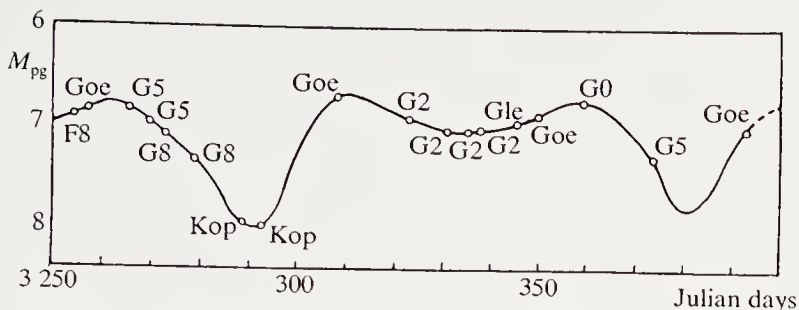
Fig. 6.3. Variation of the spectral type of U Mon. (After J.C. Pecker and E. Schatzman, in *Astrophysique Générale*, Masson 1963, Fig. 220, p. 395)

(c) The mean brightness is constant or very slowly varies with a long period.
(d) The mean spectral type is F, G, or K.

There are at present more than 100 known stars with these characteristics. The brightest, R Scuti, has magnitude 6.1 at maximum. The periods lie between 30 and 50 days, and their amplitudes are in general between 1 and 2 magnitudes. Spectral classification is difficult, because the hydrogen lines and the Ca I 4227 line are abnormal, probably indicating a large divergence from thermodynamic equilibrium in the surface layers. In contrast the G band of CH places the RV Tau stars in class Ia or Ib. In some of them, such as U Mon, TiO bands appear at the deep minima and hydrogen lines near the maxima (Figs. 6.2 and 6.3).

These are low-mass post-asymptotic-branch stars and evolving towards the white-dwarf branch.

Semi-regular Variables of Type K. This is a small group of stars related to the RV Tau stars, but having more irregular light variations. The RV Tau and semi-regular variable regions extend towards the region of the long-period variables such as Mira Ceti, red supergiants of types M, R, and N or S. This suggests that these various types of supergiants have a common mechanism which brings about instability without all having the properties needed for regular pulsation.

6.1.4 β CMa Stars

These are blue stars with very short periods, of the order of a few hours. Their light variation is slight: 0.1 magnitudes in the blue; the radial-velocity amplitude is between 5 to 100 km s^{-1}.

Most β CMa stars have two periods and show the beat between them. For example β CMa, of spectral type BII–BIII has periods $P_1 = 0.250\,022\,44$ d with amplitude $K_1 = 5.8$ km s^{-1} and $P_2 = 0.251\,3003$ d with amplitude $K_2 = 2.0$ km s^{-1} (the latter has varied). β Cep itself has a radial velocity amplitude of 45 km s^{-1}.

These stars also have quite long rotation periods ($V \sin i \simeq 5$ to 30 km s^{-1}).
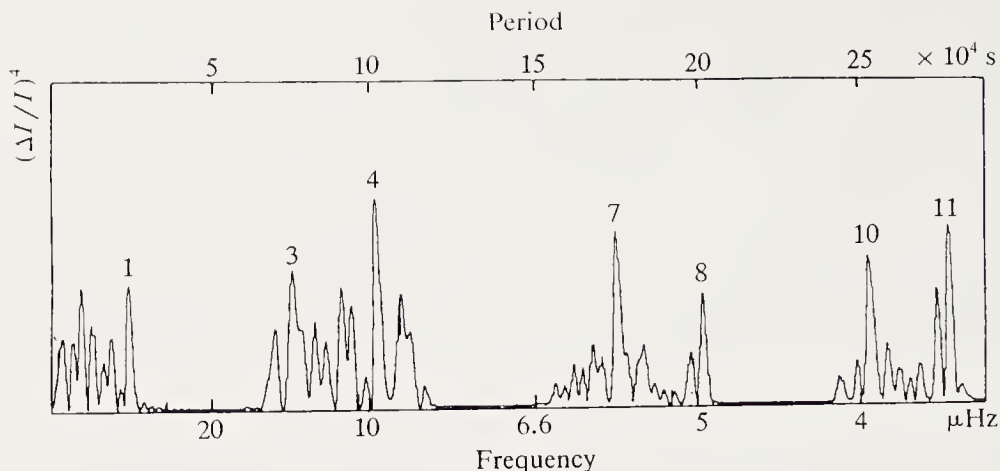
Fig. 6.4. Power spectrum of Procyon in the range 1.17–1.65 mHz. The frequency difference $\delta\nu_{nl}$ (Sect. 6.5.3) is proportional to the inverse of the frequency for modes of increasing degree, so it is convenient to take as abscissa a scale proportional to *inverse* frequencies. We thus clearly see a regular system of equidistant peaks. Not all the peaks allowed by the observing window appear because of the interruption of observations in the day. The period of the fundamental (about 2.4×10^4 s) is the beat period between successive harmonics of the same degree. This period defines the spacing $\Delta\nu = 79.4\,\mu\text{Hz}$

Other groups of variables related to the β CMa stars have been found, with spectral types B0 V to B Ia, with periods of a few hours to a day.

6.1.5 Main-Sequence Stars

Since the discovery of high-mode, high-degree non-radial oscillations in the Sun, there have been efforts to study other stars in corresponding period ranges. Gelly, Grec, and Fossat (1986) measured the spectrum of radial-velocity variations of Procyon (α CMi) (F6 IV–V) and of α Cen A (G2 V) (Fig. 6.4). Noyes et al. (1984) measured the flux variations of the calcium H and K lines in ε Eri (K2 V). Fourier analysis in the period range from 10 to 15 min. shows equal spacing of the eigenmodes of 79.4 μHz for Procyon (see the discussion of the theoretical problem in Sect. 6.5.3) and of 165.5 μHz for α Cen A. For ε Eri, the spacing is of 172 μHz. The significance of these results is still under discussion. We will see later the meaning of equidistant frequencies and the fundamental importance of these modes for the study of the internal structure. These pulsations are characterised by very small amplitudes (of the order of 1 m s^{-1} in Doppler velocity).

260

6.2 Periodic Pulsating Variables (RR Lyrae, Cepheids, Miras)

Observational studies and theoretical work on pulsating variables are so closely interwoven that it is first necessary to give some idea of the interpretation of these phenomena.

For a spherical non-rotating star, a linear analysis factorising the displacement into spherical harmonics and functions of the radial variable r gives a description of the motion in terms of a complete set of orthonormal functions. We thus distinguish between radial and non-radial oscillations. The latter are classified in p modes (the restoring force is pressure) and g modes (the restoring force is gravity). Radial oscillations, called f modes by Cowling, can be analysed at each order. The zero order, or fundamental, has a node at the centre and an antinode at the surface. The first order, or first harmonic, has a node between the centre and the surface.

RR Lyrae stars, Cepheids, and Miras oscillate in the fundamental mode, or possibly the first harmonic.

The theory of these pulsating variables allows us to interpret the raw data and we shall present the latter along with their interpretation.

The main problem is that of the pulsation itself. We have to explain why a star oscillates rather than being in hydrostatic equilibrium. We expect that stability conditions will define the boundaries of the *instability strip* within which the pulsating variables are found. More exactly, the hydrodynamic model should explain the main features of the light curve and its dependence on period. The theoretical model should allow us to use both photometric data *and* measurements of radial velocity to determine the radii of variable stars. Finally, the need for agreement between the mass deduced from the period and the theoretical mass deduced from the evolutionary track should give an idea of the rate of mass loss.

During its evolution a star enters and leaves the physical region of instability. The study of stellar evolution should indicate the evolutionary status of a given variable star. Remembering that the oscillation period is proportional to the the sound travel time from the centre to the surface, we see that the secular variation of the period gives information on how the global properties of a variable star evolve. Table 6.3 gives several examples.

We can see that the observed secular variations and the estimates from stellar-evolution theory are not in contradiction.

6.2.1 Radial Velocity, the Light Curve

Observational Data. Each class of variable star has characteristic light-curve properties. For pulsating variables the curve is asymmetric and has a sort of bump on the descending part (Fig. 6.5).

Table 6.3. Secular period variations

Star	Duration of observations	Period (days)	Variations of the period		
			Date	Days/year	s/year
δ Cep	1784[a]–1919	5.3663770	1883	-9.16×10^{-7}	-7.9×10^{-2}
T Mon	1881–1920	2.700313	1886	$+4.17 \times 10^{-5}$	3.6
SV Vul		45.2			$-(254 \pm 10)$
RR Lyrae[b]		0.5			5.1×10^{-10}

[a] Date of the discovery by Goodricke.
[b] Mean values for the RR Lyrae stars in the globular cluster ω Cen.
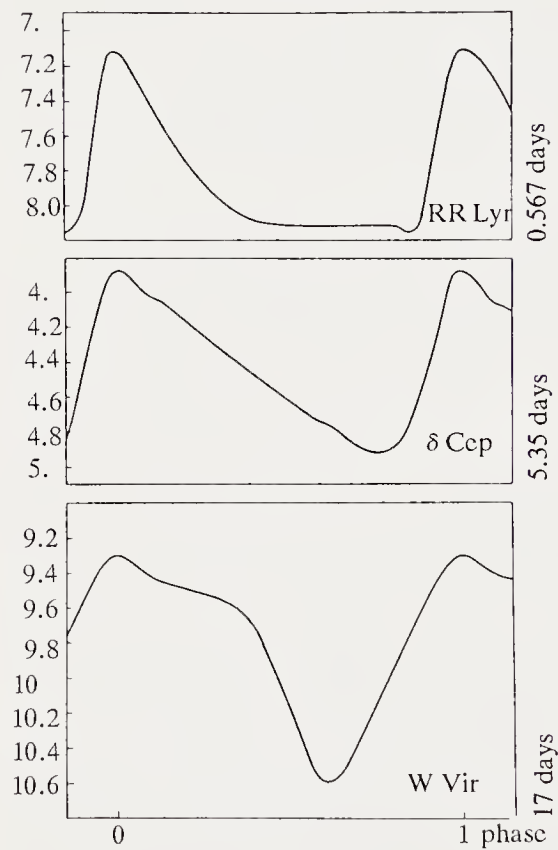
Fig. 6.5. Light curve of 3 periodic variable stars. (From J.C. Pecker and E. Schatzman, in *Astrophysique Générale*, Masson 1963, Fig. 200, p. 374)

The *asymmetry* of the light curve can be measured by the difference between the phase $\varphi(M)$ of maximum and the phase $\varphi(m)$ of minimum:

$$\varepsilon = \varphi(M) - \varphi(m).$$

When we plot the logarithm of the period against the asymmetry, we see that there is a discontinuity at about 10 days. However, if we plot for periods less than 10 days the difference

$$\varepsilon' = \varphi(M') - \varphi(m)$$
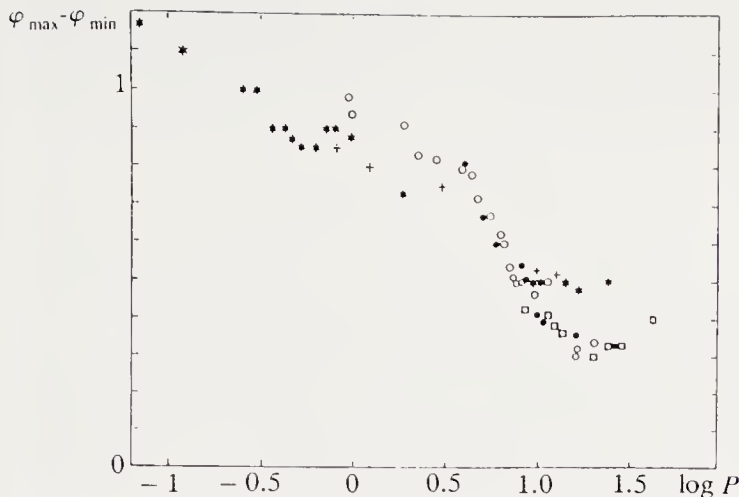
Fig. 6.6. Asymmetry of the light curves of periodic variables. • = Galactic Cepheids of δ Cephei type, ∗ = Galactic Cepheids of RR Lyrae and W Virginis type, + = Globular cluster variables, □ = Variables in Andromeda and M 33, o = Variables in the Large and Small Magellanic Clouds. (From J.C. Pecker and E. Schatzman, in *Astrophysique Générale*, Masson 1963, Fig. 202, p. 375)

between the phase $\varphi(M')$ of the bump or secondary maximum of the light curve and the phase $\varphi(m)$ of the minimum, the points are continuous (Fig. 6.6). One has the impression that the bump, which is just a detail of the light curve for $P < 10$ d, becomes the main light variation for periods $P > 10$ d, and, conversely, that the main light variation for periods $P > 10$ d is only a secondary phenomenon for periods $P < 10$ d.

Cepheids separate into two main groups: classical Cepheids (δ Cephei type), and short-period variables (RR Lyrae type), which merge into the anomalous Cepheids of W Virginis type at long periods.

Precise photoelectric photometry has revealed finer differences amongst the Cepheids: for example, the light curve of η Aql has a bump essentially absent from the light curves of δ Cep and ζ Gem. The light curve of δ Cep is extremely asymmetric, while the light curve of ζ Gem is almost symmetrical.

Radial Velocities. During pulsations spectral lines undergo periodic Doppler shifts, indicating periodic motion of the stellar surface.

The radial velocity v, measured by the position of line centre, results partly from the motion V_R of the star with respect to the Sun and partly from the average $\overline{V}$ of the radial velocities $V = dR/dt$ of different parts of the star's disc. In other words,

$$v = V_R - \overline{V}, \tag{6.1}$$

where the velocities v and V_R are taken as positive for motion away from the observer. Assuming that the variation of Cepheids is caused by a spherically symmetric pulsation of the star, we have to study the relation between the radial matter velocity $dR/dt = V$ and its average value on the disc.

Assuming for simplicity a constant relation between V and $\overline{V}$, $\overline{V} = kV$, we find V_R while imposing the condition $\int_0^P V \mathrm{d}t = 0$, expressing the fact that during a period P the material layers return to the same places:

$$V_R = \frac{1}{P} \int_0^P V \mathrm{d}t. \tag{6.2}$$

To a first approximation all the lines have the same Doppler shift at the same phase, indicating a motion of the whole stellar atmosphere.

To find the parameter k, we have to model the atmosphere and calculate the mean value of the radial velocity as a function of phase. As an example, we could suppose that the measured radial velocity is an average of the radial velocity over the stellar surface weighted by the background continuum intensity, which assumes the same limb-darkening for the line and continuum. Under these conditions we have simply

$$\overline{V} = \frac{\int_0^1 V \cos\theta I(\theta) \cos\theta \, 2\pi \mathrm{d}(\cos\theta)}{\int_0^1 I(\theta) \cos\theta \, 2\pi \mathrm{d}(\cos\theta)}. \tag{6.3}$$

Using a linear representation of the limb-darkening

$$I(\cos\theta) = I_c \left[1 - x + x \cos\theta\right]$$

we get easily (x is constant)

$$\overline{V} = \frac{4 - x}{6 - 2x} V. \tag{6.4}$$

This fraction varies slowly as x varies from 0 to 1, and it is usual to adopt the relation

$$\overline{V} = \tfrac{17}{24} V = 0.708 V \left(x = \tfrac{3}{5}\right). \tag{6.4'}$$

This is only an approximate relation. In reality the ratio $(\overline{V}/V)$ varies with phase. The exact value of $(\overline{V}/V)$ is important in the determination of Cepheid radii by the classic Baade–Wesselink method. Before discussing this method we note some features of the radial velocity curves.

(a) For most Cepheids the Doppler shifts give a regular radial velocity curve.

(b) In certain stars such as RR Lyr, W Vir, some lines are observed to become doubled at the point where the radial velocity $(\overline{V} - V_R)$ changes sign, particularly the strong hydrogen lines. Anticipating the explanation of Sect. 6.2.2, this corresponds to the passage of a travelling wave through the stellar atmosphere. The wave which has just passed through the atmosphere is still visible when the next wave appears deep in the atmosphere.

(c) In most Cepheids one sees emission lines at certain phases, always with large negative velocities, i.e. corresponding to rapid matter motion towards us, and thus towards the upper atmosphere.

RR Lyr has $H\alpha$ and $H\gamma$ in emission near phase -0.09 (the origin corresponds to maximum light). The radial velocity is -99 km s^{-1}. The emission lines appear just before the doubling of $H\alpha$ at velocities -130 and -20 km s^{-1}.

W Vir has intense hydrogen emission lines during phases 0.65–0.825. Later, when the lines become doubled at phases 0.94–0.10, the emission lines are weaker. The measured radial velocity of the emission lines varies from -80 to -85 km s^{-1}. The doubled emission lines have velocities -92 and -40 km s^{-1}. The radial-velocity curve of the emission lines joins up with that of the absorption lines, and it is possible to follow the velocity of a gas layer for 1.45 periods.

δ Cep has the H and K lines in emission at phase 0.97. The measured radial velociies are then

— line H (emission):	-47 km/sec
— line K (emission):	-46 km/sec
— line K (centre of the absorption):	$+22$ km/sec
— six other lines:	-4 km/sec.

The Baade–Wesselink Method. We let F_ν be the monochromatic flux per square centimetre of the star's surface. The monochromatic magnitude at a given time is

$$m = -2.5 \log F_\nu - 5 \log R + \text{const}. \qquad (6.5)$$

Considering two instants (1) and (2) when the star has the same colour, we can assume that the star's surface brightness is the same and that the difference in luminosity comes solely from the change in radius,

$$m_2 - m_1 = -5 \log(R_2/R_1). \qquad (6.6)$$

Assuming further that the variation of photometric radius is identical to that calculated from the radial velocities, we set $10^{-0.2(m_2-m_1)} = n$. The two radii R_1 and R_2 are related to the mean radius R and the calculated radius variations by $R_1 = R + \Delta R_1$, $R_2 = R + \Delta R_2$. We thus have

$$\frac{R + \Delta R_2}{R + \Delta R_1} = n \qquad (6.7)$$

which gives

$$R = \frac{n\Delta R_1 - \Delta R_2}{1 - n}. \qquad (6.8)$$

Table 6.4. Cepheid radii by the Baade–Wesselink method

	RR Lyr	δ Cep	η Aql
$R \times 10^7$ km	0.50 ± 0.06	3.7 ± 0.1	4.7 ± 0.2
$R/R_\odot$	7.2 ± 0.9	53 ± 2	68 ± 2

The constancy of the right-hand side with phase is a test of the pulsation hypothesis. Stebbins (1953) obtained the results of Table 6.4 from 6-colour photometry.

The values found for R are only very approximate because of the limited validity of relation (6.4′), the uncertainty in the parameter k, and the determination of ΔR_1 and ΔR_2 from radial velocities.

We can also consider the problem as follows. Theory shows that there is a linear relation between the monochromatic flux variation and the variation of the colour index (denoted CI):

$$-2.5 \Delta \log F_\nu = a\Delta(CI), \tag{6.9}$$

where the symbol Δ stands for the variation between two observations. We thus get immediately

$$\Delta \log R = 0.2[a\Delta(CI) - \Delta m]. \tag{6.10}$$

A suitable choice of the parameter a then gives a curve for $\Delta \log R$ of the same form as that deduced from the radial velocities.

For δ Cep, η Aql this procedure works if a is taken as 2.3 and 3.3 respectively, although the values calculated from model-atmosphere theory are 4 and 5.4. Using the theoretical value of a gives a radius of the same order as that from the radial velocities, but significantly smaller, and with different variations.

Christy (1968) has made a detailed comparison of the models applied to β Dor, $P = 9.84$ d, and finds from the Baade–Wesselink method a radius too large by 30 % compared with that deduced from the relation $P\sqrt{\rho} =$ constant (see below, p. 270) and from the period–luminosity relation (see p. 275). He explains the difference as being due to a variation of the depth of the photosphere with phase. The Baade–Wesselink method can only be used if significant systematic corrections are made.

6.2.2 Phase Lag and the Cause of the Instability

In an adiabatic oscillation of a gaseous mass, the relation $P \sim \rho^\gamma$ implies that the variations of velocity, density, temperature, and pressure are all in phase, and the luminosity and radius in antiphase: the luminosity should reach a maximum at the minimum radius.

In one dimension, the equations of continuity and motion are

$$\frac{\partial \rho}{\partial t} + \rho \frac{\partial v}{\partial x} = 0 \,, \tag{6.11}$$

$$\frac{\partial v}{\partial t} = +\gamma \frac{P}{\rho} \frac{\partial \rho}{\partial x} \,, \tag{6.12}$$

giving the solutions

$$\delta \rho \sim \exp i \left(\frac{\omega x}{c} - \omega t \right) \,, \tag{6.13}$$

$$\delta T \sim \delta \rho \,, \tag{6.14}$$

$$v \sim \exp i \left(\frac{\omega x}{c} - \omega t \right) \,, \tag{6.15}$$

$$\nabla \delta T \sim i \exp i \left(\frac{\omega x}{c} - \omega t \right) \,, \tag{6.16}$$

$$\delta R \sim \frac{1}{i} \exp i \left(\frac{\omega x}{c} - \omega t \right) \,, \tag{6.17}$$

showing that δR and $\nabla \delta T$, the latter being proportional to the flux, are in antiphase.

In reality, for Cepheids and RR Lyrae stars, the luminosity reaches a maximum at the maximum velocity, corresponding to a phase lag close to $(\pi/2)$ in sinusoidal motion.

To explain this requires the study of the linearised equations of motion in the non-adiabatic case. This shows that the cause of the instability and that of the phase lag are related and involve sub-photospheric ionisation zones of hydrogen, helium, and ionised helium. Here we shall try to bring out the physical basis of the instability and the phase lag, without discussing the detailed agreement with observation, which requires numerical solution of the full non-linear equations of motion.

Stellar Structure. Cepheids are stars of 6 to 7 $M_\odot$ and spectral type approximately F to K; such stars have extended low-density convection zones. At low densities the efficiency of convective energy transport is low (see Chap. 3) and the convection zone is close to radiative equilibrium. The envelope structure is determined by the absorption coefficient and ionisation equilibrium. In equilibrium the transition from partially to fully ionised hydrogen occupies a very small fraction of a scaleheight and behaves like a discontinuity. The ionisation zones of helium and ionised helium behave similarly. For hydrogen, two effects combine: the rapid increase in opacity near 10^4 K, and the sensitivity of the ionisation degree to the temperature (Fig. 6.7).
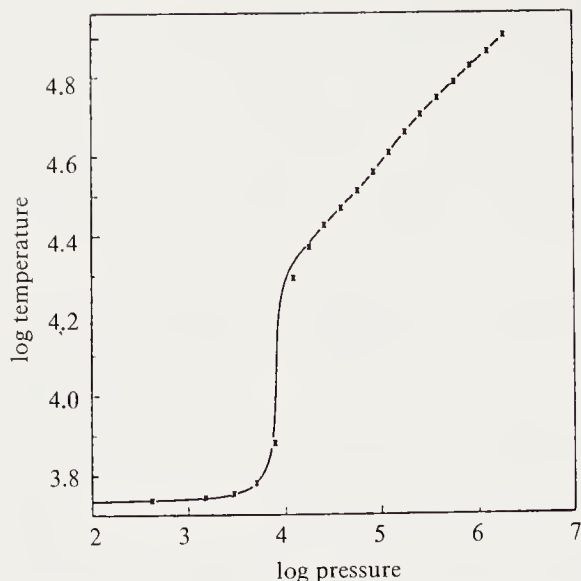
Fig. 6.7. Pressure–temperature relation in the radiative envelope of a classical Cepheid. Note the near-discontinuity in temperature, caused by the change in opacity near 10 000 K. (From R.F. Christy, *Quarterly Journal of the Royal Astronomical Society* **9**, 15 (1968). Reproduced by kind permission of the Royal Astronomical Society)

Equations. These are the standard equations of hydrodynamics. In Lagrangian coordinates the most useful independent space variable is the fractional mass

$$M(r) = \int_0^r 4\pi r'^2 \rho(r')\mathrm{d}r' , \tag{6.18}$$

where $\rho(r')$ is the density. Then the equation of motion is

$$\frac{\partial^2 r}{\partial t^2} = -\frac{GM(r)}{r^2} - 4\pi r^2 \frac{\partial P(\rho, T)}{\partial M} . \tag{6.19}$$

The diffusion (quasi-isotropic) approximation for radiative transfer holds throughout the star:

$$L(r) = -(4\pi r^2)^2 \frac{4\sigma}{3\kappa(\rho, T)} \frac{\mathrm{d}(T^4)}{\mathrm{d}M} , \tag{6.20}$$

where σ is Stefan's constant and $\kappa(\rho, T)$ the opacity. The heat-production equation is then

$$T\frac{\partial S(\rho, t)}{\partial t} = -\frac{\mathrm{d}L}{\mathrm{d}M} + \varepsilon(\rho, T) , \tag{6.21}$$

where S is the entropy and ε the rate of nuclear energy production per unit time and mass. At the free surface the boundary conditions are:

— in the equation of motion, $P(R_*) = 0$, where R_* is the stellar radius.
— in the heat equation, the surface flux, proportional to ∇T^4, is equal to the flux radiated at the surface temperature. Assuming the radiation is emitted at optical depth $\tau = 2/3$ we write

$$\left(\frac{\mathrm{d}T^4}{\mathrm{d}\tau}\right)_{(\tau=2/3)} = \frac{3}{2}(T^4)_{(\tau=2/3)} \,. \tag{6.22}$$

In a giant star, the rapid density increase towards the interior makes the amplitude of the oscillation very small inside radius $r = R_*/4$. In non-linear theory, and for numerical reasons, it is convenient to introduce a rigid boundary at a small but non-zero radius, which then contains all the energy sources. This gives the boundary conditions $\dot{r} = 0, L = \langle L(\theta) \rangle = L_0$, the average value of the luminosity. This means that the causes of the instability and the phase lag are to be sought outside the stellar core and do not involve the nuclear energy sources. Down to $r = 9R_*/10$ the envelope contains only 1% of the star's mass. This is therefore a chemically uniform region. The luminosity L_0, taken as constant at the internal boundary, is fixed by the stellar structure. In linearised theory, by contrast, the boundary conditions are written at the centre and the surface.

Linear Adiabatic Theory. Linearisation of the equations assumes that the amplitude of the oscillations is small. Linear theory can explain the causes of the instability and the period. It cannot give the detailed characteristics of the light curve or explain the finite amplitude of the oscillation. Nor can it give the red boundary (towards decreasing temperatures) of the instability strip, which involves non-linear dissipation (Sect. 6.2.5). The linear theory is however indispensable for an understanding of variable stars.

The classic dependent variable is the radius variation δr, written as

$$\delta r(t) = r\xi e^{i\omega t} \,, \tag{6.23}$$

and as independent variable we take

$$x = (r/R) \,. \tag{6.24}$$

For an adiabatic oscillation

$$P \sim \rho^\gamma \tag{6.25}$$

we do not need the heat equation (6.21); (6.18) and (6.19) now give a second-order equation

$$\frac{\mathrm{d}^2\xi}{\mathrm{d}x^2} + (4 - V(x))\frac{1}{x}\frac{\mathrm{d}\xi}{\mathrm{d}x} + \frac{V(x)}{x^2}\left(\frac{4 - 3\gamma}{\gamma} + \frac{x^3 R^3}{\gamma GM(x)}\omega^2\right)\xi = 0 \,, \tag{6.26}$$

where

$$V(x) = \rho(x)\frac{GM(x)}{P(x)xR_*} \tag{6.27}$$

is the ratio of gravitational to thermal energy.

Equation (6.26) is an eigenvalue equation for the amplitude $\xi(x)$. The boundary conditions at $x = 0$ and $x = 1$ fix the eigenvalues of the angular frequency ω. At $x = 1$ the pressure condition is

$$\frac{\delta P}{P} = -\left(4 + \frac{\omega^2 R^3}{GM}\right)\xi, \tag{6.28}$$

assuming that the pressure vanishes at the surface. From the adiabatic assumption we have the condition

$$\frac{d\xi}{dx} = \left(4 - 3\gamma + \frac{\omega^2 R^3}{GM}\right)\frac{\xi}{\gamma}. \tag{6.29}$$

As x tends to zero, the condition that ξ should remain finite is sufficient. Equation (6.26) can be expressed as

$$-O\xi + \omega^2\mu(x)\xi = 0, \tag{6.30}$$

where O is a differential operator. With the boundary conditions at $x = 0$ and $x = 1$ we have a classic eigenvalue problem for a linear self-adjoint system. This gives (Ledoux and Walraven 1958) an expression for the frequency of the fundamental, as the result of a variational principle:

$$\omega_0^2 = \min \frac{\displaystyle\int_0^1 \frac{\gamma P}{R^2}\left(\frac{d\xi}{dx}\right)^2 dx - \frac{1}{R^3}\int_0^1 x^2\xi^2\frac{d}{d\xi}[(3y-4)P]dx}{\displaystyle\int_0^1 \rho x^4\xi^2 dx}. \tag{6.31}$$

Ritter (1878–1883) investigated the case of a uniform sphere with a constant relative oscillation amplitude, finding

$$\omega_0^2 = (3\gamma - 4)\frac{4\pi G\rho}{3}. \tag{6.32}$$

The frequency found from more realistic models is always proportional to $\rho^{1/2}$. The relation $P_0\sqrt{\rho} = $ constant, where P_0 is the period, is a fundamental property of variable stars; it obviously remains to find the constant, as will be discussed later.

Applying the calculus of variations to (6.31) gives

$$\omega^2 = -\frac{\displaystyle\int_0^1 4\pi x^3\xi\frac{d}{dx}[(3\gamma-4)P]dx}{\displaystyle\int_0^1 4\pi x^4\xi\rho\, dx}. \tag{6.33}$$

The weighting function determining the period of the fundamental has its maximum near $x = 0.7$ and is small outside the interval 0.3–0.9. This implies that there is a physical parameter essentially fixing the form of the light curve. Christy (1968) has shown that the quantity $V(x)$, defined by (6.27) and calculated for $x = 0.83$, acts as this parameter.

270

Cause of the Instability. If we regard the departure from adiabaticity as a perturbation, we can follow Eddington and insert the solution of the adiabatic equation of motion into the heat equation. If the energy of the oscillation is W,

$$W = \frac{1}{2}\omega^2 \int_0^M \xi^2 r^2 \mathrm{d}M \,, \tag{6.34}$$

the rate of decrease or increase of the oscillation energy is given by K, where

$$K = \frac{1}{W}\frac{\mathrm{d}W}{\mathrm{d}t} \,, \tag{6.35}$$

where we have

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \frac{1}{2}\int_0^M (\gamma-1)\frac{\delta\rho}{\rho}\left(\delta\varepsilon - \delta\left(\frac{\partial L}{\partial M}\right)\right)\mathrm{d}M \,. \tag{6.36}$$

The meaning of this expression is that if the rate of nuclear energy production increases with compression (as is in fact true) the oscillation energy will grow; if heat is absorbed at maximum compression, the heat flux raises the oscillation energy. As we have mentioned above, the very small amplitude near the centre makes the nuclear contribution negligible in most cases.

Assuming an opacity $\kappa = \kappa_0 \rho^r T^{-s}$, we can calculate the contribution of the term $\delta(\partial L/\partial M)$. The amplitude of the oscillation is very small in the central energy-producing regions, and we can neglect the term in ε. The main term is thus

$$K = \frac{(\gamma-1)}{2\omega^2 I}\int 9\xi^2(4-3\gamma-(\gamma-1)s+r)\mathrm{d}M \,, \tag{6.37}$$

where I is the moment of inertia.

For opacity following Kramers' law, $r = 1, s = 3.5$, the integrand vanishes for $\gamma = 1.307$; if $r = 1$, $s = 3$, we find $\gamma = 4/3$ (Eddington). On the other hand, in surface layers $s \simeq -10$. Two effects can thus combine to produce the instability. On the one hand, if the opacity increases with temperature (for $T \simeq 10^4$), the maximum compression is accompanied by an increase of the opacity and heat retention which thus raises the amplitude of the oscillation (opacity mechanism or kappa mechanism). On the other hand, in helium ionisation zones the adiabatic compressibility γ decreases and approaches 1, and the integrand is positive (gamma mechanism). Regions where $\gamma \simeq 5/3$ *stabilise* the star, while surface layers destabilise the star and allow the oscillation.

Non-adiabatic Theory: The Phase Lag. The physical basis for the phase lag was found by Zhevakin (1954). It involves the deviation from adiabaticity caused by the displacement of the ionisation boundaries during the oscilla-

271

tion. We can understand this mechanism by discussing the ionisation layer of hydrogen schematically.

Regarding the transition from neutral to ionised hydrogen as a discontinuity, the incident and emergent luminosities are L_i and L_e, and because of the low mass of the hydrogen layer above the discontinuity, the latter is equal to the stellar luminosity. As the hydrogen layer has low mass it can be regarded as being in radiative equilibrium. There is then a relation between the mass of the layer and the emergent luminosity which we shall give later. The hydrogen layer retarding the escape of the luminosity L_i behaves as a low-pass filter whose time constant is the ionisation time of the hydrogen in the layer. In the Cepheid instability strip the time constant is of the order of the period, explaining the phase lag of $\pi/2$.

Because the transition layer to ionised hydrogen is very thin we can regard it as a discontinuity and use the Rankine–Hugoniot conservation equations. If v is the matter velocity with respect to the discontinuity, mass conservation gives

$$\rho v = \frac{1}{4\pi r^2} \frac{\mathrm{d}M}{\mathrm{d}t}. \tag{6.38}$$

Conservation of energy gives

$$\rho v \left(h + \tfrac{1}{2}v^2\right) - F = \mathrm{const}, \tag{6.39}$$

where h is the specific enthalpy and $F = L/4\pi r^2$ the radiative flux.

The enthalpy variation is of the order of the ionisation energy per unit mass, χ, much larger than the kinetic energy. Then we have approximately

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \frac{1}{\chi}(L_e - L_i). \tag{6.40}$$

The relation between the mass of the layer of neutral hydrogen and the luminosity L_e follows from integrating the equations of radiative equilibrium in the convective layer, adopting an opacity of the form $\kappa \sim P^{1/2}T^{1/2}$ for the typical temperatures encountered in variable stars. With effective gravity g_e we obtain for the pressure

$$P \sim g_e^{2/3}T_{\mathrm{eff}}^{-8}\left[1 - \left(1 + \tfrac{3}{2}\tau\right)^{-2}\right]^{2/3}, \tag{6.41}$$

where T_{eff} is the effective temperature. The very strong temperature dependance of the opacity leads to an asymptotic value of the pressure for large optical depth τ,

$$P \sim g_e^{2/3}T_{\mathrm{eff}}^{-8} \tag{6.42}$$

(of course, the pressure begins to grow again once the opacity starts to decrease with temperature). As the mass of the hydrogen layer is proportional

to (P/g_e) (neglecting the radius variations) we find for the mass M_H of the neutral hydrogen envelope (with $L_e \sim T_{\mathrm{eff}}^4$),

$$M_H \sim g_e^{-1/3} L_e^{-2}.$$
(6.43)

To obtain the phase lag, we assume that the luminosity L_i at the base of the discontinuity is in phase with the pressure and thus with the second derivative $\mathrm{d}^2 r/\mathrm{d}t^2$, as indicated by numerical calculations of the non-linear theory. The effective gravity includes a term due to the gravitational attraction, g, and an acceleration term $(\mathrm{d}^2 r/\mathrm{d}t^2)$,

$$g_e = g + (\mathrm{d}^2 r/\mathrm{d}t^2),$$
(6.44)

so we assume that

$$L_i \sim g_e^\alpha,$$
(6.45)

where α is of the order of $1/2$. Assuming the oscillation has the form $\mathrm{e}^{i\omega t}$ we then find that

$$i\omega\delta M = \frac{L}{\chi}\left(-\frac{1}{2}\frac{\delta M}{M} - \frac{1}{6}\frac{\delta g_e}{g} - \alpha\frac{\delta g_e}{g}\right),$$
(6.46)

where the symbol δ refers to a perturbation from equilibrium. With

$$\frac{\delta g_e}{g} = \frac{-\omega^2 \delta r}{g}$$
(6.47)

we get finally

$$\frac{\delta L_e}{L} = -i\left(-\frac{\omega^2 \delta r}{g}\right)\left(\frac{1}{2}\frac{\alpha L}{\omega M \chi}\right)\frac{1 - i(\omega M \chi)/(3\alpha L)}{1 - iL/(2\omega M \chi)},$$
(6.48)

and we see that the phase lag is controlled by the ratio $(L/\omega M\chi)$, which is equal to the pulsation period divided by 2π times the ionisation time for the hydrogen layer. For models resembling real pulsating variables such as RR Lyrae stars or Cepheids, this parameter is of the order of unity, indeed giving a phase lag of order $\pi/2$.

Of course, only a non-linear calculation can give results directly comparable with observations. The above calculation shows the role of the hydrogen-envelope mass M in determining the phase lag between the *luminosity* and the *velocity*.

6.2.3 The Period–Luminosity Relation

The first step towards the discovery of the period–luminosity relation was due to Leavitt (1908), who detected 1777 variable stars in the Magellanic clouds. She found the periods of 16 of these variables and noted in 1912 that the stars' brightness increased with period; she drew Pickering's attention

to this relation. Hertzsprung, in 1913, showed the identity of these variables with galactic Cepheids and realised the importance of the period–luminosity relation for distance determinations. This relation was the instrument used by Shapley to determine the Sun's position in the Galaxy (1918), and by Hubble in definitively establishing the extragalactic nature of M31, M33, and NGC 6822 (1925, 1926).

Much work was undertaken to find the zero point of the period–luminosity relation. The idea is to use the method of statistical parallaxes (see Chap. 1). It was necessary to wait for the first results of Baade in 1952 to show finally that the RR Lyrae stars and classical Cepheids were not a unique class. Baade could observe stars in the Andromeda galaxy (M31) down to magnitude 22.7; the RR Lyrae stars should have had magnitude 22.75 according to Shapley's 1930 period–luminosity relation. But they were invisible. This forced a correction of -1.5 magnitudes for the Cepheids. This result had already been found by Mineur (1944), but had remained unnoticed. Classical Cepheids belong to population I, while the RR Lyrae stars and dwarf Cepheids ($P < 0.26$ d) belong to populations I and II. One typically finds RR Lyrae stars in globular clusters, while dwarf Cepheids belong to the disc population. W Virginis stars, or population II Cepheids, appear to belong to the same physical family as the RR Lyrae stars, and their
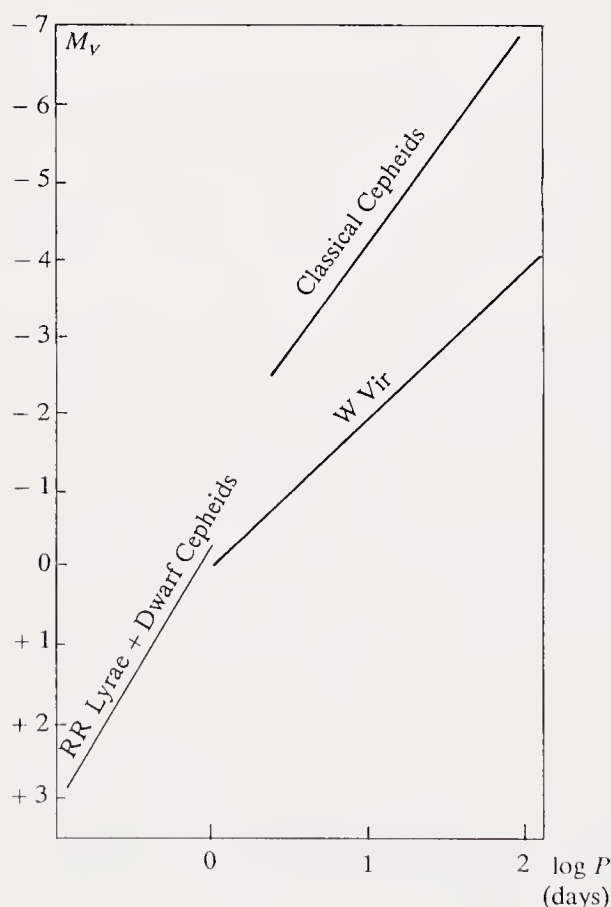
Fig. 6.8. Period–luminosity relation for period variables

period–luminosity relation extends that of the RR Lyrae stars and dwarf Cepheids with a different slope (Fig. 6.8). We adopt the expression

$$P \sim L^{\alpha} \tag{6.49}$$

with α close to 1.

The theory of the period–luminosity relation rests partly on the existence of a mass–luminosity relation and partly on the period–density relation (see below) and the simplified expression for the cause of the instability.

Pulsation occurs once the ionisation time of a helium layer is comparable to the pulsation period, $t_{ion} \gtrsim \alpha P$, $\alpha < 1$. We can use a simple self-similar model to understand how the period–luminosity relation arises. Obviously the exact solution can be found only numerically.

The condition on the ionisation time (which we have already encountered in connection with the cause of the phase lag) has the form

$$(PL/\Delta M) = \text{const}, \tag{6.50}$$

where ΔM is the mass of the neutral layer. If the ionisation occurs at a fixed temperature T_I (in reality it depends very weakly on temperature), assuming a polytropic index n for the outer layers gives

$$\Delta M \sim M^{-n} R^{n+1}. \tag{6.51}$$

The period–density relation (6.31, 32),

$$P\sqrt{\rho} = \text{const}, \tag{6.52}$$

combined with the instability condition (6.33) gives

$$L \sim R^{-n+(1/2)} M^{n-(1/2)}. \tag{6.53}$$

With the mass–luminosity relation

$$L \sim M^m, \tag{6.54}$$

we get finally

$$P \sim L^{(1/m)+(3/(2n-1))}. \tag{6.55}$$

For $m \simeq 3, n \simeq 3$ we find $P \sim L^{0.93}$, in satisfactory agreement with observation.

Actually none of the power-law relations used here is fully exact. The quantity $Q = P(\rho/\rho_{\odot})^{1/2}$ depends a little on mass and radius. From numerical results, Christy (1968) gives

$$Q = 0.022 R^{1/4} M^{-1/4} \text{ (days)}, \tag{6.56}$$

and for the fundamental

$$P \simeq 0.022 g^{-7/8} M^{-1/8}, \tag{6.57}$$

giving finally the ratio of the ionisation time t_I and the period

$$\frac{t_I}{P} = \frac{9.8 \times 10^{-5} Y}{(T_e/10^4)^{9.5}} g^{3/4} M^{-1/8} , \tag{6.58}$$

where Y is the helium concentration by mass.

These relations lead to

$$P \sim L^{11/12} , \tag{6.59}$$

$$L \sim T_{\text{eff}}^{-12} . \tag{6.60}$$

The pulsation condition $t_{\text{ion}} \gtrsim \alpha P$ is satisfied by moving from left to right in the Hertzsprung–Russell diagram. The last relation (6.60) represents approximately the blue edge of the near-vertical instability strip in the HR diagram.

6.2.4 The First Harmonic and the Structure Parameter

Homology relations immediately give $P\sqrt{\rho} = $ constant (6.31). However, variable stars are not strictly homologous (6.56). Complete model calculations reveal a parameter characterising the interior properties of the star and essential features of the pulsation.

Following Christy (1968) we denote by $V(x)$ the ratio of gravitational to thermal energy at the distance $r = xR$ from the centre (6.27). The parameter $V(x = 0.83) = V_c$ describes the properties of the Cepheids and the RR Lyrae stars. First, the quantity Q,
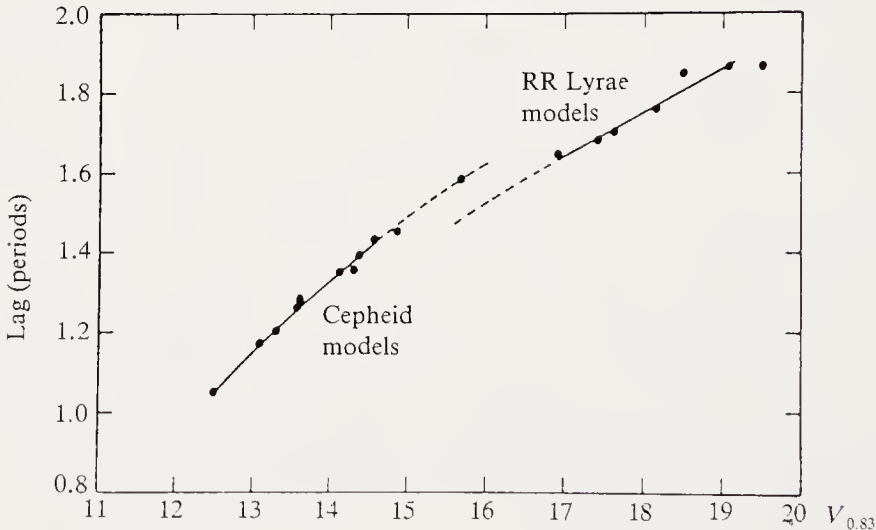
Fig. 6.9. Phase of secondary acceleration. The origin is the phase $V_{\text{rad}} = 0$ of minimum radius. The phase of secondary acceleration for different Cepheid and RR Lyrae models is plotted as a function of the structure parameter $V_c = V(x = 0.83)$ (6.53). (From R.F. Christy, *Quarterly Journal of the RAS* **9**, 15 (1968). Reproduced by kind permission of the Royal Astronomical Society)

$$Q = P\sqrt{\rho/\rho_\odot}\,,$$

is well described by the relation

$$Q = 0.222(V_c)^{-0.60}\,. \qquad (6.61)$$

The value $r = 0.83R$, close to the node of the first harmonic, thus appears as the characteristic point for the behaviour of the pulsations of Cepheids and RR Lyrae stars. This property appears clearly in the results of calculations of the motion of the star's interior. We plot on the same graph the Lagrangian variable $r(a, t)$ as a function of time for different values of the initial radius a. Once the wave propagating from the centre to the surface reaches large amplitude, it is reflected by the surface layers, propagates towards the interior, rebounds from the central regions, and emerges from the surface with a certain phase lag. This secondary wave is clearly visible in the radial-velocity curves, where it appears as a kind of secondary maximum, or a small bump.

This lag (Fig. 6.9) is an increasing function of the characteristic parameter V_c. This quantity appears to be important in fixing the properties of variable stars. It allows in particular an estimate of the star's mass (see Sect. 6.2.6).

6.2.5 The Red Edge of the Instability Strip

The blue edge of the instability strip appears to be defined by the onset of the instability in linear theory; in contrast the red edge is due to non-linear dissipative phenomena. Perturbations of the convective flux in the oscillation and dissipation due to turbulent viscosity damp the oscillations. To give some idea of the role of the turbulent viscosity ν_T, we consider its contribution to the imaginary part of the oscillation frequency,

$$\omega'_I = \frac{2}{3} \int_0^M \nu_T r^2 \left|\frac{\mathrm{d}\xi}{\mathrm{d}r}\right|^2 \mathrm{d}M \bigg/ \int_0^M r^2 \xi^2 \mathrm{d}M\,. \qquad (6.62)$$

ω'_I is positive-definite, so that the turbulent viscosity always damps. Moving to the red in the HR diagram the convective flux become more important, and turbulent dissipation becomes dominant once the characteristic convection time becomes comparable to the pulsation period. This effect defines a limit on the red side of the instability strip. The agreement with the observed limit depends on the choice of parameters describing the turbulent viscosity, implying a suitable modelling of the turbulence in the convection zone. We refer here to the discussion of Chap. 5 of the hydrodynamic of the stellar interior.

6.2.6 Masses of Pulsating Stars (Cepheids, RR Lyrae)

We see that stars on evolutionary tracks in the HR diagram enter and leave the instability strip either moving from right to left or in the opposite direction (Cepheids), or when they leave the ZAHB (RR Lyrae stars). Mass determinations for pulsating stars are simultaneously tests of stellar-evolution theory and a way of studying these stars.

RR Lyrae Stars. We have already noted (Table 6.2) that the RR Lyrae stars belong to the spherical component of the stellar population (population II). One typically finds them in globular clusters, allowing unambiguous determination of their evolutionary status. Figures 6.10a and b (Renzini 1977) show schematic HR diagrams for a typical globular cluster (e.g. M3) and the diagram calculated for stars evolving without mass loss.

The main characteristics of the HR diagram (e.g. for M3) are: (a) the horizontal branch is brighter than the stars of the knee in the main sequence by about 3.4 magnitudes; (b) the horizontal branch is quite extended in effective temperature, containing stars from various parts of the RR Lyrae instability strip; (c) the asymptotic giant branch (AGB) is close to the red-giant branch (RGB) without exceeding its peak luminosity.

Evolutionary sequences calculated at constant mass ($\dot{M} = 0$), constrained by the 3.4 magnitude difference between the horizontal-branch stars and those of the knee, have the following properties: (a$'$) the luminosity function of the RGB is in excellent agreement with observations; the end of the RGB coincides to within 1/4 magnitude with the observed peak of the RGB; (b$'$) contrary to observation, the horizontal branch is only populated to the red side of the instability strip (the first question mark on the figure); (c$'$)
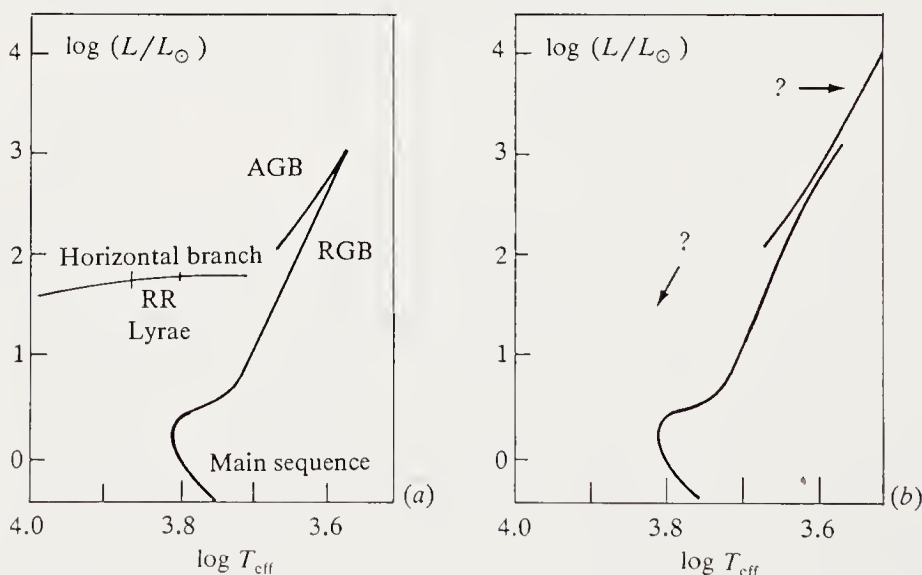
Fig. 6.10. (a) HR diagram for a typical globular cluster. (b) Diagram calculated for stars evolving without mass loss. There are no stars to the red side of the horizontal branch; the asymptotic giant branch extends too high in luminosity

the asymptotic giant branch (AGB) extends 2.5 magnitudes above the end of the red-giant branch.

The condition on the magnitude difference between the knee stars and those of the horizontal branch gives an initial mass of $0.85\,M_\odot$ for the stars reaching the horizontal branch; mass loss of $0.2\,M_\odot$ leads to a mass of $0.65\,M_\odot$, the mass given by the pulsation properties. Assuming a mass loss rate (Reimers, see Chap. 4)

$$\dot{M} = -4 \times 10^{-13}\,\eta\,\frac{L}{gR}\,(M_\odot/\text{year}) \qquad (6.63)$$

we find $\eta \simeq 0.4$. The discussion by Renzini shows that this value of η is fixed within a narrow interval ($\pm\,10\,\%$).

The same mass loss rate limits the maximum luminosity of the AGB. We may conclude that the disagreements (b') and (c') are both explained by mass loss.

The extent of the horizontal branch implies the presence of a second parameter (stellar mass? degenerate helium core mass? core rotation?). We should add to this the effect of metallicity. The horizontal branch is bluer for lower metal abundance, in agreement with the results of Faulkner for population I stars (Chap. 3). We note that this metallicity effect implies a tight constraint on the mass loss. Clusters with low Z would have a "red" horizontal branch for $\Delta M < 0.1\,M_\odot$; metal-rich clusters would have a "blue" horizontal branch for $\Delta M > 0.3\,M_\odot$.

W Vir Stars (Population II Cepheids). As we have seen, these are pulsating variable stars with period comparable to those of the Cepheids. But these stars have much smaller radii, clearly showing that their masses are very different, and the stars have become unstable at a quite different stage of their evolution. Models give a mass of $0.9\,M_\odot$ for W Vir.

These stars are found in metal-poor clusters (small Z) with a "blue" horizontal branch. They are stars which have left the peak of the AGB and are now crossing the instability strip.

Cepheids. The various methods of determining masses for Cepheids are in relatively serious disagreement. The main disagreement is that between the mass deduced from evolution and that deduced from the pulsation. The ideal method would be to find the mass of a Cepheid in a binary. Unfortunately binaries containing Cepheids are too wide to allow an orbital solution. In two cases radial-velocity measurements lead only to the mass ratio of the two components (S Mus, V636 Sco).

We may appeal to the following methods:

(a) *Evolution.* During the passage through the instability strip a star of mass M has a luminosity L. Measurement of L (which involves the calibration of the distance scale) gives the mass.

(b) *Pulsation*. Measuring the luminosity and the effective temperature gives the radius. Using the relation $P(\rho/\rho_\odot)^{1/2} = Q$ gives the mass.

(c) *The bump in the light curve*. As we have seen, the phase of the bump in the light curve gives the structure parameter V_c, and thus the parameter $Q = P(\rho/\rho_\odot)^{1/2}$. A radius measurement gives the mass.

(d) *The Baade–Wesselink method*. This gives a value of the radius independent of the luminosity and effective temperature. Using $P(\rho/\rho_\odot)^{1/2} = Q$ we derive the mass.

Obviously there are numerous problems in interpreting the observational data: calibration of the distance scale, the relation between the radial velocity and the derivative $\dot{R}$, and the relation of the effective temperature to observed photometric and spectroscopic quantities. From an observational viewpoint the two main uncertainties are the value of the distance modulus of the Cepheids and the value of the effective temperature; theoretically the main uncertainties are the opacity and chemical composition. Comparison of the mass M_{Ev} deduced from the evolutionary track, and the theoretical mass M_{Th} found from the mass–luminosity relation associated with the relation $P\sqrt{\rho} = $ constant, and the mass M_{puls} found from various properties of the pulsation allow us in principle to decide the importance of the distance scale and the mass loss rate. The result of the comparison depends essentially on the distance modulus; at present the basis of the determination of distance moduli is the distance to the Hyades. Too small a distance to the Hyades leads to an underestimate of the radii of Cepheids and the expression for M_{puls} gives too small a mass, by about 40 %. Increasing the distance modulus by $\Delta m = 0.26$ brings the mass M_{puls} into line with M_{Ev} and M_{Th} (see Sect. 3.7).

The evolutionary position of the Cepheids indicates that they are helium-burning stars, on their way towards the AGB (Fig. 3.25), and in the main are crossing the instability strip from right to left. The only possible epoch of mass loss is that of the very brief passage along the giant branch, implying that very little mass is lost. This theoretical argument favours a higher distance modulus. However, following a revision (1974) the smallest distance modulus appears to be the correct one (1984). The lowest of the masses M_{puls} for $P \gtrsim 6$ d appears to be in agreement with mass determinations for Cepheids in binaries ($M \simeq (5\pm1)\,M_\odot$). The frequency of Cepheids depends on their mass, through the initial mass function; a little larger for lower masses appears to accord better with the observed statistics. The origin of the disagreement should be sought in the theoretical models, and would be resolved by crossing of the instability strip at a higher luminosity than present theory predicts. The explanation seems to rest on the value of the opacity (see Sect. 3.7.2).

6.2.7 Long-Period or Red Variables

These variables have large amplitude and long period and are essentially of type M (to which we should add types R and S). It is usual to place in this group the Miras, having periods of 300 days (4566 stars in the Kukarkin catalogue), and semi-regular variables of types K5 to M5, with periods between 30 and 1000 d (9221 objects). They are distinguished from the Miras by their amplitudes, which are less than 2.5 magnitudes.

Miras show strong emission lines at certain phases, the intensity of the lines increasing with the amplitude of the light curve. Observation of Miras in globular clusters shows a relation between the period and the chemical composition (the [Fe/H] ratio) (Fig. 6.11). This supports the idea that Miras of a given period constitute a homogeneous group in age and chemical composition.

Pulsation Mode and Mass. For stars whose energy source is core helium burning (AGB), linear theory gives the quantity $Q = P(\rho/\rho_\odot)^{1/2}$, (with Q expressed in days) as a function of mass for the fundamental and first harmonic. Conversely, given a calibrated distance to the Miras, it is possible to find an empirical value of Q by making an assumption about the mass. The distance determination rests on the hypothesis of Eggen (1975) that the Miras have the same mean luminosity as the non-Miras of the red-supergiant branch.

The value of Q as a function of period (Fig. 6.12) for different assumed masses (0.8, 1, and $2\,M_\odot$) (dashed curves), and the theoretical value of Q for the fundamental and the first and second harmonics for various masses, give the following conclusion: Miras of the old disc population, with periods between 250 and 500 days, pulse in the first harmonic. They are on the AGB, and their mass is that expected for old stars which have lost mass.
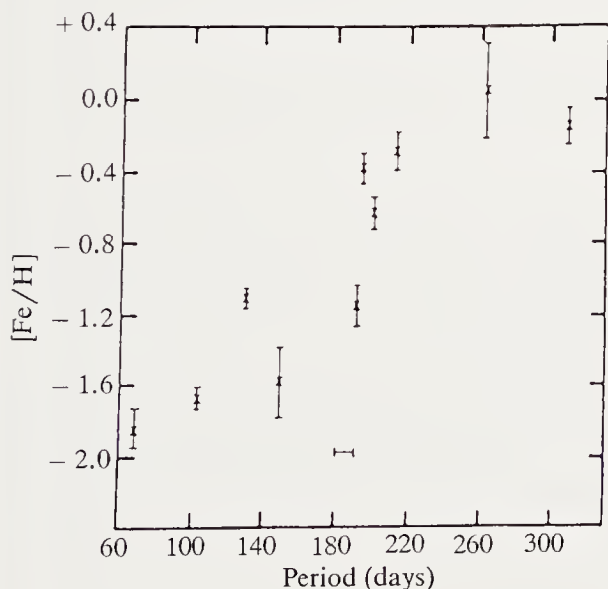
Fig. 6.11. Relation between the period and the iron abundance in Miras from several clusters. (A mean value of the period has been taken for each cluster). (From M.W. Feast, in *Physical Processes in Red Giants*, ed. by I. Iben and A. Renzini, Reidel 1981, p. 198. Reproduced by kind permission of Kluwer Academic Publishers)
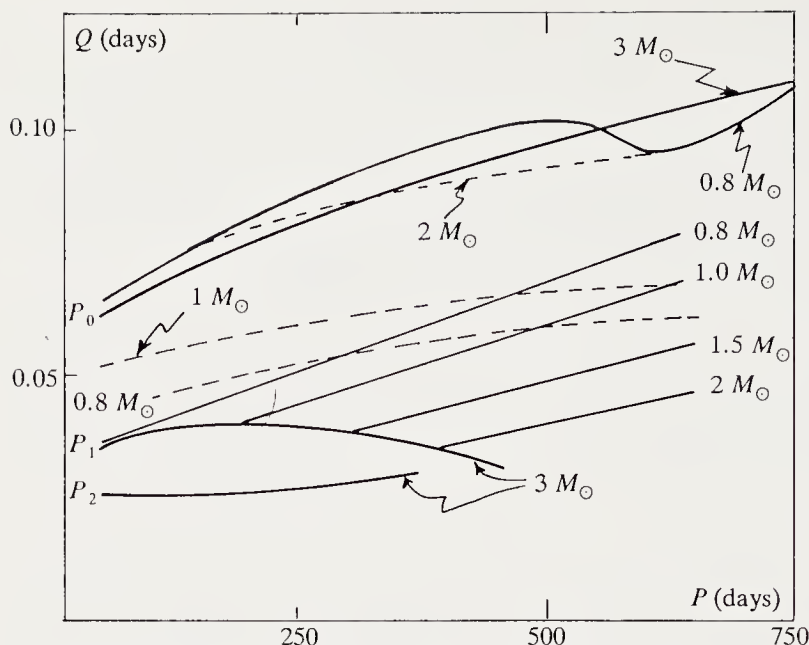
Fig. 6.12. Red variables. The relation $Q = P(\rho/\rho_\odot)^{1/2}$ for the fundamental and first harmonics, as functions of period. The *dashed curves* are deduced from observation for various assumptions about the mass. The *continuous curves* are theoretical ones for the fundamental and first two harmonics for various masses. Comparison of theoretical and observed curves favours the idea of oscillation in the first harmonic of stars which have lost mass. (From P.R. Wood, in *Physical Processes in Red Giants*, ed. by I. Iben and A. Renzini, Reidel 1981, p. 208. Reproduced by kind permission of Kluwer Academic Publishers)

The Evolutionary Phase. Calculations of the evolution along the asymptotic giant branch (AGB) reveal the following properties:

— The stars evolve upwards along the AGB while losing mass.
— At a certain luminosity the stars become unstable and oscillate in the first harmonic.
— For a larger luminosity the fundamental oscillation begins to dominate. There is a relaxation oscillation which after a few periods leads to the ejection of the envelope and the formation of a planetary nebula.

This model predicts an instability region in the $(M, \log L)$ plane. The position of the instability region given by the theoretical models does not agree with observation (which requires a higher luminosity), but this could be due to the poorly known molecular opacities for the outer envelope and to the difficulty of treating the convection zone.

Theoretical models of helium-shell-burning stars predict runaway burning (the *helium flash*) accompanied by an increase of the surface luminosity above the steady value. These episodes last about 600 years. The radius change associated with the luminosity increase causes a rapid change of period.
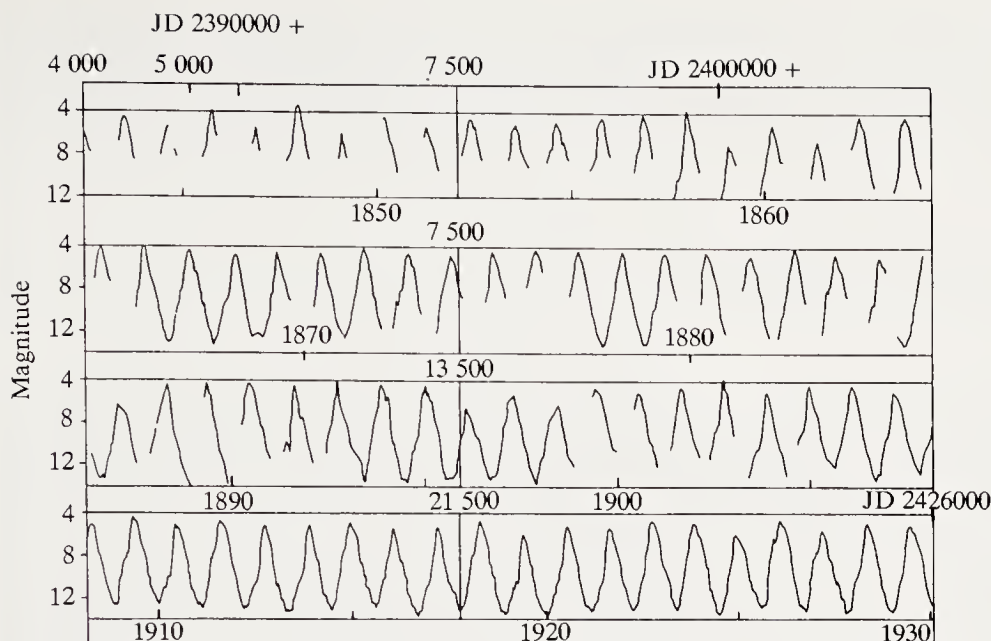
Fig. 6.13. Light curve of χ Cyg. (From J.M. Perdang, in *Chaos in Astrophysics*, ed. by J.R. Buchler et al., Reidel 1985, p. 32. Reproduced by kind permission of Kluwer Academic Publishers)

Secular period variations have long been known in R Hya, R Aql, and W Dra. In R Aql, $(\partial P/\partial t) \simeq 0.485$ days/year, about 10^7 times greater than for RR Lyrae. Combining Eggen's period–luminosity relation for Miras with the secular period variation gives an (L, t) diagram in good agreement with theoretical predictions of the helium flash.

Chaotic Behaviour. Besides secular variations, the light curves can show significant changes in the phase of maximum (e.g. χ Cyg, Fig. 6.13). The behaviour of these oscillations resembles that of dynamical systems. Recent developments in the study of non-linear systems appear to suggest physical causes for the preference for chaotic rather than limit-cycle behaviour of the oscillations (see Sect. 6.4).

6.2.8 δ Scuti Stars

General Properties. These are variable stars with very small amplitudes (a tenth of a magnitude or less) close to the main sequence, at the intersection with the Cepheid instability strip. The prototype is δ Scu, with an amplitude of $\Delta m = 0.194$ magnitudes. The spectral type of the δ Scu stars is A3–F6, and their chemical composition is normal population I.

δ Scuti and Am (Metallic-Line) Stars. These two classes occupy the same region of the HR diagram. The difference between them appears to be related to their chemical composition. In the Am stars, elemental separation

by gravity or radiation pressure produces spectral anomalies; moreover, helium is separated too. Once the helium content of the outer layers falls below 10 %, the kappa-mechanism becomes ineffective and the star can no longer pulse.

6.3 Other Variables

6.3.1 β CMa Stars

Stars in this group have spectral type B0.5 to B5 (15 000 to 25 000 K); their periods lie between 3 and 7 hours. The light-curve amplitude is small, less than 0.1 magnitudes, but they have large radial-velocity amplitudes, between 5 and 50 km s^{-1} or even more.

These stars have several periods, and beats between them. They are slow rotators, with equatorial velocities between 15 and 60 km s^{-1}, while normal stars of the same spectral type have typical equatorial velocities of 165 km s^{-1}. Their oscillations are *non-radial* (see Sect. 6.5) with significant secular variations.

Particularly remarkable is α Vir (B1V). The amplitude of oscillation decreased steadily from 1968 to 1972 and the star has now almost stopped pulsing. It is a 10 $M_\odot$ star (in a binary) with standard chemical composition.

The cause of the oscillations of β CMa stars is still unknown. It is possible that the observed oscillations are modulated by the rotation.

6.3.2 White Dwarfs

Pulsating white dwarfs are divided into two groups.

ZZ Ceti Stars. These are pulsating DA white dwarfs with a luminosity of about $10^{-2} L_\odot$ and an effective temperature between 12 000 and 10 000 K. More than 20 such stars are known at present, all close to the Sun.

The observed periods, in the range 100–1200 s, are longer by two orders of magnitude than the period of the fundamental. The simplest assumption is that they pulse in non-radial g modes (see below). All the ZZ Ceti stars show several periods, with light-curve amplitudes lying between 0.003 and 0.3 magnitudes.

The ZZ Ceti stars have masses between 0.4 and 0.8 $M_\odot$; they probably have a carbon–oxygen core surrounded by a helium layer containing about 1 % of the mass and surrounded by a hydrogen layer, whose mass is between 10^{-14} and 10^{-4} times the total mass. The driving of the g modes is mainly due to hydrogen ionisation.

DB and DO White Dwarfs. These stars have almost no hydrogen and may drive non-radial g modes through the ionisation of HeII. At present only a few such stars are known, with significant differences. The power spectrum

given by Fourier analysis reveals many periods between 2 and 15 min. As for the DA white dwarfs, it appears that linear theory is inadequate for explaining the observed pulsations. A third instability band has been found for higher-temperature DO stars.

6.4 Variable Stars and Dynamical Systems

6.4.1 Variable Stars as Dynamical Systems

Irregular behaviour is observed in many variables (amplitude, phase of maximum, variation of Fourier amplitudes). This suggests that the description of pulsating stars by linear equations may be incomplete. Dynamical systems involve non-linear restoring forces and dissipative effects. There are no analytical solutions for such systems. They have a characteristic feature: although described by regular equations they can have solutions which behave irregularly in time. Irregular oscillations of a system, or turbulence, show the same features. In astrophysics such behaviour is seen in irregular pulsing of Miras, irregular red variables, and ZZ Ceti stars, and turbulence in convection zones.

Here we give an elementary introduction to dynamical systems and show how we can relate an astrophysical phenomenon (irregularity of variables) to a simple model of a dynamical system.

The equations governing a dynamical system are deterministic; the system may nevertheless have unpredictable behaviour. Already in 1892 Poincaré showed that solutions starting from infinitesimally separated points might nevertheless diverge exponentially from each other. The study of dynamical systems has developed rapidly in recent years because of the great variety of problems which can be described in this way.

6.4.2 The One-Zone Model

We represent the star as having a static core producing a constant energy flux (because of the very small oscillation amplitudes in central regions), surrounded by a hydrogen layer of constant density and uniform temperature, surrounded by a radiative region. In Miras we have a carbon–oxygen core with a helium-burning shell; the outer parts of these stars are immense convection zones in which ionisation equilibrium brings the adiabatic compressibility below 4/3. We can show that for an equation of state with this property the equation of hydrostatic equilibrium has three solutions; two are *dynamically stable*, corresponding to two minima of the potential, and one solution is *dynamically unstable*, corresponding to a maximum of the potential.

Non-linear dynamical systems with at least three fixed points (two stable points and one unstable saddle point) have trajectories which are extremely

sensitive to small differences in physical parameters. In the absence of dissipation, the stable points lead to limit cycles while the unstable point repels the trajectories exponentially. We can describe the situation in physical terms as follows.

A star in dynamically stable hydrostatic equilibrium but *thermal disequilibrium* will evolve away from this equilibrium on a thermal timescale. *Thermal disequilibrium* means that the right-hand side of the thermal-balance equation

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{1}{T}\left(q - \frac{\partial L}{\partial m}\right)$$

does not vanish, where q represents local energy sources and $(\partial L/\partial m)$ the divergence of the flux. As a result the specific entropy S depends on time. If there is sufficient energy available, and the conservation of thermal energy pulls the star in the right direction, two processes now compete: thermal disequilibrium drives the star from one hydrostatic-equilibrium state to another, and the dynamical instability near the unstable-equilibrium point drives the star towards one or other of the stable hydrostatic equilibria. A great variety of oscillatory motions are then possible, including relaxation oscillations. In particular, if the dynamical and thermal timescales are comparable, irregular variations can occur.

We write the equations for a one-zone model. The mass variable m is bounded above at $R(t)$. We let S be the specific entropy of the zone, which surrounds a fixed core of mass $M_c = M - m$, radius R_c, and luminosity L_c. Then the zone's behaviour is governed by the equations

$$\frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = g(R, S), \qquad (6.64)$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \varepsilon h(R, S), \qquad (6.65)$$

where g is the total acceleration,

$$g = -\frac{GM}{R^2} + 4\pi R^2 \frac{P(\rho, S)}{\Delta m}, \qquad (6.66)$$

while the quantity Δm differs from m and gives the relation between the uniform pressure in the zone and its effective inertia. h is the entropy production rate,

$$h = \frac{(L_c - L)}{mT}, \qquad (6.67)$$

and ε is a scale parameter, equal to the ratio of the dynamical and thermal timescales, the former appearing through the equation of motion and the latter through h.

The functions $P(\rho, S), T(\rho, S)$ are given by the equation of state taking account of the ionisation change during the motion, and ρ is given by mass conservation for the one-zone model, $4\pi\rho(R^3 - R_c^3) = 3m$.

The luminosity L at the outer boundary of the zone is fixed by radiative transport across a layer of mass $\Delta m'$,

$$L = \frac{(4\pi R^2)^2}{\kappa(\rho, T)\Delta m'} \frac{4\sigma}{3} (T^4 - T_*^4),$$ (6.68)

where σ is Stefan's constant, κ the opacity, and T_* the temperature (assumed constant) of the outer radiative zone. Finally the opacity is given by a power law:

$$\kappa(\rho, T) = \kappa_0 \rho^k T^n.$$ (6.69)

At constant entropy we can obtain the total acceleration as a function of the radius (Fig. 6.14) for a particular choice of model parameters (see figure caption). The fixed point is described by $g = 0, L = L_c$; it corresponds to the curve $S = S_3$ in Fig. 6.14. In the interval $S_{crit1} < S < S_{crit2}$ the hydrostatic-equilibrium equation has three solutions, with the intermediate solution being unstable, as perturbations away from the equilibrium give the same sign of radius and acceleration variations.

Plotting the hydrostatic-equilibrium radius as a function of the entropy in the thermal equation, we get the entropy production rate as a function of specific entropy (Fig. 6.15). We note that on the branch of minimum radius R_{min}, the entropy production rate is positive and drives the zone towards the maximum radius, while on the branch of maximum radius R_{max} the
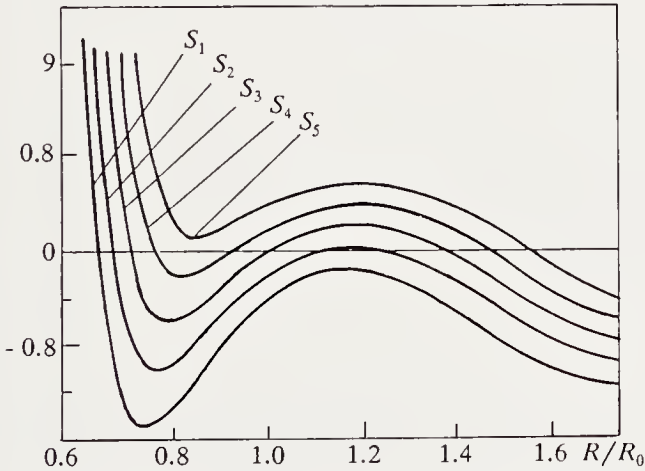
Fig. 6.14. Total acceleration as a function of radius in a one-zone model. The ionisation of He$^+$ fixes the behaviour of the pressure $P(\rho, S)$ and determines the form of the curves, with a maximum and a minimum. The curve $S = S_3$ is chosen so that $g = 0$ for $L = L_{core}$ (see text). (From J.R. Buchler and O. Regev, *Astrophys. J.* **263**, 312 (1982). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1982 The American Astronomical Society)
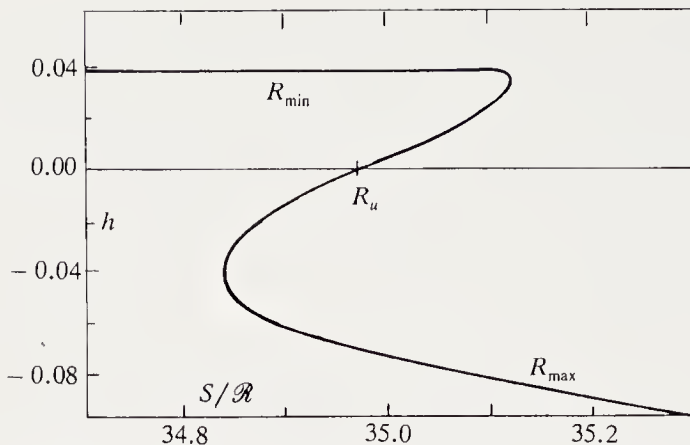
Fig. 6.15. Entropy production rate as a function of the specific entropy in a one-zone model

entropy production is negative and drives the zone towards the minimum radius.

The extremal solutions of hydrostatic equilibrium combine dynamical stability and thermal instability; the intermediate solution is simultaneously dynamically unstable and thermally secularly unstable.

6.4.3 The Moore–Spiegel Model (1966)

The Moore–Spiegel model is the first study of a non-linear one-zone model, having the characteristics of an unstable physical system. We can find the equation governing the Moore–Spiegel dynamical system by expanding the equation of motion and the energy equation in the neighbourhood of the intermediate solution defined in the last section. Setting $x = R - R_0, \lambda = S - S_0$, we have

$$x'' + A_1\lambda + (B_0 + B_1\lambda)x + (D_0 + D_1\lambda)x^3 = 0. \tag{6.70}$$

The non-derivative term is of algebraic degree 3 in x. It has the behaviour seen in Fig. 6.14. For $\lambda = 0$, there are 3 fixed points, $x = 0$ and $x = \pm(-B_0/D_0)^{1/2}$, corresponding respectively to the intermediate dynamically unstable solution and the two hydrostatically stable solutions.

The energy equation thus reduces to

$$\dot\lambda = -K\varepsilon x, \tag{6.71}$$

where K is a measure of the departure from adiabaticity. Reducing the equation of motion to lowest order in λ, we get the Moore–Spiegel equation:

$$\ddot x + (B_0 + 3D_0x^2)\dot x - K\varepsilon x = 0. \tag{6.72}$$

Examination of the adiabatic solutions facilitates an understanding of the non-adiabatic ones. We have adiabaticity for $K = 0$. Then we have
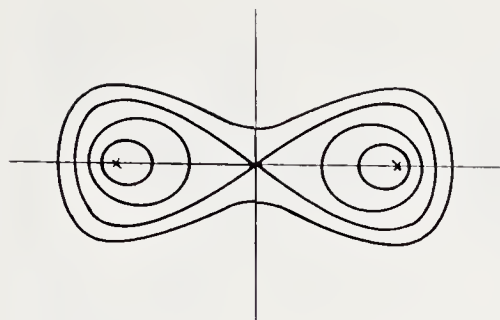
$$\ddot{x} + (B_0 + 3D_0 x^2)\dot{x} = 0 \,, \tag{6.73}$$

which can be integrated to give

$$\ddot{x} + (B_0 x + D_0 x^3) = b \,. \tag{6.74}$$

This equation has 3 fixed points. It suffices to examine the case $b = 0$. We can represent the motion in the phase space $(\dot{x}, x)$. Near $x = 0$ we have a saddle point, which repels; near $x = \pm(-B_0/D_0)^{1/2}$ we have two minima. We then have

$$\dot{x} = \pm\sqrt{E - B_0 x^2 - \frac{D_0 x^4}{2}} \,, \tag{6.75}$$

which we can write as

$$\dot{x} = \pm\sqrt{E + V} \,, \tag{6.76}$$

where V is the potential $-B_0 x^2 - D_0 x^4/4$. Depending on the value of E, we get the various curves of constant energy shown in Fig. 6.16.

When dissipation is introduced ($K \neq 0$) we find a new property of the trajectories in phase space: period doubling.

For $K = 0$ we have a periodic orbit, of period T; for $K \neq 0$ and small, the orbit is still periodic, but closes only after a period $2T$ (Fig. 6.19). This is the first bifurcation. As K is increased, a new bifurcation appears. The orbit is still periodic but we now have periods $2T$ and $4T$, with the orbit closing after $4T$. Increasing the parameter K gives more bifurcations, with the distance $(K_{n+1} - K_n)$ from one bifurcation to the next decreasing. There is a critical value $K_{crit} = K_\infty$ above which the orbits are no longer regular at all; the dynamical system has become chaotic.

6.4.4 A Schematic Red Variable

We return to the system (6.64, 65), where the only parameter is ε, the ratio of the dynamical and thermal timescales.

Depending on the value of the parameter ε, we get different types of oscillations (Fig. 6.17). For small ε there are many dynamical oscillations per thermal cycle ($1/\varepsilon$ dynamical oscillations per thermal relaxation cycle)
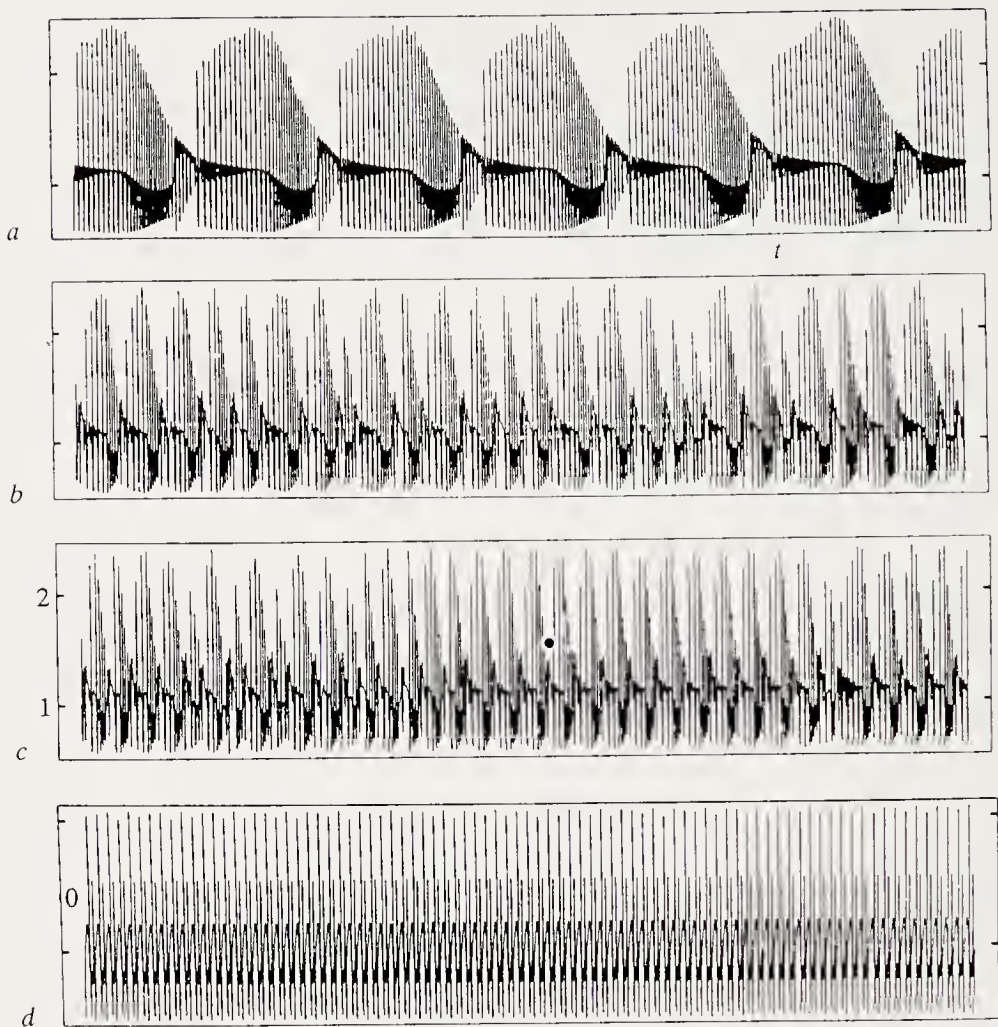
Fig. 6.17. Radius as a function of time for various values of the parameter ε: (a) $\varepsilon = 0.04$; (b) $\varepsilon = 0.15$; (c) $\varepsilon = 0.205$; (d) $\varepsilon = 4.0$. The timescale of an oscillation in (a) is of the order of a year. (From J.R. Buchler and O. Regev, *Astrophys. J.* **263**, 312 (1982). Reproduced by kind permission of The Astrophysical Journal, published by The University of Chicago Press; © 1982, The American Astronomical Society)

(case (a), $\varepsilon = 0.04$). In the phase diagram (Fig. 6.18) the trajectory winds around two fixed points. This can be understood by reference to the figure giving the entropy production rate as a function of the entropy of the layer (Fig. 6.15). The short period of the dynamical oscillation shows that the star is close to hydrostatic equilibrium for most of the time. Starting from the upper branch, the star is thermally unstable. It evolves, close to hydrostatic equilibrium, along the curve $h(S)$ (Fig. 6.15), until the maximum entropy is reached. At this point, hydrostatic equilibrium disappears. The system then passes into a phase of dynamical oscillations, oscillating in the potential well centred on the maximum-entropy branch (lower branch of the $h(S)$ curve). This oscillation is damped and the layer evolves with decreasing entropy towards the critical entropy minimum. Here hydrostatic equilibrium once again disappears and there follows a phase of dynamical oscillation around
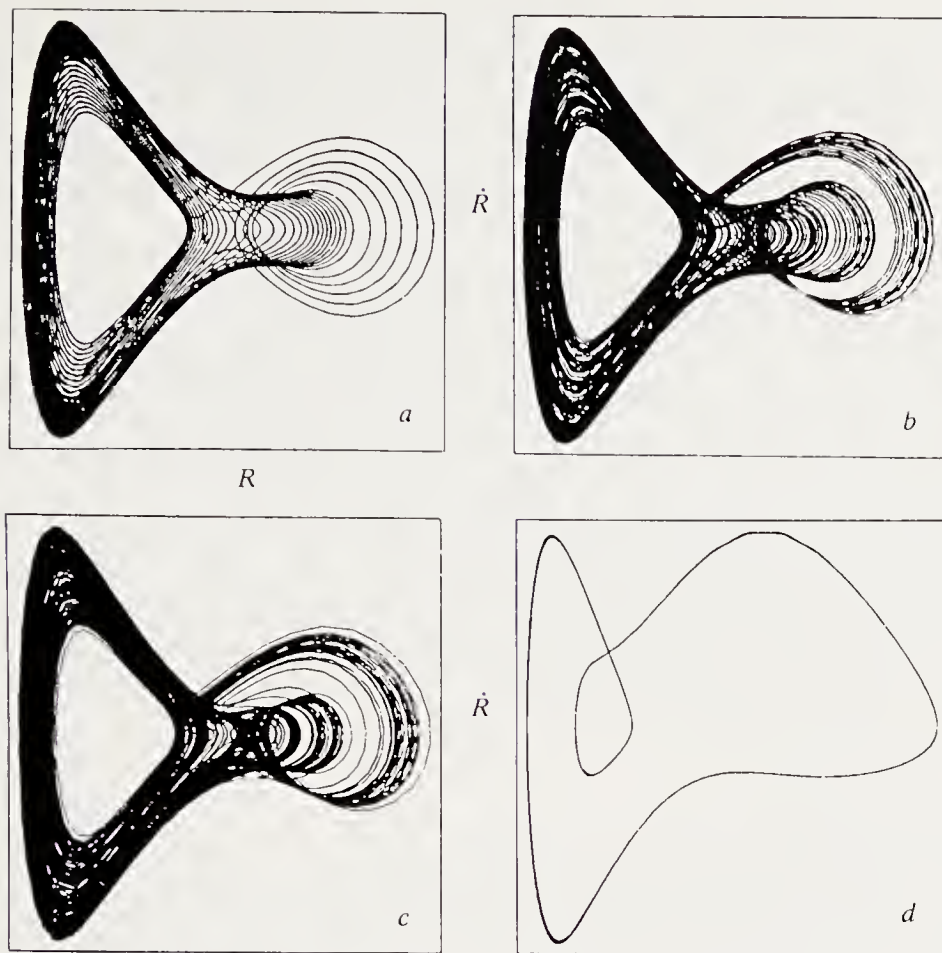
Fig. 6.18. Phase diagram for the oscillation of Fig. 6.17, with the same values of ε. In (a) the system is dominated by dynamical oscillations around one or other of the hydrostatic equilibria; in (b) and (c) the oscillations are irregular; in (d) there is a limit cycle. (From J.R. Buchler and O. Regev, *Astrophys. J.* **263**, 312 (1982). Reproduced by kind permission of The Astrophysical Journal, published by The University of Chicago Press; © 1982 The American Astronomical Society)

the potential minimum associated with the upper branch of the $h(S)$ curve (Fig. 6.15).

The one-zone model can thus have two oscillation modes with different amplitudes, according as the damped oscillation occurs around the entropy maximum or minimum. It is tempting to see here the possibility of two stars of the same mass and same state of evolution having different oscillation amplitudes, corresponding to two different epochs in different regimes. This would then have the appearance of hysteresis.

For smaller values of the ratio ε the two oscillation regimes would last longer (having more dynamical oscillations per thermal relaxation time). With increasing ε, the number of dynamical oscillations per thermal cycle decreases, and the oscillations suddenly become aperiodic for a certain value of ε. There has not been a complete examination of the model and the presence of a sequence of period-doublings has not been verified. For the values

of ε chosen by Buchler and Regev (1982) there are aperiodic oscillations for the case (b) ($\varepsilon = 0.15$) and (c) ($\varepsilon = 0.205$). For larger values of ε the system is dominated by the thermal timescale, and we find another periodic system, with a limit cycle. In case (d) ($\varepsilon = 4.0$) the limit cycle makes 2 loops around the 2 hydrostatic-equilibrium points. For larger values of ε the limit cycle makes a single loop around the two fixed points.

This one-zone model is clearly very crude, and in particular does not describe the movement of the ionisation front through the stellar envelope during the pulsation. However, the main effect of ionisation in lowering the adiabatic compressibility below 4/3 is well represented. From a physical point of view the outer convective layer of low-mass supergiants has density and entropy profiles which are very flat, so that modelling it as a region of constant density and entropy is not absurd.

It is not really possible to identify the various modes of oscillation of the one-zone model with the radius variations of a red variable, regular, or irregular. It is however very suggestive to compare the aperiodic variations of the one-zone model with those of the regular or irregular red variables; we might attribute Mira oscillations to an irregular oscillation which is close to a limit cycle.

6.4.5 n-Zone Models

When a stellar envelope is represented by n layers, each layer is characterised by three quantities, radius, velocity and entropy, so that the discretised system has a $3n$-dimensional phase space. To study this system we start from an envelope in hydrostatic equilibrium, and then examine its hydrodynamic behaviour. Buchler and Kovacs (1987) did this for a sequence of envelopes of mass $M = 0.6\,M_\odot$, with chemical composition $X = 0.745$, $Z = 0.005$, luminosity $L = 400\,L_\odot$, and effective temperatures from $\log T_{\mathrm{eff}} = 3.71$ to $\log T_{\mathrm{eff}} = 3.64$ (Table 6.5). The heat transfer is treated in a simplified fashion and reduced to radiative transfer. The number of layers is $n = 60$.

The results of this calculation are extremely rich and simulate a sequence from the W Vir stars to the RV Tau and semiregular variables.

The two-dimensional representation (velocity as a function of radius) already shows period-doublings (models a, b, c, d) and chaos (models e and f) (Fig. 6.19).

The best insight is given by the *Poincaré representation*. Given a trajectory in an N-dimensional space, we cut it by a surface of $N - 1$ dimensions. A periodic trajectory will return to the same point after a certain number of intersections. Chaos produces a uniform distribution of points. In the present case we construct a 3-dimensional trajectory by taking the radius R of one of the layers (here $n = 55$) at the points $x = t, y = t + k, z = t + 2k$, where k is chosen to correspond approximately to a phase shift of order $\pi/2$. The intersecting surface is the plane $x = y$. About 100 intersection points are continuously distributed on a curve.

Table 6.5. Properties of model red supergiants

Models (1)	$\log T_{\text{eff}}$ (2)	P_0 (3)	σ_0 (4)	σ_1 (5)	P_0^{NL} (6)
a	3.71	9.730	0.034	-0.033	11.587
b	3.69	11.689	0.031	-0.026	26.196
c	3.67	14.043	0.029	-0.019	59.040
d	3.66	15.405	0.028	-0.015	125.784
e	3.65	16.913	0.027	-0.012	...
f	3.64	18.587	0.026	-0.009	...

Column (3): period of the fundamental (in days) calculated from a linear analysis

Column (4): growth rate of the fundamental

Column (5): growth rate of the first harmonic

Column (6): period of the fundamental (in days) given by non-linear calculation
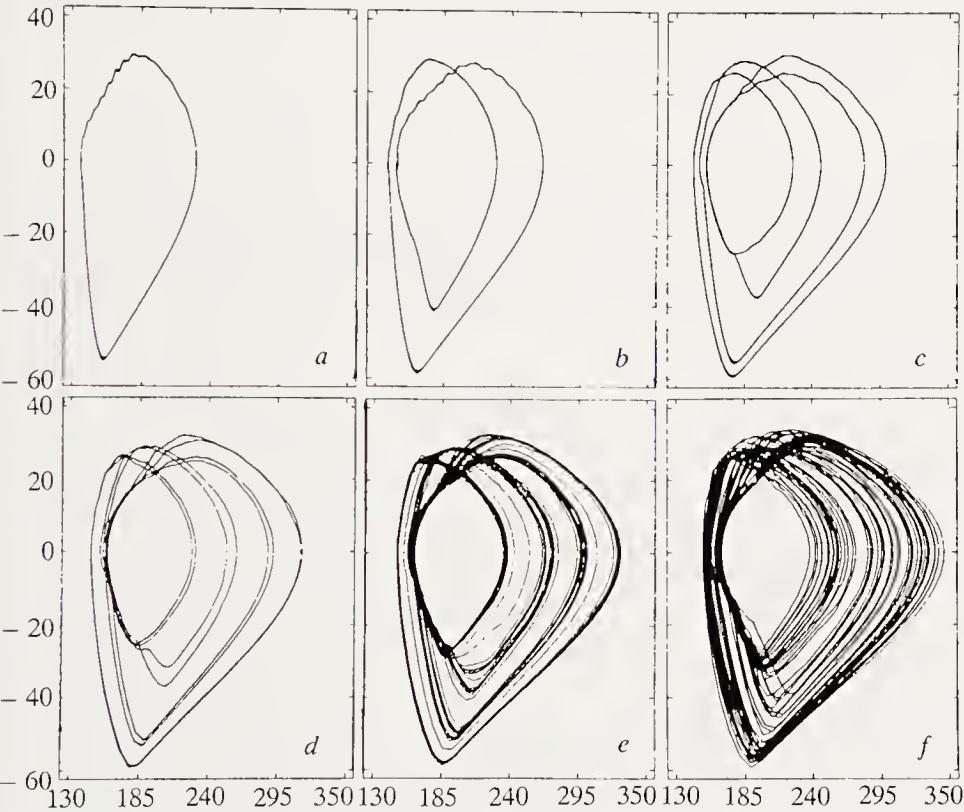
Fig. 6.19. Two-dimensional phase-space representation (R, v of layer $n = 55$) of the envelope models (Table 6.5) of Buchler and Kovacs. We note the period-doublings (a, b, c, d) before chaos (e, f). (From J.R. Buchler and G. Kovacs, *Astrophys. J.* **320**, 257 (1987). Reproduced by kind permission of the Astrophysical Journal, published by the University of Chicago Press; © 1987 The American Astronomical Society)

The light curve of model (f) resembles that of an RV Tau star remarkably closely. A 3-dimensional representation is thus enough to characterise

the main properties of a model with 180 dimensions. This suggests that this complex dynamical model can be represented by a small number of coupled differential equations. But the nature of the physical variables which would be governed by this system remains to be clarified: this would be an important step towards understanding the irregularities of variable stars and the passage from periodicity to chaos.

6.4.6 White Dwarfs of ZZ Ceti Type

Fourier analysis of the light variations of white dwarfs sometimes shows chaotic oscillations (e.g. GD 66) or the presence of subharmonics indicating the transition to chaos. These stars appear to have the properties of dynamical systems, and the one-zone model can in principle explain their properties. In particular, small differences of the parameter ε (the ratio of dynamical and thermal timescales) cause large changes in the character of the pulsations.

6.5 Non-radial Oscillations

6.5.1 Modes of Oscillation

The linearised equations of motion show the existence of normal modes of oscillation of a spherical star with amplitudes ξ proportional to spherical harmonics $Y_l^m(\theta, \phi)$, with $l = 0, 1, 2 ...$, and $-l \leq m \leq l$ and frequency σ,

$$\xi \sim Y_l^m(\theta, \varphi)e^{i\sigma t} = P_l^m(\cos\theta)e^{im\varphi}e^{i\sigma t} . \tag{6.77}$$

For a non-rotating star there is degeneracy in m, i.e. the frequency is independent of m.

One usually denotes the order of the radial mode by n, and the degree by l. In the presence of rotation or a magnetic field the degeneracy in m is broken. Radial modes correspond to the case $l = 0$.

There are two main problems associated with non-radial oscillations. The first is to identify the mode; the second is to find a way of stimulating the oscillation.

The restoring force governs the oscillation. The pressure changes occurring in radial oscillations act here too. But as for waves on the sea, gravity is also a restoring force. We distinguish between pressure modes (p modes) and gravity modes (g modes). Radial modes are called f modes.

Two characteristic local frequencies govern the oscillation properties of the medium: (a) the Lamb frequency L_l, associated with the sound speed and the horizontal wavelength and given by

$$L_l^2 = \frac{l(l+1)c^2}{r^2} , \tag{6.78}$$

where c is the sound speed; pressure modes correspond to frequencies above L_l, and gravity modes to frequencies below. The second is the Brunt–Väisälä frequency N, which is the adiabatic oscillation frequency of a bubble of gas,

$$N^2 = g \left(\frac{1}{\Gamma_1} \frac{\mathrm{d} \ln P_0}{\mathrm{d}r} - \frac{\mathrm{d} \ln \rho_0}{\mathrm{d}r} \right), \tag{6.79}$$

where Γ_1 is the coefficient of adiabatic compressibility and P_0 and ρ_0 the unperturbed pressure and density. The frequencies of propagating acoustic pressure modes necessarily exceed N, and those of propagating gravity waves are necessarily less than N.

For a given frequency σ, these two frequencies define the regions of the star where pressure or gravity modes can propagate. Figures 6.20a and b give an example of a propagation diagram.

If $\sigma > L_l$ and $\sigma > N$, we have a region where pressure modes can propagate; if $\sigma < L_l$ and $\sigma < N$, gravity modes propagate. If $L_l < \sigma < N$, pressure modes are damped, and if $N < \sigma < L_l$, gravity modes are damped. In convection zones, $N^2 < 0$, and gravity modes cannot propagate.

6.5.2 Observational Data

It is interesting to note that theoretical studies of non-radial oscillations of stars preceded the observational recognition of the importance of such motions. The first theoretical calculations were by Lord Kelvin (1863). Pekeris (1938) obtained the exact analytic solution for non-radial adiabatic oscilations for a uniform compressible fluid. Cowling (1941) extended these solutions to polytropes.

Work on non-radial oscillations was stimulated by difficulties in interpreting the radial velocity and line-profile variations in β CMa or β Cep stars. Ledoux (1951) showed that it was possible to explain some spectral characteristics of these stars in terms of non-radial modes. The discovery, in several stages, of the 5-min oscillation in the Sun (Evans and Michard 1962; Leighton, Noyes, and Simon 1962) led rapidly to helioseismology. The importance of this method of probing the interior of the Sun led to similar studies for some bright stars, such as α Cen A, Procyon, and ε Eri.

The most important of these observations, by Deubner (1975, 1977), and Rhodes, Ulrich and Simon (1977), clearly showed the properties of high-degree modes, because of the high instrumental resolution. Global observations of the Sun in radial velocity or photometry show the presence of modes of high order but low degree, $l = 0, 1, 2$, and 3. The highest-degree modes remain for the moment undetectable in global observations because of the cancellation between rising and falling motions or between local increases and decreases of brightness.

The low-degree modes are detected in radial-velocity fluctuations using resonant scattering in the analysis of solar line profiles. The light of the full solar disc (or a part of the disc) is focussed on to a transparent cylinder
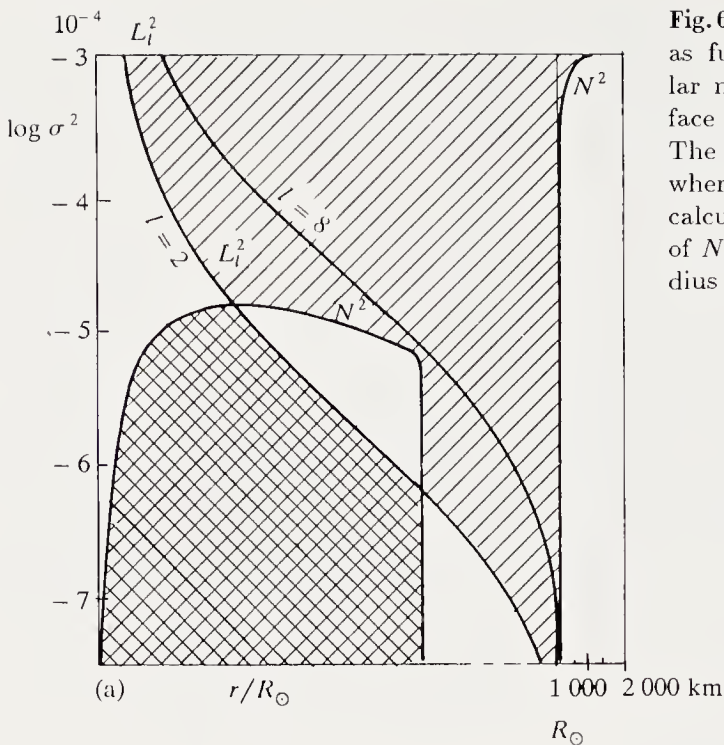
Fig. 6.20. (a) Curves of N^2 and L_l^2 as functions of radius for a solar model (the scale at the surface is expanded by a factor 40). The *hatched regions* denote zones where waves can propagate (L_l^2 is calculated for $l = 2$). (b) Curves of N^2 and L_l^2 as functions of radius (L_l^2 calculated for $l = 8$)
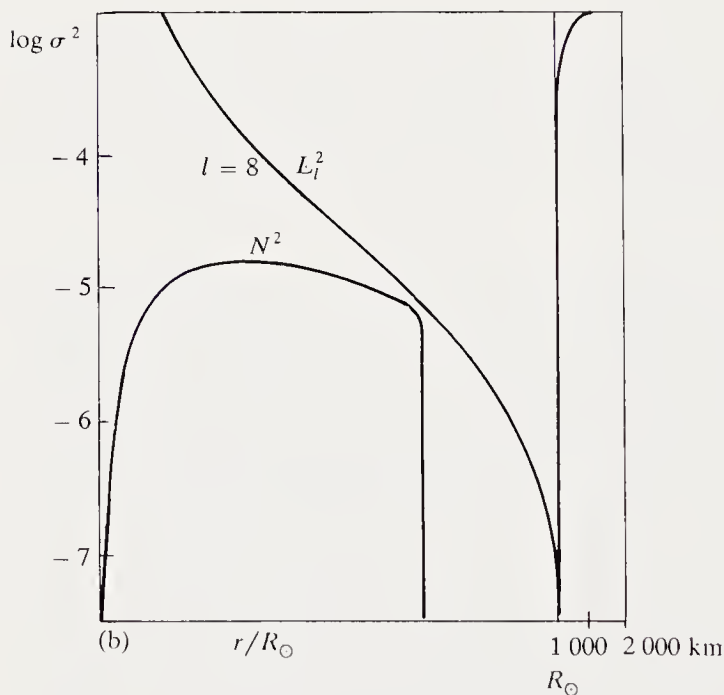
containing strontium, potassium, or sodium gas. Light which is resonantly scattered has an intensity proportional to the incident intensity at the wavelength of the scattering element in the laboratory standard. This wavelength can be altered by using a magnetic field parallel to the observation axis of the detector, and the measurement is made using a circular polaroid which selects one of the two σ components separated by the Zeeman effect
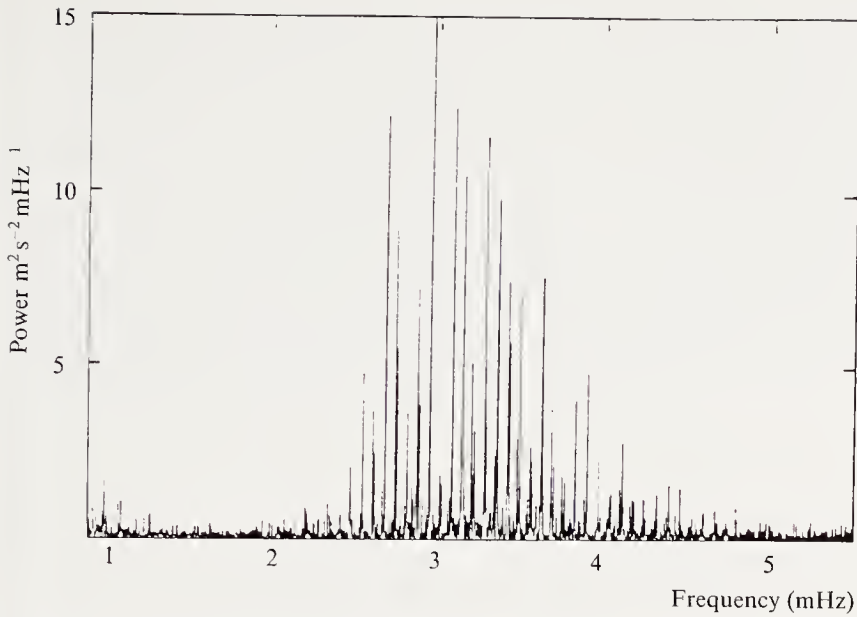
Fig. 6.21. Power spectrum of South Pole solar data. The data are from Doppler-shift measurements of the whole disc. The frequency resolution is $\Delta\nu = 1.97~\mu$Hz. (From E. Fossat, *Sol. Physics*, **82**, 55 (1983). Reproduced by kind permission of Kluwer Academic Publishers)

(Roddier 1965). The noise level in this method is at present of the order of $1~\mathrm{cm~s^{-1}}$.

Continuous observation of the Sun at the South Pole for 120 hours using sodium optical resonance yielded 80 modes when Fourier-analysed. Their periods lay between 3 and 10 min (actually 4.5 mHz to 2.5 mHz; Fig. 6.21). The same modes were identified by the *SMM (Solar Maximum Mission)* satellite from intensity fluctuations of the Sun. The amplitude of these modes is $\Delta I/I = 2$ to 3×10^{-6}.

Table 6.6 summarises the situation with respect to non-radial oscillations.

We shall not consider all the cases mentioned in Table 6.6 in the following. We give the basic theoretical picture and confine ourselves to the solar case.

6.5.3 The Linear Theory of Non-radial Oscillations

We confine ourselves here to the linear theory of adiabatic oscillations, even though it is clear (in particular from observations of variable white dwarfs) that the chaotic character of some oscillations can only be explained by a non-linear theory.

Non-adiabaticity occurs in stellar atmospheres, just where the surface boundary conditions have to be applied. Correct surface boundary conditions are quantitatively important. But the zero boundary conditions $P = \rho = 0$ preserve the general qualitative features of the oscillations. For

comparing the frequencies of high-n modes the surface boundary conditions are unimportant.

We start from the four basic equations (continuity, motion, energy, and Poisson equation). These constitute a fourth-order system of differential equations with respect to time.

In the linear approximation, all the perturbed quantities can be expanded in spherical harmonics using the associated Legendre polynomials $P_l^m(\mu)$ [1]. As an example, Fig. 6.22 shows the the amplitudes of various modes (l, m). If ξ is the displacement vector, with radial and horizontal components ξ_r, ξ_h related by the continuity equation, the displacement vector is given by

$$\boldsymbol{\xi} = \left[\xi_r(r), \xi_h(r)\frac{\partial}{\partial\theta}, \xi_h(r)\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi} \right] Y_l^m(\theta,\varphi)\mathrm{e}^{i\sigma t}. \tag{6.80}$$

The Y_l^m are normalised and related to spherical harmonics by

$$Y_l^m(\theta,\varphi) = (-1)^{(m+|m|)/2} \left(\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!} \right)^{1/2} P_l^m(\cos\theta)\mathrm{e}^{im\varphi}, \tag{6.81}$$

---

[1] We recall here some properties of Legendre polynomials. They are solutions of the differential equation

$$(1-z^2)\frac{\mathrm{d}^2 u}{\mathrm{d}z^2} - 2z\frac{\mathrm{d}u}{\mathrm{d}z} + \left[l(l+1) - \frac{m^2}{1-z^2} \right] u = 0.$$

We also have

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{\mathrm{d}^m}{\mathrm{d}x^m} P_l(x),$$

$$P_l^{-m}(x) = (-1)^m \frac{\Gamma(l-m+1)}{\Gamma(l+m+1)} P_l^m(x).$$

The first few Legendre polynomials are:

$$P_0 = 1,$$
$$P_1^0 = x = \cos\theta,$$
$$P_1^1 = -(1-x^2)^{1/2} = -\sin\theta,$$
$$P_2^0 = \tfrac{1}{2}(3x^2 - 1) = \tfrac{1}{4}(3\cos 2\theta + 1),$$
$$P_2^1 = -3(1-x^2)^{1/2}x = \tfrac{3}{2}\sin 2\theta,$$
$$P_2^2 = 3(1-x^2) = \tfrac{3}{2}(1 - \cos 2\theta),$$
$$P_3^0 = \tfrac{1}{2}(5x^3 - 3x) = \tfrac{1}{8}(5\cos 3\theta + 3\cos\theta),$$
$$P_3^1 = -\tfrac{3}{2}(1-x^2)^{1/2}(5x^2 - 1) = -\tfrac{5}{8}(\sin\theta + 5\sin 3\theta),$$
$$P_3^2 = 15(1-x^2)x = \tfrac{15}{4}(\cos\theta - \cos 3\theta),$$
$$P_3^3 = -15(1-x^2)^{3/2} = \tfrac{15}{4}(3\sin\theta - \sin 3\theta),$$
$$P_4^0 = \tfrac{1}{8}(35x^4 - 30x^2 + 3) = \tfrac{1}{64}(35\cos 4\theta + 20\cos 2\theta + 9).$$

Table 6.6. Observational data explained or explicable by non-radial oscillations[a]

Type of star	Phenomenon	Proof of non-radial oscillation	Mode
β CMa	Pulsating variables	Beat? Profile variations	$l \sim 2$?
O, B stars	Line-profile variations	Profile variations	$l \sim 2$
DA white dwarfs DB white dwarfs	Periodic light curve	Period value, multiplicity of periods	g modes with small l
Cataclysmic variables	Periodic light curve	Values of periods, phase changes of oscillation during eclipse	g modes with small l
Sun	5-min. oscillation	(k, σ) diagnostic diagram	p modes with large l
	Global oscillations	Values of periods	p (and g?) modes with small l
Ap stars	Global oscillations	Values of periods	$l = 0, 1, 2$
Blue supergiants (α Cyg type)	Light curve, semi-regular radial-velocity variations	Multiplicity of periods, width of lines	$l \sim |3, 4$ p or g modes
δ Scu stars	Pulsating variables	Multiplicity of periods	p modes
Red giants and supergiants	Chromosphere, Wilson–Bappu effect[b]	Heating of outer atmospheric layers	p modes with large l

[a] From Unno, Osaki, Ando, and Shibahashi (1979).
[b] Width of calcium H and K lines increases with luminosity.

and the equations of motions reduce to the three relations

$$\frac{1}{\rho}\frac{\mathrm{d}p'}{\mathrm{d}r} + \frac{g}{c^2\rho}p' + (N^2 - \sigma^2)\xi_r + \frac{\mathrm{d}\Phi'}{\mathrm{d}r} = 0, \tag{6.82}$$

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}(r^2\xi_r) + \frac{1}{\Gamma_1}\frac{\mathrm{d}\ln p}{\mathrm{d}r}\xi_r + \left(1 - \frac{L_l^2}{\sigma^2}\right)\frac{p'}{\rho c^2} - \frac{l(l+1)}{\sigma^2 r^2}\Phi' = 0, \tag{6.83}$$

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\Phi'}{\mathrm{d}r}\right) - \frac{l(l+1)}{r^2}\Phi' - 4\pi G\rho\left(\frac{p'}{\rho c^2} + \frac{N^2}{g}\xi_r\right) = 0, \tag{6.84}$$

for p' (pressure perturbation), ξ_r, and Φ' (gravitational potential perturbation), where L_l and N are the Lamb and Brunt–Väisälä frequencies defined above.

Two extreme cases are particularly important: (a) surface waves (large l), and (b) global oscillations (small l but large n).

High modes can be studied by the WKB (Wentzel–Kramers–Brillouin) method: the asymptotic solutions allow a discussion in terms of the trapping
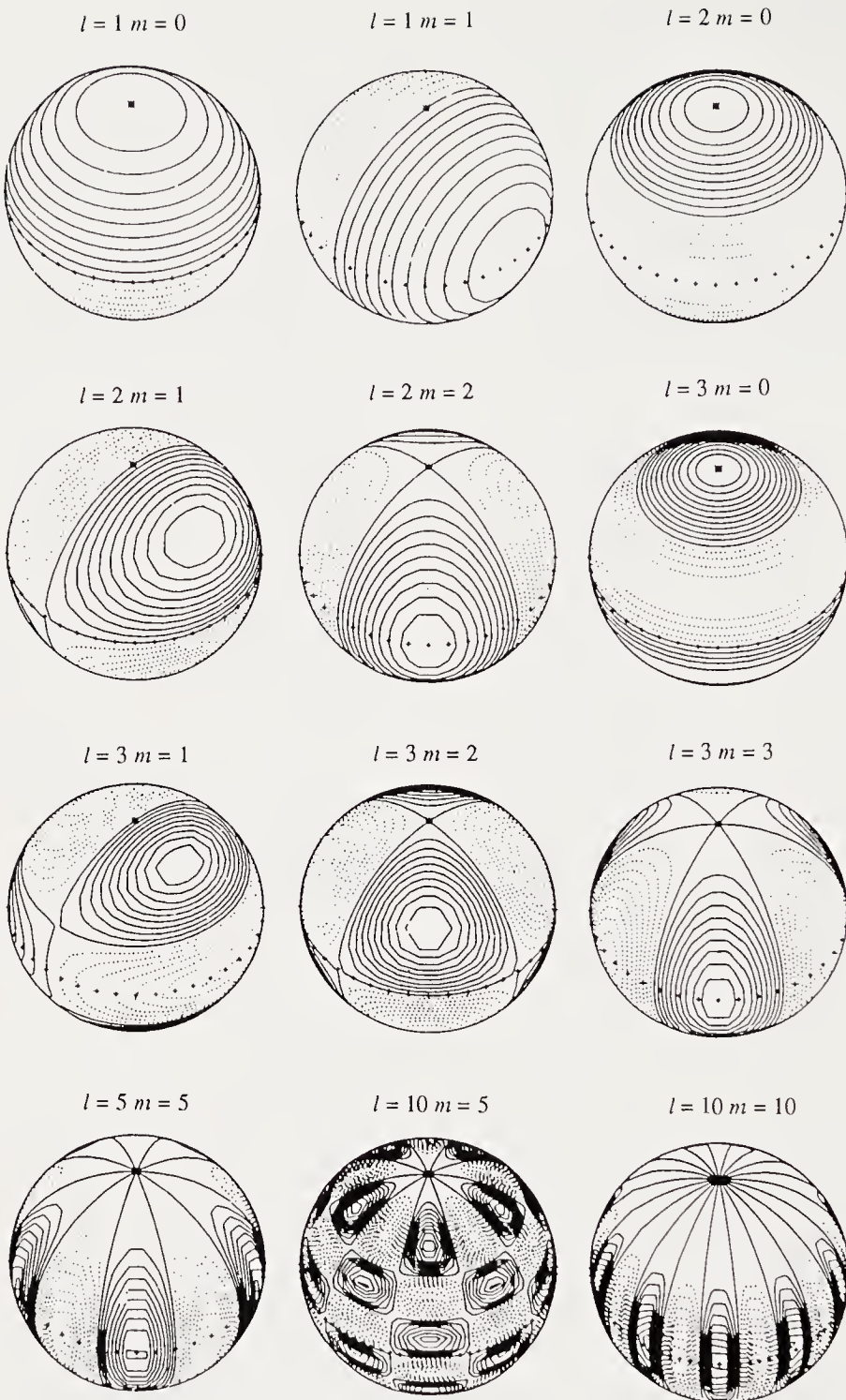
Fig. 6.22. Amplitudes of modes $l = 1$, 2 and 3 and three higher-degree modes

and propagation mechanisms of waves inside the star. For high modes (n or l large) it is possible to neglect the perturbation of the potential Φ. The system then reduces to two differential equations. We introduce new variables:

$$\tilde{\xi} = r^2 \xi_r \exp\left(-\int_0^r \frac{g}{c^2} dr\right), \tag{6.85}$$

$$\tilde{\eta} = \sigma^2 r \xi_r \exp\left(-\int_0^r \frac{N^2}{g} dr\right) = \frac{p'}{\rho} \exp\left(-\int_0^r \frac{N^2}{g} dr\right), \tag{6.86}$$

and the system reduces to the canonical form

$$\frac{d\tilde{\xi}}{dr} = h(r)\frac{r^2}{c^2}\left(\frac{L_\rho^2}{\sigma^2} - 1\right)\tilde{\eta}, \tag{6.87}$$

$$\frac{d\tilde{\eta}}{dr} = \frac{1}{r^2 h(r)}(\sigma^2 - N^2)\tilde{\xi} \tag{6.88}$$

with

$$h(r) = \exp\left[\int_0^r \left(\frac{N^2}{g} - \frac{g}{c^2}\right) dr\right] > 0. \tag{6.89}$$

To a first approximation the WKB method gives a solution

$$\tilde{\xi}(r), \tilde{\eta}(r) \sim \exp\left(i\int k_r dr\right) \tag{6.90}$$

with

$$k_r^2 = \frac{1}{c^2\sigma^2}(\sigma^2 - L_l^2)(\sigma^2 - N^2). \tag{6.91}$$

The importance of the sign of k_r^2 is apparent when one plots a *propagation diagram* (drawn here for the Sun), where the limiting frequencies L_l^2 and N^2 are shown as functions of radius (Figs. 6.20a and b).

Surface Waves (Large l). The Brunt-Väisälä frequency vanishes at the boundary of the convection zone. Writing

$$N^2 = g\left(\frac{1}{\Gamma_{\mathrm{rad}}} - \frac{1}{\Gamma_{\mathrm{ad}}}\right)\frac{1}{H(r)}, \tag{6.92}$$

where H is the local scaleheight,

$$H(r) = \frac{\Re T}{g\mu}, \tag{6.93}$$

we immediately see that N^2 grows as we move inwards from the convection zone, reaches a maximum, and then decreases because of the growth of the scaleheight with temperature (Figs. 6.20a, b).

The Lamb frequency

$$L_l^2 = \frac{l(l+1)c^2}{r^2} = \frac{l(l+1)}{r^2}\frac{\gamma\Re T}{\mu}, \tag{6.94}$$

on the other hand, has a minimum at the temperature minimum. In the atmosphere the curve $L_l^2(r)$ only lies above that for $N^2(r)$ for very high values $l \gtrsim 3000$.

The propagation diagram (Figs. 6.20a, b) for the whole Sun (with L_l^2 plotted for $l = 2$ and $l = 8$) gives the propagation regions of the p and g modes (shown hatched) as the domains $k_r^2 > 0$. The g modes cannot have frequencies greater than 4 mHz (or periods less than 250 s). They are confined to the solar interior, below the convection zone. By contrast, the p modes are confined between the solar atmosphere and a deep layer of the Sun, deeper for higher-order modes. The frequency rises with the degree of the mode, and high-degree modes penetrate very little into the interior: these are surface modes.

Assuming a temperature distribution

$$\frac{\Re T}{\mu} = \frac{\gamma - 1}{\gamma} g \Delta r \tag{6.95}$$

near the convection zone, and using the approximate solution

$$\int_a^b \frac{1}{c} (\sigma^2 - L_l^2)^{1/2} \mathrm{d}r = n\pi , \tag{6.96}$$

where a and b denote the boundaries of the p-mode propagation zone ($L_l^2 = \sigma^2$ and $r = R$), we find

$$\sigma^2 = \frac{4(\gamma - 1)n\sqrt{l(l+1)}}{R} g . \tag{6.97}$$

With $\gamma - 1 = 0.25$, high values of $l = \pi R k_h / 4$ (where k_h is the horizontal wavenumber) obey

$$\sigma = 0.014\sqrt{n}\sqrt{k_h / Mm^{-1}} \, \mathrm{s}^{-1} . \tag{6.98}$$

We thus get an approximate solution for high-degree modes which agree with the general trend of the observations (Deubner's diagnostic plot, Fig. 6.23). A full calculation shows that the surface waves give a diagnostic for the structure of the convection zone, which is penetrated by these high-degree modes.

Global Oscillations. The asymptotic form of the eigenfrequencies for low-degree oscillations makes it easy to see their significance. As we have seen, modes with $l = 0, 1, 2, 3$ are so far the only ones to be detected in global observations of the solar disc.

For frequency ν_{nl} we have the expansion

$$\nu_{nl} = \left(n + \frac{l}{2} + \varepsilon\right)\nu_0 - \frac{(l(l+1) + B)A}{n + \frac{l}{2} + \varepsilon} , \tag{6.99}$$

302
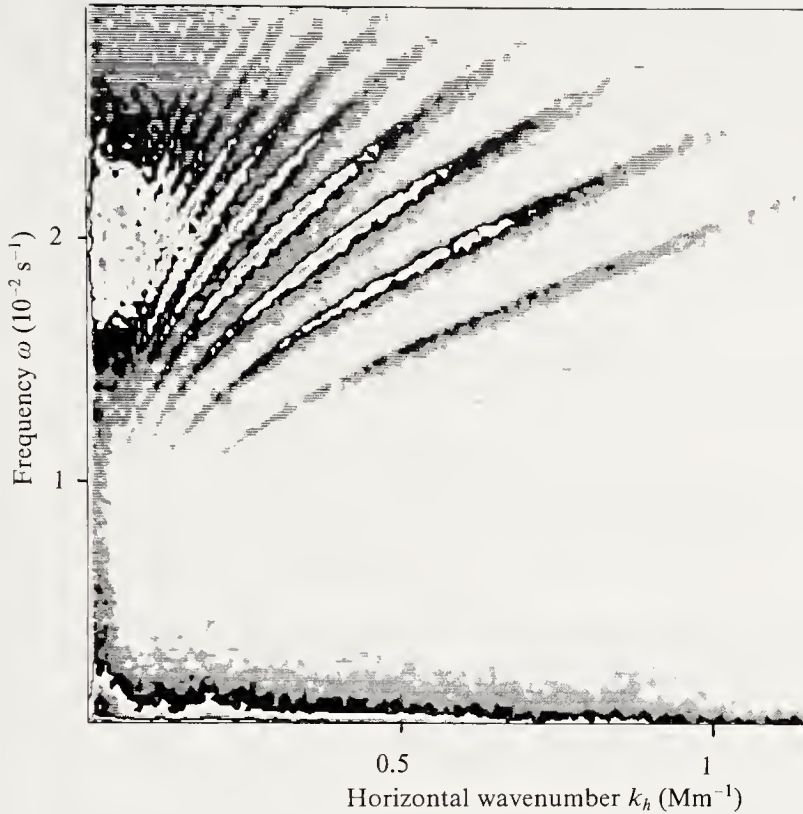
Fig. 6.23. Diagnostic plot for high-degree, high-order p modes showing the power spectrum obtained by F. Deubner (kindly supplied by the author)

where ε is of the order of unity and B, which has a complicated form, is also of the order of 1.

A is given by

$$
A = \frac{1}{2\pi^2} \left[\frac{c(R_\odot)}{R_\odot} - \int_0^R \frac{1}{r} \frac{\mathrm{d}c}{\mathrm{d}r} \mathrm{d}r \right] . \tag{6.100}
$$

The principal term is

$$
\nu_0 = \left(2 \int_0^R \frac{1}{c} \mathrm{d}r \right)^{-1} \tag{6.101}
$$

so that $2\nu_0$ is the propagation time of an acoustic wave from the centre to the surface of the Sun. An important probe of the solar interior is obtained by comparing the frequencies ν_{nl} for values (n, l) and $(n-1, l+2)$:

$$
\nu_{n,l} - \nu_{n-1,l+2} = \delta\nu_{nl} = \frac{(4l+6)A}{n + \frac{l}{2} + \varepsilon} , \tag{6.102}
$$

where A is related to the gradient of the sound speed near the core. The global Sun spectra show clearly only the modes of degree 0, 1, 2, and 3. The
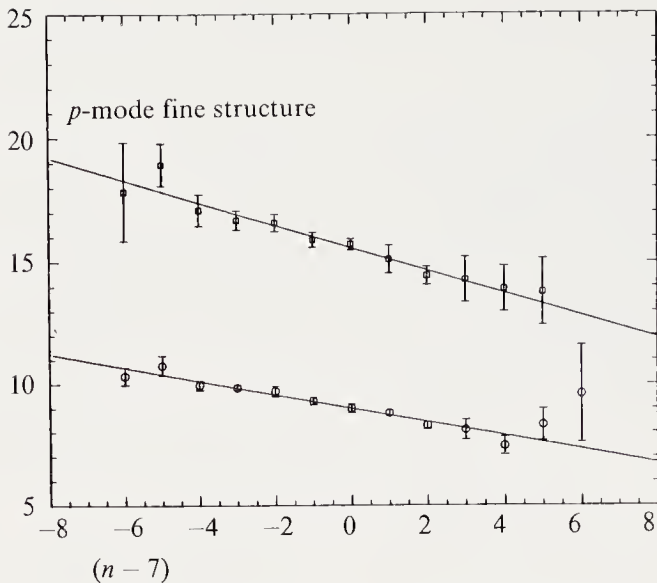
pattern of the spectrum repeats modulo $\nu_0 = 135 \ \mu$Hz. The pattern itself shows the modes $(n - 1, 3)$, $(n, 1)$, $(n, 2)$, $(n + 1, 0)$ with the succession of frequencies

$$0, \ \frac{10A}{n + \frac{1}{2} + \varepsilon}, \ \frac{\nu_0}{2} + \frac{6A}{n + \frac{1}{2} + \varepsilon}, \ \frac{\nu_0}{2} + \frac{12A}{n + \frac{1}{2} + \varepsilon}\left(1 - \frac{1}{2n}\right)$$

assuming that n is large. This spacing can easily be read off the spectrum, allowing the values of n and l to be identified. We thus plot the frequency differences $\delta\nu_{nl}$ as functions of frequency or of the order n of the mode (Fig. 6.24) and compare with the theoretical frequency differences which decrease as the frequency rises. Adjustment of the mode of order 21, arbitrarily chosen, allows one to compare the measured value of $D_0 = \delta_{nl}/(4l + 6)$, $D_0(n, 0) = 1.52 \ \mu$Hz, with the theoretical value $D_0 = 1.51 \ \mu$Hz for the "standard" solar model.

Study of the Solar Interior. The approximate formula (96) for the frequency may be rewritten

$$\int_{r_p}^{R_\odot} \frac{1}{c}\left(\sigma^2 - \frac{l(l + 1)c^2}{r^2}\right)^{1/2} dr = (n + \alpha)\pi \,, \tag{6.103}$$

where α allows one to connect the pressure-wave propagation region (from r_p to $R_\odot$) and the evanescent wave zone, and must be found empirically from the observations.

This relation can be approximated by the empirical relation

$$\frac{\pi(n + \alpha)}{\sigma} = F\left(\frac{\sigma}{L}\right) \tag{6.104}$$

with $L^2 = l(l + 1)$.

Thus we plot $\pi(n + \alpha)/\sigma$ as a function of (σ/L) to give α by empirical fitting. Christensen-Dalsgaard et al. (1985) give $\alpha = 1.58$.

Setting

$$\frac{\sigma}{\sqrt{l(l+1)}} = x \tag{6.105}$$

we can write

$$\pi\frac{n + \alpha}{\sigma} = \int_{r_p}^{R_\odot} \sqrt{\frac{r^2}{c^2} - \frac{1}{x^2}}\,\frac{\mathrm{d}r}{r} = F(x). \tag{6.106}$$

We can thus invert (6.103) to give

$$r = R_\odot \exp\left[-\frac{2}{\pi}\int_{c(R_\odot)/R_\odot}^{c(r)/r}\left(\frac{1}{x^2} - \frac{r^2}{c^2}\right)^{-1/2}\frac{\mathrm{d}F}{\mathrm{d}x}\mathrm{d}x\right], \tag{6.107}$$

where $F(x)$ is given by observation; we thus get a relation $r = f(c/r)$ which gives $c(r)$, which we can compare with the sound speed in theoretical models. The differences between the sound speeds (deduced from helioseismology and given by the standard model) are less than 2 %, implying a temperature deficit of 4 % in the model near $r = 0.4R_\odot$.

Rotation. We mention here only the lifting of the degeneracy in m by rotation. The separation by $\pm m\Omega_\odot$ of the oscillation frequencies in principle gives Ω as a function of depth into the solar interior. However, inversion of the relation giving the frequency separation is difficult because of the relatively noisy data. We can conclude only that the $\Omega(r)$ law is close to rigid rotation in the range $0.2R_\odot < r < R_\odot$ (Fig. 6.25).
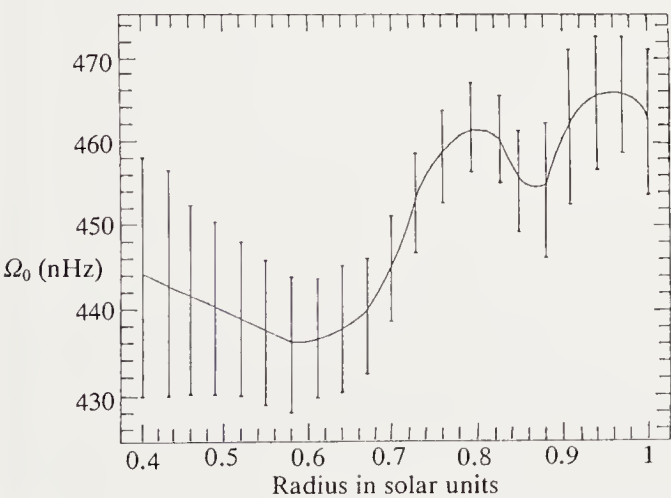
Fig. 6.25. Angular rotation frequency in the solar interior, as a function of the distance from the axis. (From W. Dziembowski et al. 1989)

Asteroseismology. Studies similar to those for the Sun have been made for Procyon and α Cen A (Gelly, Grec, and Fossat 1986), and for ε Eri by Noyes et al. (1984). These show the same system of equidistant modes as in the Sun. These studies, which are still preliminary, have not yet produced results consistent with other measures of the internal structure of these stars.

7. Solar and Stellar Activity

The surfaces of the stars, particularly that of the Sun, are the sites of time-dependent phenomena which, at least for the Sun, are directly linked to the heterogeneous nature of its surface and the appearance of magnetic fields (see Chap. 2). The most striking sign of solar activity is the existence and temporal evolution of sunspots, which possess intense magnetic fields (up to 4500 G). While the spots are located at the photosphere, activity is also seen in higher layers; another characteristic phenomenon is the strengthened emission of some lines or continua (the *level* of activity) (Fig. 7.1): in the Sun, this strengthening is seen in heterogeneous chromospheric structures (plages). In whole-disc observations of the Sun (i.e. viewed as a star) over several years this strengthening is more significant at certain phases of the activity *cycle*.

The term "activity" actually refers to two kinds of phenomenon. One is a type of surface variability caused by structure due to the magnetic field: the Sun and stars, besides being oscillating mechanical systems and thus intrinsic variables (see Chap. 6), are also *magnetic variables*. On the other hand, activity may also mean the strengthening of emissions which show the existence of temperatures in "active regions" which are locally enhanced above those of "quiet regions"; activity then means *heating of the layers* where these emissions appear.

7.1 Indicators of Activity

7.1.1 Sunspots and Starspots

We have seen in Chap. 2 that spots are scattered over the Sun's disc between latitudes $+30°$ and $-30°$, their number and positions varying over time. Each spot persists for a few (2 or 3) rotations, but spots are not rigidly fixed to the same place on the Sun: at the beginning of the cycle they appear at high latitude, generally in pairs; the pair migrates towards the equator or the poles depending on its latitude, indicating circulation in the form of rolls. The mean latitude at which spots appear decreases as the cycle progresses.

A plot of spot latitude against time has a characteristic form: this is the Maunder or butterfly diagram (Fig. 7.2). We can count the number of
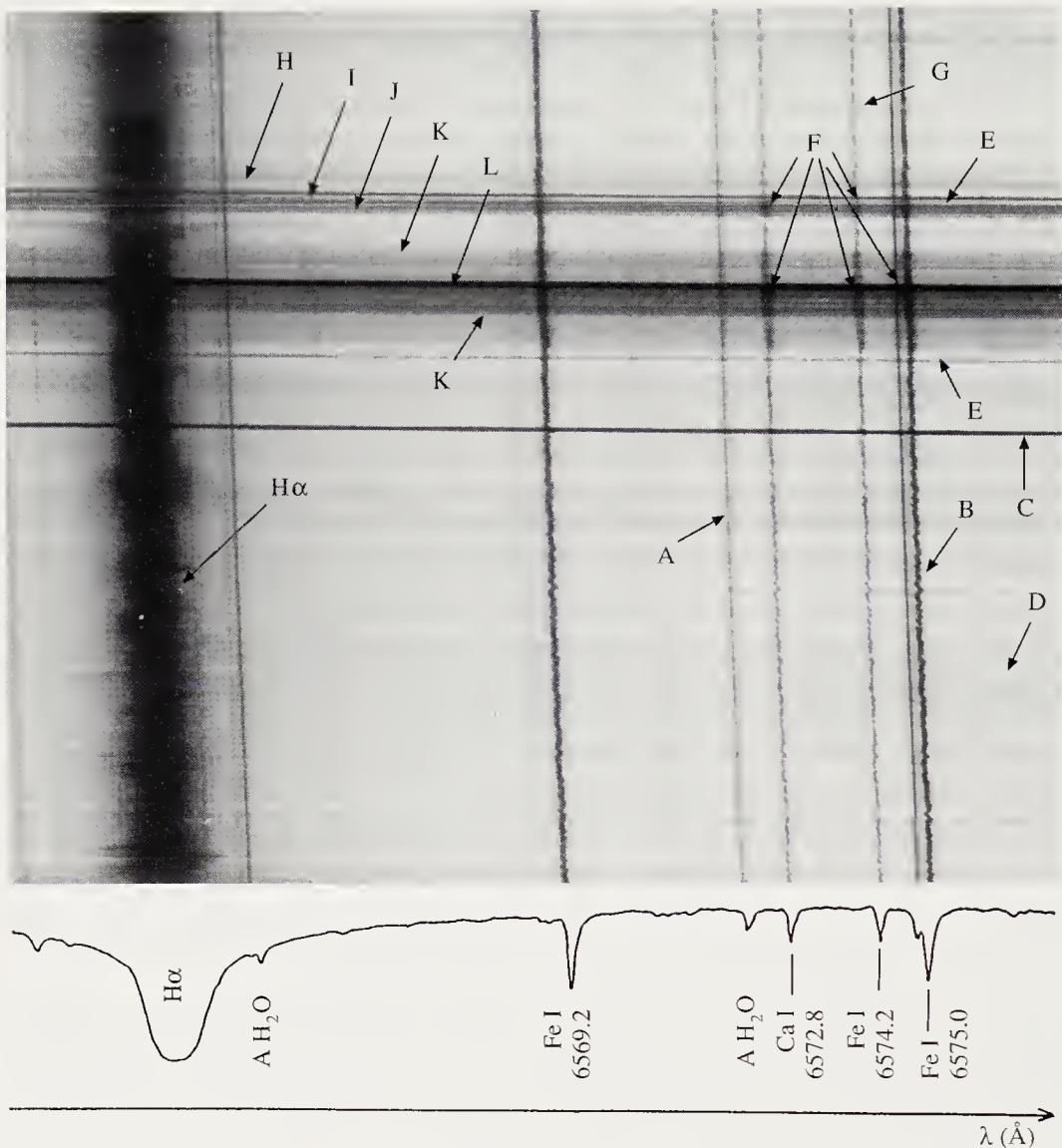
Fig. 7.1. Solar spectrum near an active centre. The intensity profile of the main spectral lines is shown below the photograph. Hα is the intense hydrogen line at 6562.8 Å. The line marked C is a reference wire. The photograph shows properties appearing in various forms in most stars. The zig-zags of the line B and the intensity fluctuations marked as D are caused by the solar granulation: the Doppler effect of moving granules and the intensity differences associated with rising and falling motions. A denotes a telluric line, which of course shows no zig-zags. Many traces of solar activity are visible, small spots EE, a large spot L with penumbra KK, another spot I, a bright facula J, Zeeman broadening caused by a magnetic field at F, and the decreased intensity of line G due to another magnetic effect. H indicates an emission line, very broadened by the rapid motion of the emitting gas, and called a "moustache". (Photograph by the Observatoire de Paris, obtained from the spectrograph of the solar tower of the Observatoire du Pic du Midi on 1987 June 25)

sunspots visible on the Sun's disc at a given time and make annual averages. This number (called the Wolf number) varies with time and is the oldest measure of solar activity. It can be recovered uninterruptedly from the time
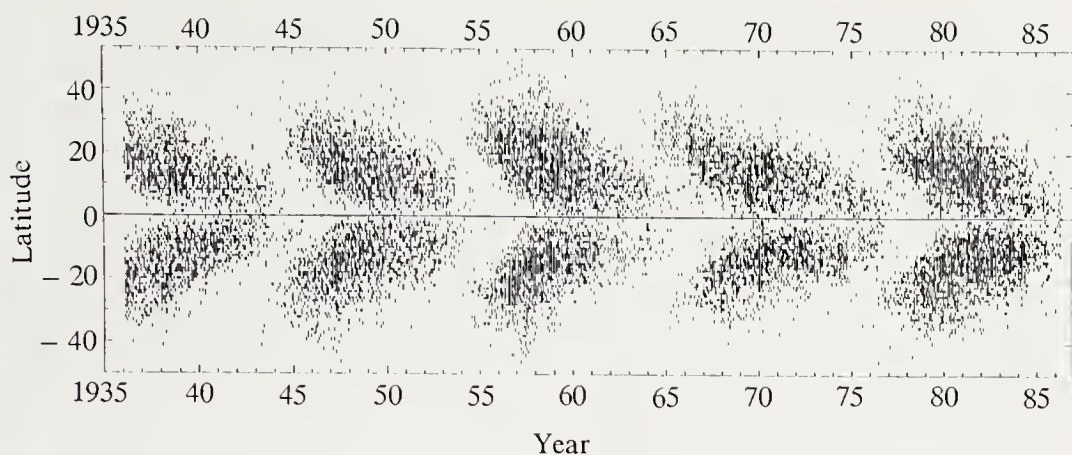
Fig. 7.2. Butterfly or Maunder diagram from 1935 to 1987. (R.K. Ulrich, private communication)
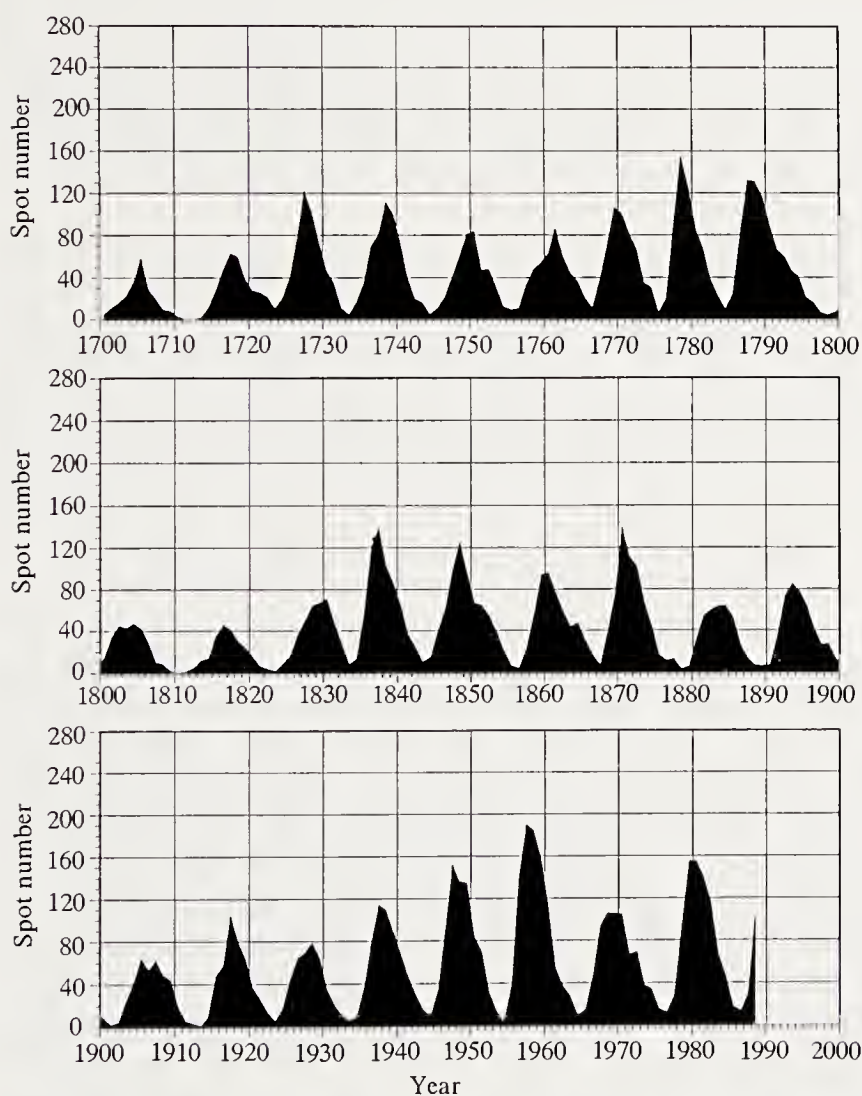
Fig. 7.3. Amplitude of the solar cycle: variation of the Wolf number with time. The annual average spot number is plotted against the year for the epoch 1700–1988. (World Data Center for Solar–Terrestrial Physics, Boulder)

of the first telescope observations in about 1610, and has been recorded systematically since the 19th century. This series of observations shows an essential property first recognised by Schwabe in 1844: the spot number varies with an eleven-year period, although the amplitude is not constant (Fig. 7.3). The Sun thus passes alternately through a phase of maximum activity (average spot number between 40 and 180) and then a minimum (about 5 spots). But the maxima can vary in intensity and the minima may be more or less inactive. Eleven years is the dominant period, but at times the cycle can be as short as 8 years, while at others it is as long as 15 years. The solar cycle can also be irregular: the most famous example is the Maunder minimum between 1645 and 1715.

As we have seen, the spots are one of the sites where new magnetic field emerges; spots migrating as a pair always have opposite polarities. Following the spots in time reveals another period in solar activity: those spots born with positive polarity in the northern hemisphere in one cycle have negative polarity in the next, implying a period of *22 years*. Studies of the large-scale magnetic field also show a 22-year periodicity in the polarity of the solar poles, the North Magnetic Pole becoming the South Pole and vice versa after 11 years. The reversal of polarity between North and South Poles occurs with a delay of up to a year.

Spots are thus a powerful way of monitoring polarity variations of the solar magnetic field and also the stability of the solar activity cycle.

However it is noticeable that the spots never cover in total more than 0.4% of the visible disc, with the largest spots occupying at most 1.5×10^{-3} of the visible hemisphere. If we observe the Sun without spatial resolution, the variations of the Sun's integrated luminosity caused by the presence of spots and their irregular distribution in longitude as the Sun rotates are of the order of a few percent at most. These variations are caused by the fact that the spots are cooler than the surrounding photosphere ($T_{\text{eff}} \simeq 4200$ K in the umbra, compared with 5700 K).

Are stars also covered with spots? This is more difficult to prove than for the Sun, because stellar discs are not resolved. But there are stars where the area covered by spots on the visible disc is much greater than on the Sun, causing a measurable periodic photometric variation as the spots cross the line of sight as the star rotates (> 0.05 mag, and sometimes several tenths of a magnitude). We observe late-type stars in binaries of the BY Dra and RS CVn type which have large-scale spots on their surfaces (up to 30 or 50% of the disc area). These are sometimes near the poles, and have a sufficiently asymmetric longitude distribution as to cause a rotational modulation of the apparent brightness (Fig. 7.4). Analysis of the light curve shows that the spots migrate differentially over the disc in time. The T Tauri stars also have surfaces covered with spots.

This gives us an indirect way of resolving the discs of some stars, since we can use photometry to produce a crude map of the longitude distribution of starspots. On the other hand, it has not yet been possible to show that
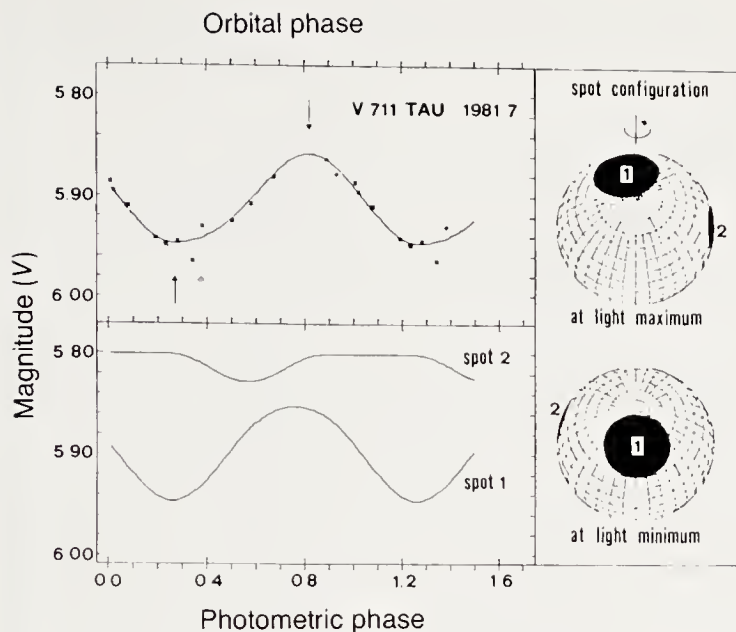
Fig. 7.4. Evidence for starspots. The light curve of the RS CVn star V711 Tau is interpreted as being due to the superposition of two large spots (called 1 and 2). Two epochs are shown. (From M. Rodono et al., *Astron. Astrophys* **165** 135 (1985) Fig. 10, p. 147)

the spots are magnetic; they cannot be individually resolved, and we do not know if the spots are analogous to those of the Sun, only much larger, or whether they are collections of many smaller spots.

Of course, the detection of starspots in white light depends on the photometric quality of the measurement. As for solar oscillations, photometry from space will be needed to detect spots at a much lower contrast level than on the Sun ($\Delta m \simeq 0.001$).

7.1.2 Spectroscopic Activity Criteria in the Visible and Ultraviolet

Observations of lines formed in the solar chromosphere and transition zone show that the intensity of the emission in these lines increases in moving from a quiet region (e.g. the centre of a supergranule) to the boundary of a supergranule or to a plage, and *a fortiori* to a flare. Depending on the part of the disc being observed, the Sun is more or less active in these lines. Similarly, coronal lines are intensified in coronal arches or loops.

Stars of spectral type close to that of the Sun (roughly F0 to M) emit energies in these same lines which can greatly exceed (factors of 100 to 1000) those of the average Sun or even the active Sun. Although their discs are not resolved, it is tempting to assume that these stars are covered with active regions, like the Sun.

Which lines are most useful as indicators of stellar activity? These are usually resonance lines, that is, lines involving the ground state of an atom or ion and a nearby excited state, and lines of abundant elements. Depend-

Table 7.1. Lines showing the existence of chromospheres, transition zones and coronae

(A) *Transition zones and coronae*

Ion	Line λ (Å)	Transition	Formation temperature T_e (K)
Mg X	609.76 624.93	2s ^{2}S-2p ^{2}P^o —	1.2×10^6
Ne VII	465.221	2s^2 ^{1}S-2s 2p ^{1}P^o	5×10^5
O VI	1 031.912 1 037.613	2s ^{2}S-2p ^{2}P^o —	3.2×10^5
O IV	554.51	2s^2 2p ^{2}P^o-2s 2p^2 ^{2}P	1.5×10^5
N V	1 238.821 1 242.804	2s ^{2}S-2p ^{2}P^o —	1.6×10^5
He II	303.78	1s ^{2}S-2p ^{2}P^o	$\sim 8 \times 10^4$
He II	1 640.332 1 640.474	2p ^{2}P^o-3d ^{2}D —	$\sim 8 \times 10^4$
C IV	1 548.185 1 550.774	2s ^{2}S-2p ^{2}P^o —	1×10^5
Si IV	1 393.755 1 402.770	3s ^{2}S-3p ^{2}P^o —	6×10^4
C III	1 175.711 1 174.933	2s 2p ^{3}P^o-2p^2 ^{3}P —	5×10^4

ing on the dominance condition of their source function and optical depth these appear either completely in emission, or with a central emission (self-reversal) within wide absorption wings. The condition for an optically thick line to be in emission at the wavelength λ of its profile is

$$S_\lambda(\tau_\lambda \simeq 1) \geq S_c(\tau_c \simeq 1), \qquad (7.1)$$

where S_λ and S_c are the source functions in the line and the continuum and τ_λ and τ_c are the optical depths. If S_λ is collision-dominated, i.e, photons are emitted essentially at the expense of electron energy, the line may appear in emission (Thomas 1957).

Table 7.1 gives a list of these lines and the temperature of the region where they are formed: corona, transition zone, or chromosphere. Most of them are observable in the space ultraviolet, explaining the great importance for the study of stellar activity of the *Copernicus* and *IUE* (*International Ultraviolet Explorer*) satellites, launched in 1972 and 1978 respectively. EUV lines, with $\lambda < 900$ Å (Mg X, Ne VII, O IV, He II λ 304 Å) have not yet been observed in stars.

It is worth stressing that observations of the Ca II H and K lines from the ground have revealed many of the characteristic features of solar and stellar activity: the existence of stellar chromospheres, the discovery of stel-

Table 7.1 (continued)

(B) *Chromospheres in the solar sense*

Ion	Line λ (Å)	Transition	Formation temperature T_e (K)
H I	1 215.67	1s ^{2}S-2p ^{2}P°	$1\text{-}2 \times 10^4$
	6 562.8 (Hα)	2p ^{2}P°-3d ^{2}D	$6\text{-}10 \times 10^3$
	3 970.07 (Hε)	2p^2 P°-7d ^{2}D	
He I	10 830	1s 2s ^{3}S-1s 2p ^{3}P°	2×10^4
	5 875.7	1s 2p ^{3}P°-1s 3d ^{3}D	
C II	1 334.532	2s^2 2p ^{2}P°-2s 2p^2 ^{2}D	2×10^4
	1 335.708		
Si III	1 206.51	3s^2 ^{1}S-3s 3p ^{1}P°	3.5×10^4
	1 892.03	3s^2 ^{1}S-3s 3p ^{3}P°	
Si II	1 304.372	3s^2 3p ^{2}P°-3s 3p^2 ^{2}S	1.2×10^4
	1 309.277		
	1 526.708	3s^2 3p ^{2}P°-3s 4s ^{2}S	1.2×10^4
	1 533.432		
Ca II	3 933.66 (K)	4s ^{2}S-4p ^{2}P°	$6\text{-}8 \times 10^3$
	3 968.47 (H)		
	8 542.09	3d ^{2}D-4p ^{2}P°	
	8 662.14		
	8 498.02		
Mg II	2 795.53 (k)	3s ^{2}S-3p ^{2}P°	$6\text{-}8 \times 10^3$
	2 802.70 (h)		
O I	1 302.169	2p^4 ^{3}P-2p^3 3s ^{3}S°	6×10^3
	1 304.875		
C I	1 560.683	2s^2 2p^2 ^{3}P-2s 2p^3 ^{3}D°	6×10^3
	1 657.008	2s^2 2p^2 ^{3}P-2p 3s ^{3}P°	

lar activity cycles by O. Wilson (1978), indirect evidence of stellar plages from the rotational modulation of emission in the H and K lines (Vaughan et al. 1981), evidence for amplified nonradial oscillations in the same lines.

Figures 7.5a and b show the Ca II H and K and the Mg II h and k lines, as observed by the OSO-8 satellite, in both a quiet region and an active region of the Sun.

Figures 7.6a and b show the same lines, but observed in stars, where the disc is unresolved. This suggests the notion of more (ε Eri) or less (τ Cet, not shown) active stars, depending on the mean level of intensity of the lines averaged over the disc.

We note finally that the lines of neutral helium at 5876 and 10 830 Å, which figure in Table 7.1, are not resonance lines. But the excitation energy required to form them is large enough that we conclude that they are not formed in the photosphere, at least in solar-type stars. Calculation supports this idea. These lines are observed either in emission or in absorption.
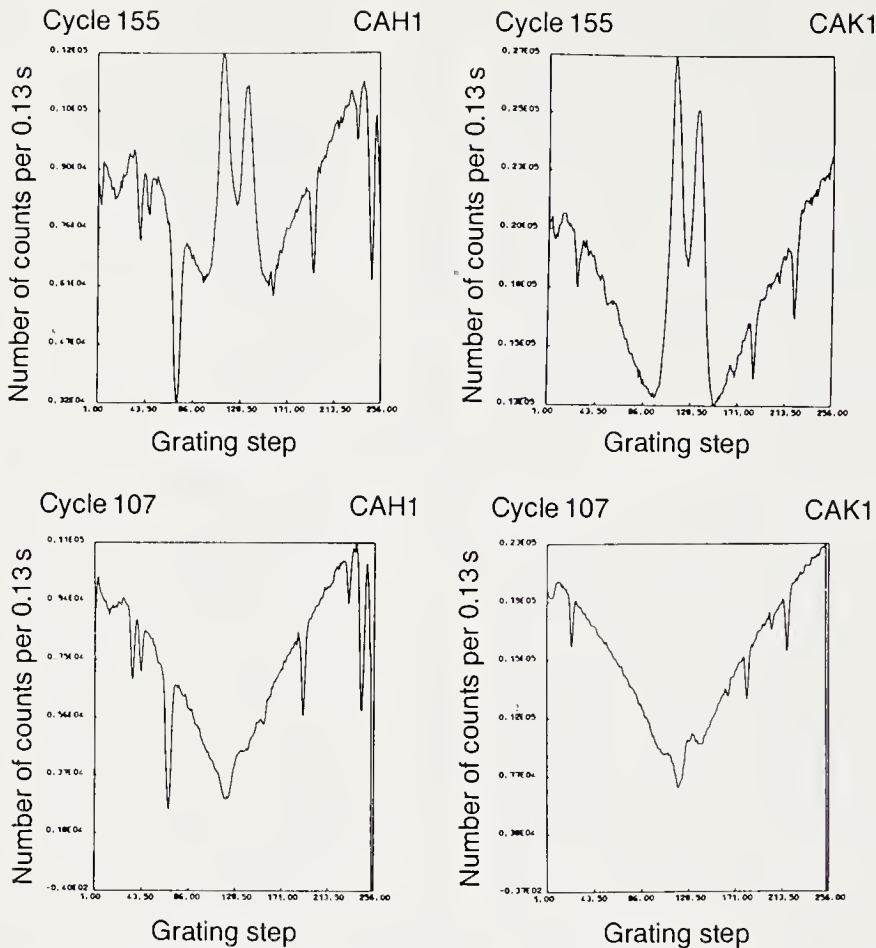
Fig. 7.5a. Ca II H and K lines observed by *OSO-8* in an active (*upper panel*) and quiet (*lower panel*) region of the Sun, of size $2'' \times 10''$. The spectral resolution is 0.02 Å. (Ph. Lemaire, private communication)

7.1.3 X-rays

Although the Sun has been known as an X-ray source since 1945, it was the *Skylab* orbital platform (1973) which provided the main results on solar X-ray emission. Figure 7.7 shows that the Sun seen in X-rays is strongly structured: we distinguish dark regions, or *coronal holes*, and bright tubular regions, or *coronal loops*. Moreover many *bright points* cover the Sun in X-rays. Each of these structures, which show that the corona is heterogeneous just like the lower layers of the atmosphere, can be matched to a feature in the chromosphere or transition zone, and even the photosphere. Thus coronal loops have footpoints in chromospheric active regions, each of which lies above a complex of spots. As we have seen in Chap. 2, coronal holes cover the photospheric and chromospheric regions of the quiet Sun, as can be identified from solar images in the He I 10 830 Å line.

It is not possible to measure directly the magnetic field at the coronal level, since this requires measurements of the polarisation in the UV or EUV lines, where polarisers are not efficient. However, it is possible to link the
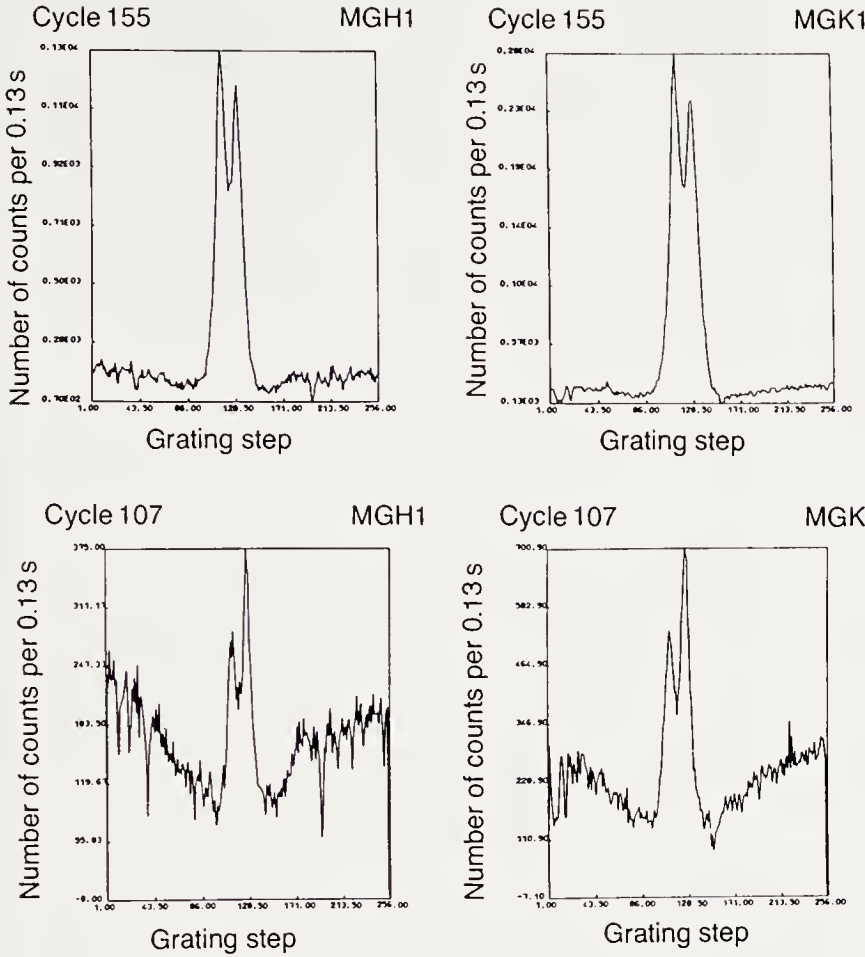
Fig. 7.5b. Mg II h and k lines observed under the same conditions as the Ca II lines. The spectral resolution is 0.025 Å. (Ph. Lemaire, private communication)

coronal loops to regions of closed magnetic field, while coronal holes correspond to regions with open fieldlines and one dominant polarity. The X-ray bright points can be linked to their Ly α counterparts down to the level of the supergranules (lower chromosphere). This identification of the magnetic configuration as a function of altitude is based on models of the magnetic field which assume that at the photosphere the field has the orientation and strength actually measured at a large number of points on the surface, and that it extends as a potential field (zero-current model). Figure 7.8 shows such a model of the coronal field.

The *Einstein* satellite (1980) showed that most stars are also X-ray sources in the 0.4–3 keV range. The X-ray luminosity in this spectral band, called L_x, is in many stars much higher than that of the Sun. It reaches 10^{33} erg s^{-1}, while the Sun, if it were totally quiet, would emit $\sim 5 \times 10^{27}$ erg s^{-1}, and if totally covered with active regions, $\sim 2 \times 10^{29}$ erg s^{-1}. The detection threshold for *Einstein* was 10^{27} erg s^{-1}.

The main properties of stellar X-rays, as revealed by *Einstein*, are as follows (Vaiana et al. 1981):

Normalised flux
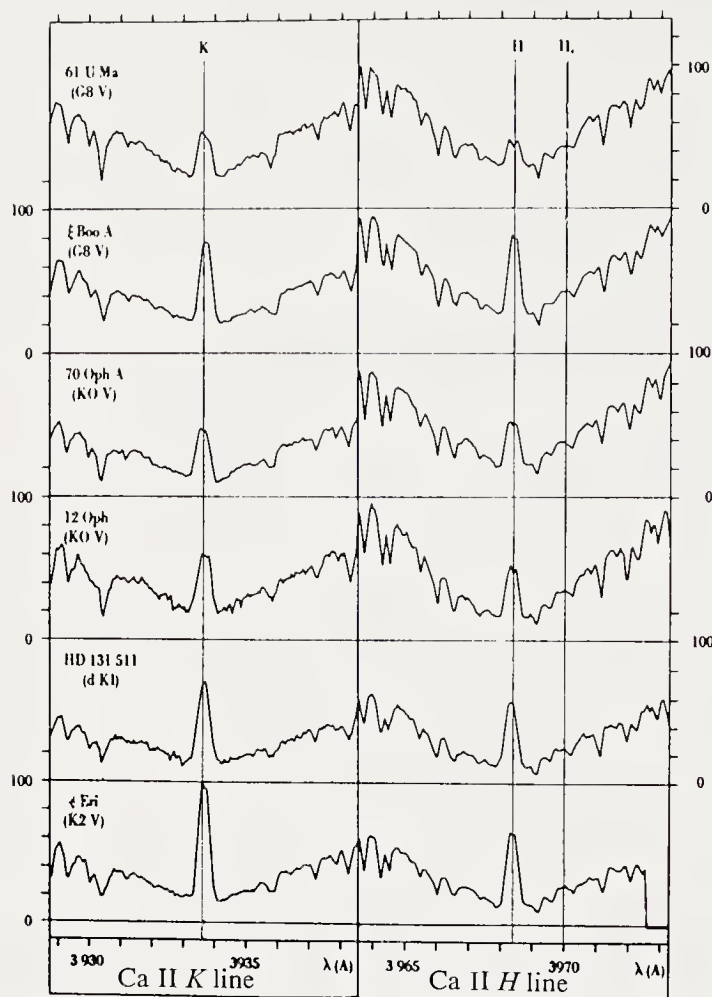
Fig. 7.6a. Ca II H and K lines observed for a series of solar-type stars. The spectral resolution is 0.12 Å. (From J.L. Linsky et al., *Astrophys. J. Supp.* **41**, 47 (1979). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1979 The American Astronomical Society)

— Stars of every spectral type and luminosity class emit X-rays, with the exception of cool supergiants. A-type stars have the lowest luminosity L_x.

— The ratio of the X-ray luminosity to the bolometric luminosity is constant ($L_x/L_{bol} = 10^{-7}$) for hot stars (O and B).

— The X-ray luminosity is enhanced and practically constant for F0–F5 stars. Stars cooler than late F have L_x varying by several orders of magnitude (typically 3) for a given spectral type.

— Dwarf M stars emit a large fraction of their luminosity in X-rays, with L_x/L_{bol} of the order of 0.1.

The *EXOSAT* satellite (1983) extended the observed energy range down to 0.04 keV, and confirmed the results of *Einstein*; it also carried out several studies of time-varying X-ray emission from "normal" stars (we do not
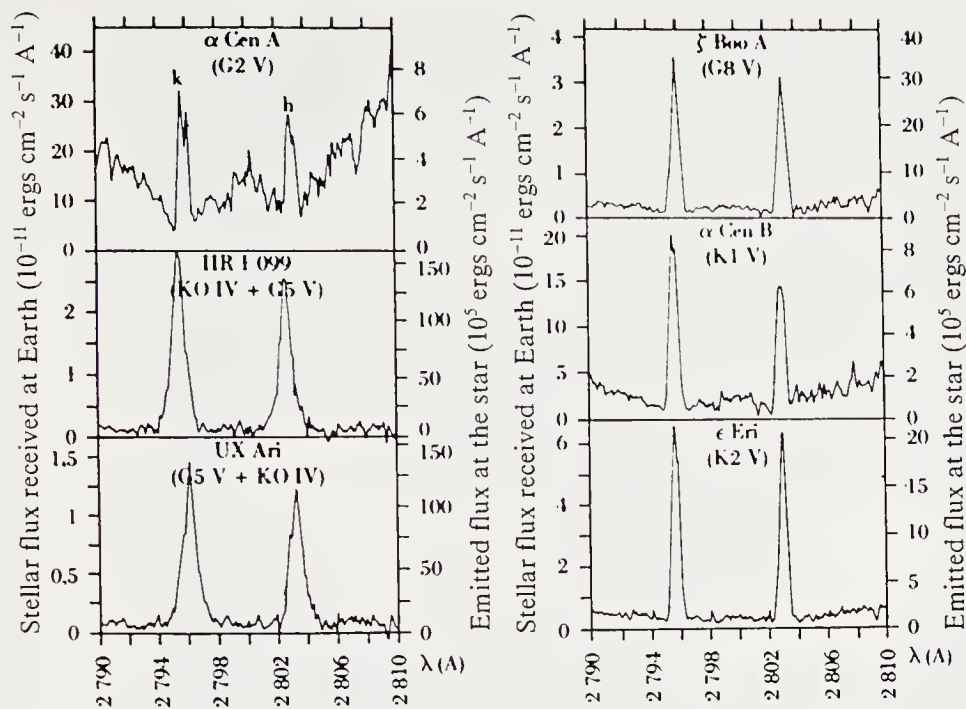
Fig. 7.6b. Mg II h and k lines observed by the IUE satellite. The spectral resolution is 0.2 Å. (From G. Basri and J.L. Linsky, *Astrophys. J.*, **234**, 1023 (1979). Reproduced by kind permission of *The Astrophysical Journal* published by The University of Chicago Press; © 1979 The American Astronomical Society)

consider here binaries with a compact component, possibly possessing an accretion disc). Like the Sun, stars are found to be variable X-ray sources when one can measure them. The *ROSAT* satellite (1990) works in the range 0.1–2 keV and has better sensitivity than *Einstein* and *EXOSAT*.

7.1.4 Radio Emission

The Sun emits radio waves which are detectable from Earth (from wavelengths of a few millimetres to several tens of metres), or from space (wavelengths above 30 m). There is a thermal component (bremsstrahlung radiation from 1 eV electrons), but the radio emission from active regions requires high-energy electrons (10 keV to 10 MeV). These are accelerated in magnetic fields and spiral around the fieldlines. In some cases the magnetic field causes anisotropies in the electron distributions, leading to coherent emission. Radio emission from active solar or stellar phenomena is via cyclotron and gyroresonant processes if the electrons are non-relativistic, and synchrotron emission if they are relativistic. Coherent plasma emission, which is independent of the magnetic field, can also occur (see Dulk 1985).

Solar radio emission has several components at wavelengths greater than a few cm; all of these originate in the corona, at various distances from the optical limb of the Sun. The components are:
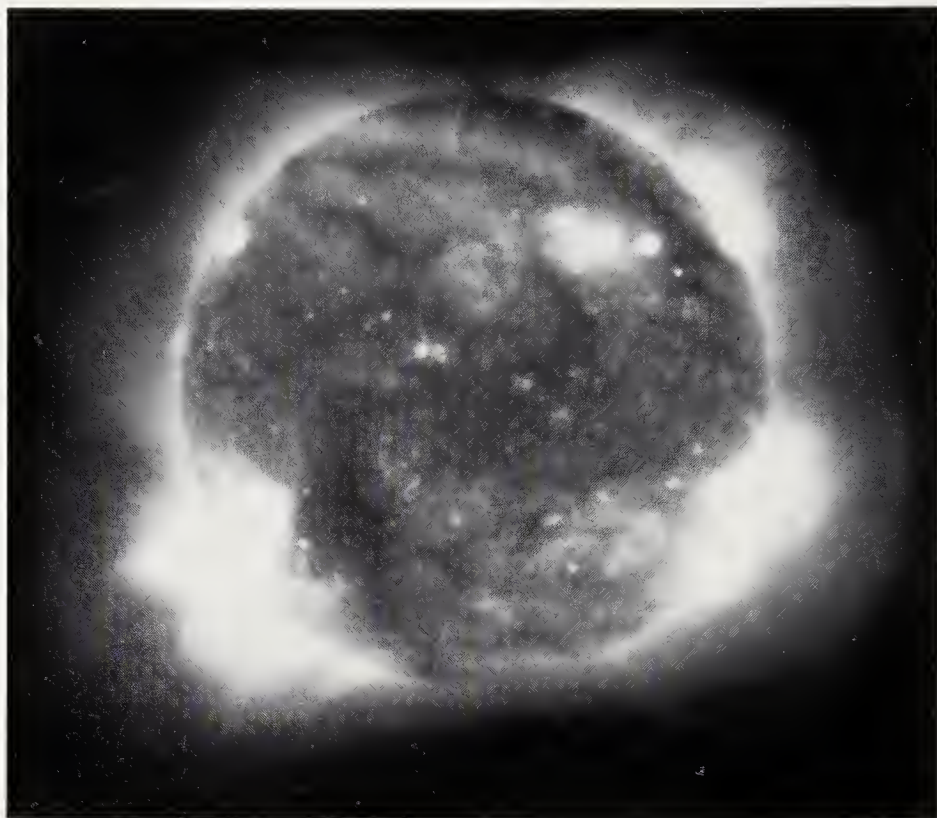
Fig. 7.7. Image of the Sun in X-rays. Several active regions are visible at the limb. The disc itself shows coronal loops in emission and quite a large number of bright points (photograph by D. Moses, American Science and Engineering, obtained from a rocket flight on 1987 December 11)

— A continuum which at $\lambda = 10$ cm is completely correlated with the spot number and shows slow periodic variations following the 11-year cycle. This emission is via bremsstrahlung or gyroresonant processes or both.

— Sporadic rapidly varying emission associated with flares (see Sect. 7.3). It is observed at all wavelengths above active regions where flares occur.

— A burst-type component, whose brightness temperature reaches 10^7 to 10^9 K and which is highly polarised.

High-resolution maps of the Sun at 169 MHz (or 1.77 m) again show the heterogeneity of the corona (Fig. 7.9) and its variability with the activity cycle.

Stellar radio emission was almost undetectable until the advent of sensitive (mJy) telescopes capable of high angular resolution ($1''$ to $10''$). The VLA (Very Large Array) in New Mexico began operating in 1982 and found that stars of all spectral types are radio sources at centimeter wavelengths. Figure 7.10 shows a detection of the active star χ^1 Ori (G0V) at 6 cm.

Like other solar and stellar activity indicators, radio emission is time-variable. In the Sun the main timescales are the periods of the activity cycle

CL = 103°

N   N   N   N   N

Pre-mission

S   S   S   S   S

24 May   21 June   18 July   14 August   10 September
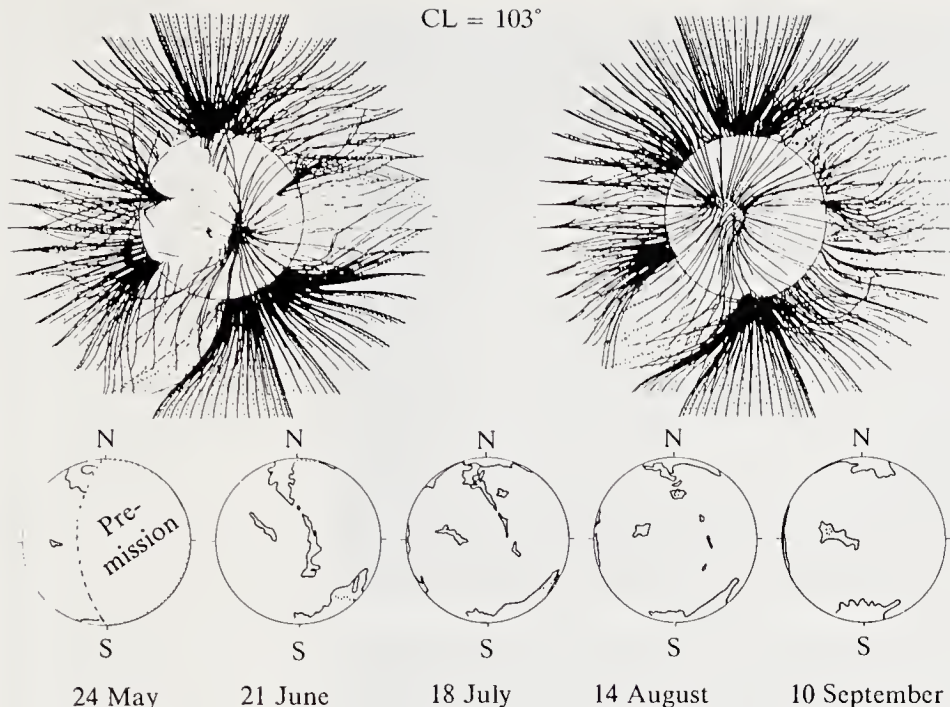
Fig. 7.8. *Top*: the magnetic field calculated for the solar corona. The open fieldlines are calculated from data two revolutions apart. *Bottom*: coronal-hole boundaries centred at the same longitude (103°), as observed by *Skylab*. The calculations correspond to 1973 June 21 (*left*) and 1973 August 14 (*right*). (From R.H. Levine et al., *Astrophys. J.* **215**, 363 (1977). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1977 The American Astronomical Society)

and of the rotation. Characteristic timescales have not yet been established for other stars.

7.2 Timescales of Magnetic Variability of the Sun and Stars

7.2.1 Rotational Modulation of Activity Indicators

We have already seen that starspots are *detectable* in photometry because their asymmetric distribution on the surface causes a rotational modulation of the star's light.

Similar modulations are seen in the intensity of other activity indicators. One is therefore led to assume that these indicators come from active regions which are asymmetrically distributed on the star's surface. This reasoning relies on the solar analogy (Fig. 7.11): this figure shows monochromatic images of the Sun in the K_3 line separated by about one half of the synodic rotation period $P = 27.25$ d. It is clear that one of the Sun's hemispheres has more active regions than the other, and that these active regions persist
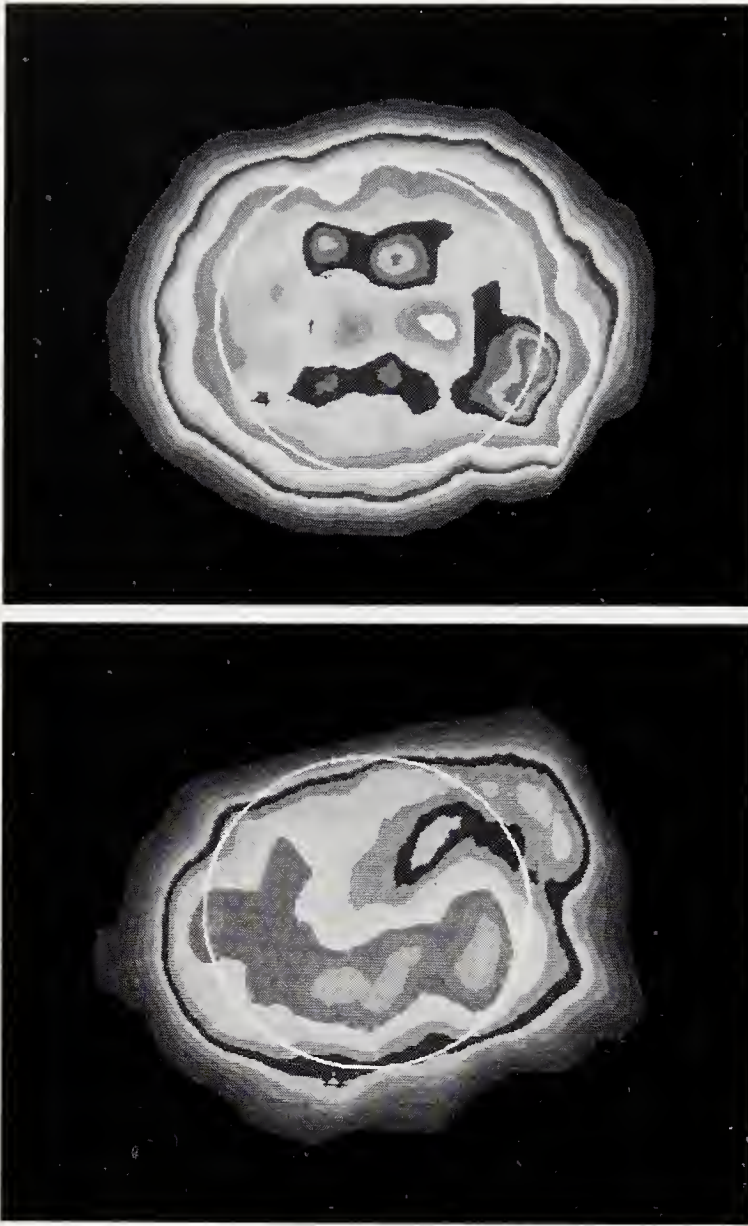
Fig. 7.9. Maps of the Sun at 169 MHz (1.77 m). Images from the radioheliograph at Nançay. The *white circle* represents the visible solar disc: North is at the top and East at the right. *Top*: (1980 July) appearance of the Sun near solar maximum. We note an equatorial coronal hole. *Bottom*: (1984 July), appearance of the Sun near solar minimum. The emission is decreased near the poles by coronal holes. (P. Lantos, private communication)

for several rotations (5 or 6). If the active phenomena varied more rapidly than the rotation period there would be no detectable rotational modulation of the emission: this is true of flares for example.

The necessary longitude asymmetry and persistence of active regions over several rotations certainly holds in many stars. For example, the disc-integrated Sun shows a rotational intensity modulation of all the X-ray, UV, and visible activity indicators. Similarly, Fig. 7.12 shows a sample from observations at Mount Wilson of the Ca II K line in 99 late-type stars
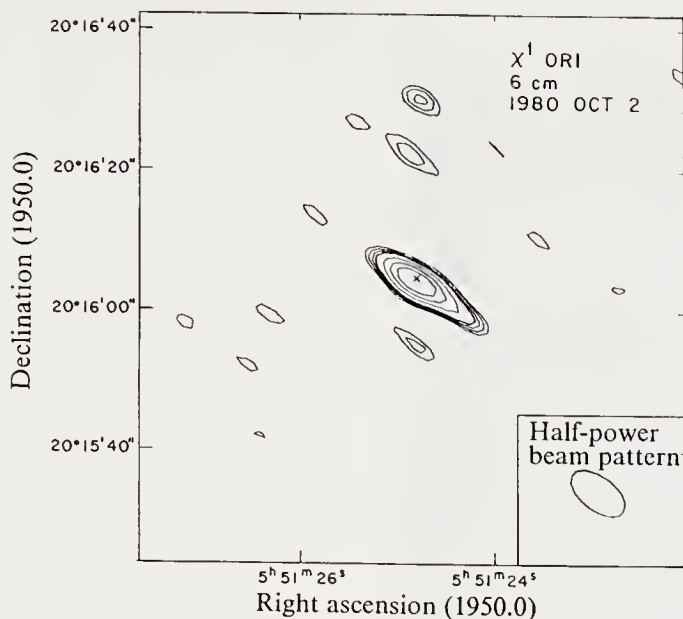
Fig. 7.10. Detection of the 6 cm radio continuum of the stars χ^1 Ori (G0V) by the VLA. The optical position of the star is marked by a *cross*. (From D. Gary and J.L. Linsky, *Astrophys. J.* **250**, 284 (1981). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1981 The American Astronomical Society)

(Vaughan et al. 1981). In this systematic programme each star was observed once per night. The observed emission variations with time give the rotation period of the star directly. For solar-type stars these periods lie between 1 and 40 days.

7.2.2 Stellar Activity Cycles

These are exhibited in spotted stars and in late-type stars, mostly on the main sequence. The distinction between these two groups of stars is not fundamental but introduced by the different modes of detection: the cycles are seen in white light in the first case and in Ca II emission lines in the second.

Using photographic-plate collections one can study the brightness variations of certain stars over more than 100 years. In this way a period of 60 years has been assigned to the K5e dwarf BD + 26°730 by Hartmann et al. (1981).

In the disc-integrated Sun, the fluxes of the H and K emission lines vary with amplitudes of the order of 20 % through the eleven-year sunspot cycle. Similarly, variations of the Ca II K-line chromospheric activity indicator have been followed at Mount Wilson Observatory since 1966 for 91 late-type main-sequence stars (types F to M). Figure 7.13 shows the great variety of long-term behaviour of these variations. In 24 years 85 % of these stars were found to vary; 60 % of them are apparently or definitely periodic. Several show chaotic fluctuations: they are the youngest stars of the sample. The
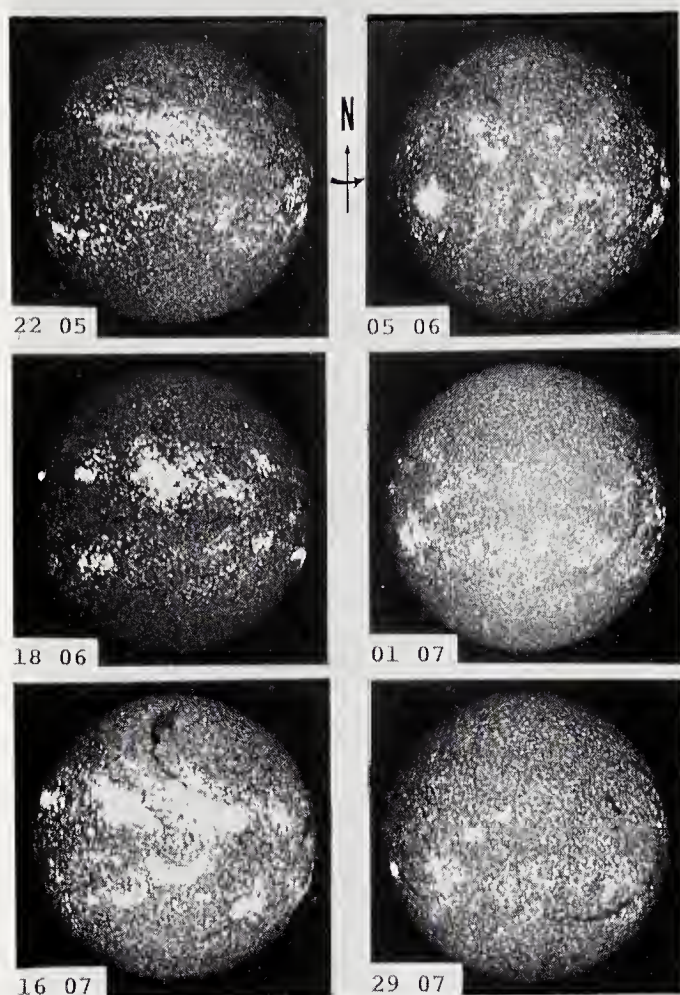
Fig. 7.11. Spectroheliograms in the K_3 component of the Ca II K line over 3 solar rotational periods (1982 May–July). Each image is separated from the previous one by one half of a rotation period. The hemisphere seen in the left-hand column is more active than that seen at right. 1982 was a solar maximum. A very large active region developed towards mid-1982 and lasted for several rotations. (M.J. Martres, private communication)

10–15 % which are constant in time may be in a state like the Maunder minimum experienced by the Sun in the 17th century. The cycles are in general 8 to 12 years long, i.e. similar to the solar cycle. Some of the stars show several periods; some have periods less than 5 years, very different from the Sun. The periods found (on a baseline of 20 years only) do not depend on the mass, the rotation period, the age, or any basic stellar characteristic.

These activity cycles are not immediately identifiable with the solar cycle, since in stars we do not yet have the possibility of detecting a *magnetic* cycle in flux or polarity. By analogy with the Sun, however, they are very important, in that the observed periods and modes of the cycles imply empirical constraints on dynamo theories of magnetic field generation (see Chap. 5).
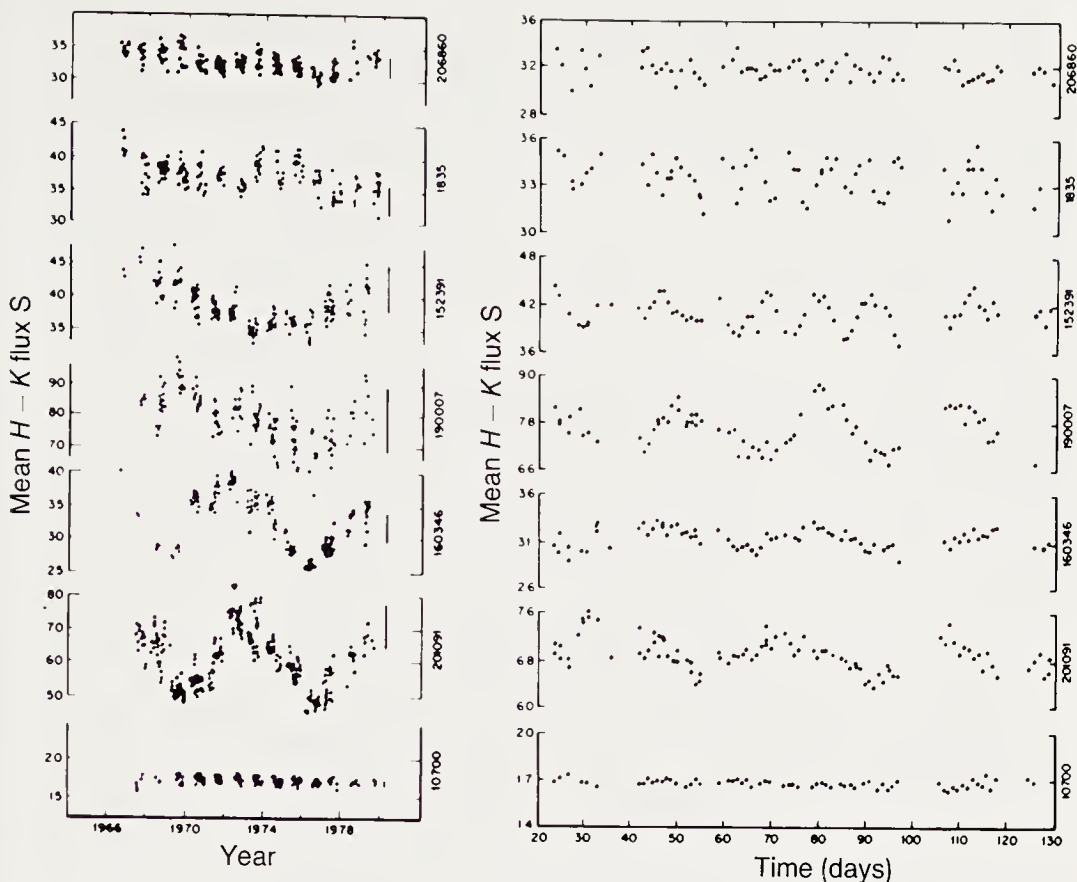
Fig. 7.12. *Right*: rotational modulation of Ca II K line emission. The average flux in the emission of the H and K lines is plotted against time. *Left*: long-term variations of the same quantity, from Wilson (1978). The stars are specified by their HD numbers. From Vaughan et al., *Astrophys. J.*, **250**, 276 (1981). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1982 The American Astronomical Society)

7.3 Solar and Stellar Flares

7.3.1 Solar Flares

Flares are the most abrupt and transient of variable solar phenomena. They are observed as intense brightenings of the Hα line, in white light, and most of all in soft and hard X-rays, radio, and the UV. Flares erupt and develop (impulsive phase) in one to two minutes, then decrease (gradual phase) over several tens of minutes. They are quite localised on the solar surface and appear along the neutral line separating two regions of opposite magnetic polarity inside an active region.

Their frequency varies with the solar cycle: they are more numerous (2 per hour on average) near solar maximum.

It is assumed that the energy of a flare ($\sim 10^{31}$–10^{32} erg) is magnetic in origin; the magnetic field dissipates and releases energy $B^2/8\pi$ per unit
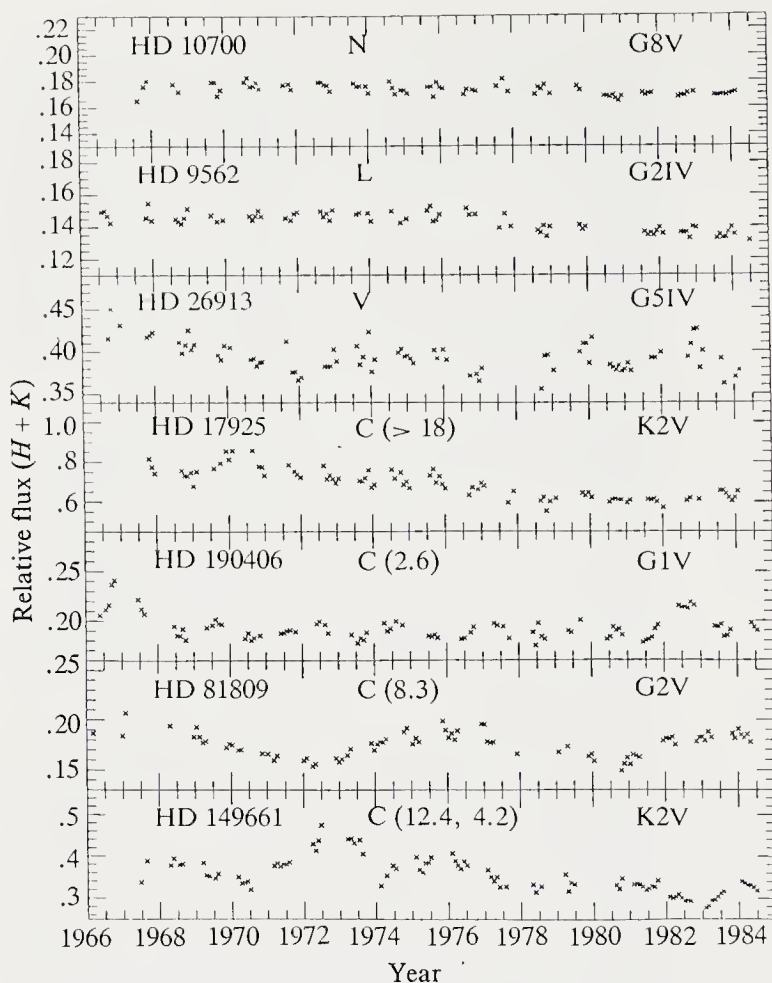
Fig. 7.13. Stellar activity cycles. 7 dwarf stars whose chromospheric activity was followed for 18 years. Meaning of symbols: N – no variation; L – linear trend; V – variable without clear period; C – periodic or probably periodic cycle. The periods are given in years in brackets. (From S. Baliunas, in *Advances in Space Research* **6**, 8 (1986) p. 231 (Pergamon), reproduced by kind permission of the Committee for Space Research COSPAR)

volume by the Joule effect. When a flare is triggered, electrons and ions are accelerated. The accelerated electrons emit the X-rays and radio radiation seen after the impulsive phase. In the gradual phase the emission is probably thermal; the emitting plasma has a temperature of the order of 10^7 K. The accelerated electrons and ions (a few tens of keV for the electrons, MeV to GeV for protons) are ejected from the Sun and propagate in interplanetary space.

7.3.2 Stellar Flares

Like solar flares, stellar flares are transient events, developing in 1 to 1000 seconds and declining over 1 to 100 minutes. By analogy with the Sun, they result from the magnetic variability of the stars. But unlike the variations considered in Sect. 7.2 they do not as yet appear to show any periodicities.

They were first detected in U-band photometry. This showed the existence of two classes of flare stars: the UV Ceti and BY Dra stars, both active as defined in Sect. 7.1. The UV Ceti stars are M dwarfs with emission lines (dMe), and the BY Dra stars are binaries of type K or M with spotted surfaces.

During flares in the optical continuum there is also a strengthening of the Balmer lines of H I and of the He I lines; the enhanced emission persists for several hours, then decreases more slowly than the emission observed in the broadband continuum.

The *Einstein* and *EXOSAT* satellites observed flares in the X-ray domain, and in some cases simultaneous *UBVRI* photometry was obtained. In BY Dra the peak of the visible emission precedes the X-ray peak by 4 to 8 minutes. In addition, the optical flare develops more rapidly than the X-ray flare (respective growth times are 1 and 10 min, decay times 5 min and 1 hour).

Study of several dMe stars simultaneously in the visible and X-rays gives a ratio E_{opt}/E_x of emitted energies between 0.1 and 1.

Since flares have become detectable in all wavelength regions (X, UV, radio) other classes of active stars have been found to have flares: the evolved RS CVn binaries, consisting of a solar-type dwarf and a G or K subgiant; the pre-main-sequence T Tauri stars; and giants and supergiants such as α Ori (M2 I ab), π Aur (M3 II) and R Aql (gM5e–8e).

Table 7.2 summarises the parameters obtained from an analysis of the X-ray and UV data from the Sun and from dMe and RS CVn stars (Byrne 1989). We note that stellar flares each emit a total energy which can exceed that of even a large solar flare by factors of 100 for a dMe star, and 10^4 to 10^5 for an RS CVn star. The mechanisms responsible for stellar flares are probably of magnetic origin, as in the Sun, but the reasons for their much greater efficacy are not understood.

Table 7.2. Parameters of solar and stellar flares in the X-ray and UV regions

	dMe stars	RS CVn stars	Sun
Emission measure (EM) in soft X-rays (cm^{-3})	10^{51}–10^{53}	10^{53}–10^{54}	10^{47}–10^{50}
Temperature (K)	1–5×10^7	6–10×10^7	2–5×10^7
N_e (cm^{-3})	1–8×10^{11}	3–9×10^{11}	1–10×10^{11}
Total energy in soft X-rays (erg)	10^{31}–10^{34}	10^{34}–10^{36}	10^{28}–10^{30}
Emission measure (EM) deduced from C IV (cm^{-3})	10^{49}–2×10^{50} (?)	10^{51}–10^{54}	10^{46}–10^{49}
N_e in the transition zone (cm^{-3})	$\leqslant 10^{12}$	5–10×10^{10}	1–5×10^{10}
Total energy in the transition zone (erg)	10^{31}–10^{33}	10^{35}–10^{37}	10^{27}–10^{29}

7.4 Stellar Magnetic Fields

Can we verify that stellar activity, like solar activity, has magnetic origins? This requires *direct measurement* of the magnetic field; then it will be possible to look for correlations, for instance between the strength of an indicator and the magnetic field intensity. Where such correlations exist they can be used to find the field *indirectly*.

7.4.1 Direct Measurements of the Magnetic Field

Most of these use the Zeeman effect, which is seen in three ways: (a) separation of a line into distinct Zeeman components, whose separation can be measured; (b) polarisation, generally elliptical, of the light emitted in the Zeeman σ components; (c) linear polarisation of the light emitted in the Zeeman π components. Figure 7.14 shows a decomposition of the field into a component $B_{\parallel}$ along the line of sight and a component $B_{\perp}$ orthogonal to it. If $\theta = 0$ (longitudinal field), only the σ components exist, and they are circularly polarised. If $\theta = \pi/2$ (transverse field), the σ and π components are linearly polarised.
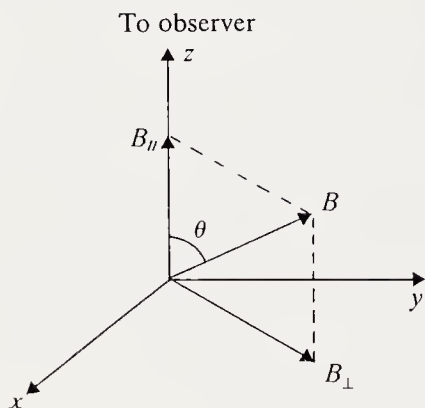
Fig. 7.14. Decomposition of the magnetic field B into longitudinal ($B_{\parallel}$ or B_z) and transverse ($B_{\perp}$) components

The Zeeman effect causes a separation $\Delta\lambda_B$ between the central group of π components and the centroid of the σ^+ or σ^- components (polarised respectively right and left circular):

$$\Delta\lambda_B(\text{Å}) = 4.67 \times 10^{-13}\lambda_0^2 g B, \tag{7.2}$$

where λ_0 is the wavelength of the line (in Å), g the Landé factor, and B the field strength (in G) of the magnetic field.

For a line at $\lambda_0 = 6200$ Å, with $g = 2.5$, $\Delta\lambda_B$ corresponds to a velocity displacement of $\Delta v = c\Delta\lambda_B/\lambda_0 = 2.2(B/1000\,\text{G})$ km s^{-1}. Δv is of the same order as the rotational or turbulent line width; this makes the measurement of stellar magnetic fields very difficult even if $B = 1000$ G. As a result very few stars (apart from the Sun in sunspots) have fields strong enough

326

and sufficiently coherent over large scales for the Zeeman splitting to be measurable in unpolarised light. This can be done for some Ap stars and white dwarfs.

In cases where the field has an arbitrary local orientation we can measure it by either Babcock's or Robinson's methods. The first uses the measurement of circular polarisation in the wings of absorption line. The light from the star in a magnetically sensitive line (i.e. large Landé factor) passes through a polarisation analyser, formed from a quarter-wave plate with its neutral line inclined at 45° to the slit of the spectrograph, followed by a Nicol (Iceland spar birefringent polariser). The Nicol is oriented so that the splitting of the emergent ordinary and extraordinary rays, corresponding to two images of the star with orthogonal polarisations, is parallel to the slit.

We thus get a spectrum made up of two bands, one showing the right circularly polarised Zeeman components and the other the left circularly polarised ones. Since in general the line of sight does not coincide with the field direction, the emergent σ components have intensities depending on θ (Fig. 7.14). Seares (1913) showed that

$$I_\sigma = \tfrac{1}{4}(1 \pm \cos\theta)^2 \,, \tag{7.3}$$

while measurement of the π component gives $I_\pi = \tfrac{1}{2}\sin^2\theta$. The splitting $\Delta\lambda_B$ is proportional to the average field intensity projected along the line of sight, or *effective field*

$$B_{\mathrm{eff}} = \int B\cos\theta I\,\mathrm{d}s \Big/ \int I\,\mathrm{d}s = \int B_\parallel I\,\mathrm{d}s \Big/ \int I\,\mathrm{d}s \,, \tag{7.4}$$

where $\mathrm{d}s$ denotes a surface element of the stellar disc through which the field is $\boldsymbol{B}$.

Babcock's method gives good results for Ap stars since the field geometry is simple, either dipolar or quadrupolar. The magnetic dipole can either be aligned or inclined to the rotation axis, and the measured field varies with the rotation period. For these stars, the precision of the method has been brought to about 5–10 G by Brown and Landstreet (1981), who measured the polarisation in 1500 lines simultaneously.

Measurement of linear polarisation (in general the rate of polarisation) is difficult because its value is very low. However, such measurements give very useful information on the field geometry.

It is clear that the detection of a polarised signal by itself reveals the presence of a magnetic field. It is equally clear that for a solar-type geometry, i.e. with many zones of opposite polarity visible on the stellar disc, there will be cancellation of the polarised signals, and the net result will be close to zero if one observes the integrated disc. In solar-type stars Babcock's technique almost always fails.

Robinson's method uses two photospheric lines as similar as possible (same element, same degree of ionisation, same multiplet, and same oscil-

lator strength) except for the Landé factor. The line with smaller g is little affected by the field, while that with large g is broadened in magnetic regions, by an amount proportional to λ^2 and the modulus of $\boldsymbol{B}$, whatever the polarity of the region. We thus compare the half-width of these two lines in unpolarised light and, following Robinson (1980), postulate that the profile of the broadened line is the convolution of the unbroadened line with a function of the area occupied by magnetic regions on the stellar surface. These regions are assumed to have a field of the same strength. The observed line profile F_{obs} is interpreted in a two-component model, i.e. $F_{obs} = fF_m(B) + (1-f)F_q(B=0)$ where F_m is the profile arising from magnetic regions covering a fraction f of the surface, and F_q is the profile arising from non-magnetic quiet regions. The derivation of F_m and F_q gives the value of $|B|$ and the filling factor f, i.e. the relative area covered by magnetic regions.

This method can be used in the visible or near IR, where the factor λ^2 amplifies the Zeeman effect. It requires very high signal-to-noise spectra of very high spectral resolution, that the star does not rotate too rapidly ($v \sin i < 10$ km s^{-1}), that the lines are not blended, but above all that we can find pairs of lines obeying the severe conditions of identical properties up to the factor g. These conditions should ensure that the lines are formed in the same atmospheric layers, whether or not the region is magnetic. It so happens that this is rarely the case.

The results existing in 1990 for the fields of some 30 slowly rotating solar-type stars do however result from this method, which is difficult to apply (Saar 1988).

(a) On the main sequence, fields are detected in stars of spectral types between G0V and dM 3.5e, with rotation periods between 3 and 15 days.

(b) There have been no detections for solar-type giants, except for the RS CVn star λ And.

(c) The detected fields have intensities from 600 to 3000 G; in the most active stars these values are exceeded: 2500 ± 300 G in EQ Vir (dK 5e), 3800 G in AD Leo (dM 3.5e).

(d) Too few stars have been measured to establish a definitive relation between the fieldstrength B and the emission in the Ca II K line: Marcy (1983) finds $L_{CaII} \sim |B|^\alpha$ with $\alpha = 0.5$, differing from the quiet Sun where $L_{CaII} \sim |B_z|$, with B_z the local longitudinal component of the field (Skumanich et al. 1975), but close to the relation established by Schrijver et al. (1989) for active regions of the Sun, where they find $\alpha = 0.6$.

The Zeeman–Doppler imaging method (Sect. 7.4.4) works for rapidly rotating solar-type stars.

7.4.2 Indirect Methods of Measuring the Magnetic Field

A relation between a sensitive activity indicator such as Ca II emission and the magnetic fieldstrength is particularly interesting: given the difficulty of direct field measurements in solar-type stars one is tempted to use the intensity of a suitably chosen line to deduce $|B|$. This is the method of the *spectroscopic magnetograph*. The problem is the calibration of such a relation. Calibration from the Sun has the advantage of being local, i.e. we can easily distinguish magnetic and non-magnetic regions. On the other hand, a calibration like that of Marcy explicitly assumes that $|B|$ is the same in all magnetic regions, and that the line with small g factor is completely insensitive to B.

7.4.3 Other Approaches

The Zeeman effect gives photospheric magnetic fields, either $B_\parallel$ or $|B|$. In principle we could deduce the coronal field from the radio intensity, if we have good reason to attribute the emission to a non-thermal process (e.g. if the source varies rapidly, $\tau < 10$ min, or the radiation is polarised). This method has the difficulty of requiring an assumption about the field geometry (we have to construct a model of the source) and assumes that we have identified the emission mechanism. As a result the interpretation of non-thermal radio emission is not yet very developed.

Clearly there are not enough measured fields for active stars in 1992 to allow answers to some of the questions suggested by the solar analogy: are there various lengthscales on which the field emerges? Is there a dipole component as well as a toroidal one? How does the measured fieldstrength depend on fundamental parameters such as mass, age, and rotation of the star? How does the energy balance in the outer layers of stars depend on the heating and confinement of the plasma by magnetic fields?

7.4.4 Sizes of Stellar Active Regions

A knowledge of filling factors, i.e. the area of active regions on unresolved stellar discs, is a first step towards discovering the characteristic scales of activity. The filling factor found by Marcy was about 0.16 for G0V stars, and 0.54 for stars of types K0–K5V. For F stars, whose high degree of activity is attested by their enhanced X-ray luminosity, there is no rotational modulation in white light or chromospheric lines, and no field measurement has been possible (but these stars rotate rapidly): in this case the active regions may be distributed fairly uniformly and close enough to each other so that the filling factor is quite close to 1.

There is another way of estimating the size of active regions, without relating them to the presence of a magnetic field, though. This is *Doppler imaging*. Figure 7.15 shows the basis of this method, which is derived from study of the rotational modulation. Here we examine the effect on the line
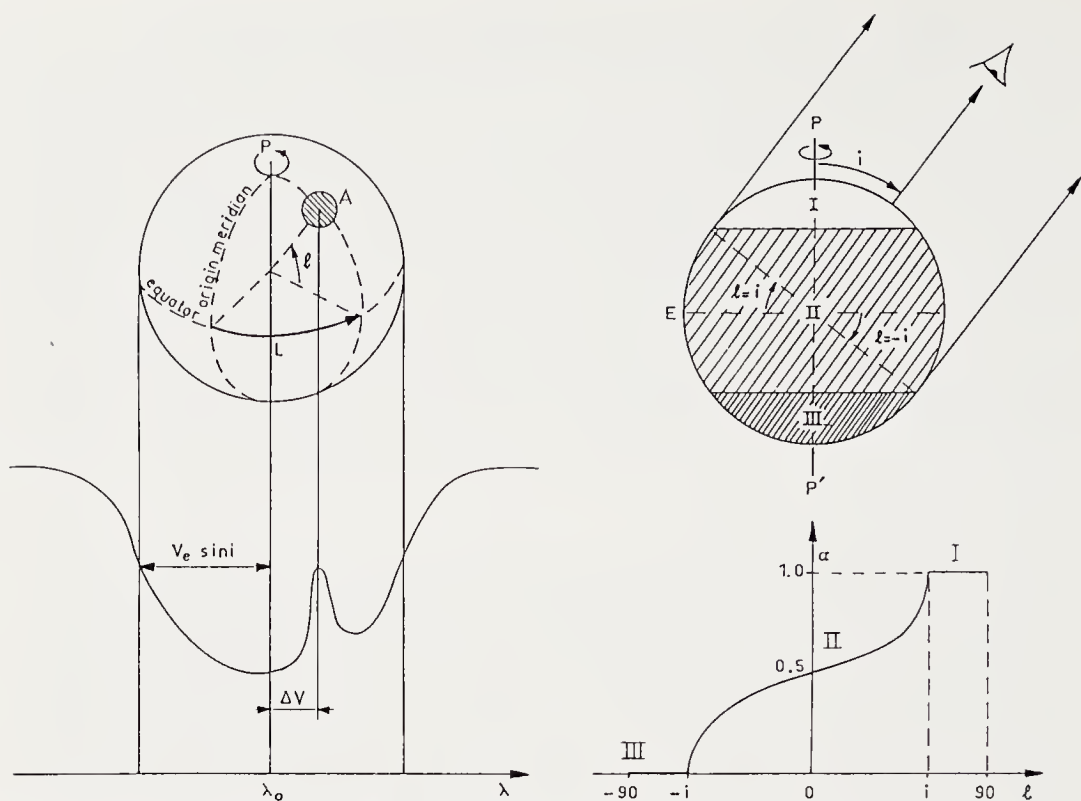
Fig. 7.15. Principle of Doppler imaging. *Left*: the star's disc has an active region A (latitude l, longitude L) and its effect ΔV on the profile of a line is a function of rotational phase ϕ; it is given by $\Delta V(\phi) = V_e \sin i \cos l \sin(2\pi\phi + L)$, where i the inclination of the line of sight to the rotation axis. *Right*: the visibility α of a region of latitude l. (From P. Gondoin, *Astron. Astrophys.* **160**, 73 (1986))

profile of the passage of spots or active regions (as in the figure) across the line of sight. If the star satisfies certain conditions on (a) the inclination of the rotation axis to the line of sight (angle i) and (b) the projected rotation velocity $V \sin i$, we can work out the contribution of regions like A to the total radiation. The contribution of A is Doppler-shifted over half a rotation (if A is the only active region we will see nothing over the next half period); moreover A produces lines and a continuum differing in intensity from those of the neighbouring photosphere. This is seen in the profile for the integrated disc of the star, as shown on the figure. The visibility function α is also shown as a function of the latitude of region A.

This method was successfully used first for Ap stars, then for active solar-type stars, both for absorption lines (photosphere) and emission lines (chromosphere). This allows one to show that the layering which occurs in the Sun is also characteristic of stellar active regions, with plages above the spots.

The complementary Zeeman–Doppler imaging method (Semel 1989; Donati et al. 1989) produces a two-dimensional map of the disc in the Stokes parameter V and has allowed the detection of magnetic regions in the active

component of the RS CVn binaries V711 Tau (HR 1099), II Peg, σ^2 Cr B, UX Aql. All these stars have $V \sin i > 20$ km s^{-1}.

7.5 Sources of Stellar Activity: Convection, Rotation, Primordial Fields (Empirical Aspects)

7.5.1 Activity Indicators on the HR Diagram

Most of the stars we have mentioned up to now as being active have been solar-type; but we noted that X-ray emission is present in all spectral types, being particularly weak for spectral type A. Does solar-type activity extend right across the HR diagram?

Stars with similar activity to that of the Sun are grouped along the main sequence from type F5 to M. Main-sequence stars from F5 to A7 are almost certainly similar to the Sun in their activity. At spectral type A7 the star α Aql (Altair, A7V) shows intense emission in the Ly α line. Hotter dwarfs have no more Ca II K or Mg II k emission, nor emission in typical transition zone lines such as the $\lambda 1550$ line of C IV.

Figure 7.16 gives an HR diagram. The shaded zone extending above the main sequence shows where evolved stars are probably active in a way similar to that of the Sun. This zone is abruptly terminated by a division (Linsky and Haisch 1979), whose existence is established by UV and X-ray indicators. To the left of this line, and in all of the shaded area, stars have an ensemble of outer layers (chromosphere – transition zone – corona) like the Sun; if they have a wind, like the Sun, it is optically thin at the same wavelengths; it is therefore not seen spectroscopically. In contrast, to the right of this line, stars have only weak X-ray emission (the temperature of the emitting plasma is less than 100 000 K). There are no lines typical of the transition zone; moreover some lines (Mg II h and k) have blue-shifted components or asymmetries indicating ejection of matter in a wind which is optically thick in at least some lines, with a low temperature ($\sim 10^4$ K, compared with the 10^6 K of the solar wind). The division thus marks a boundary for hot coronae and separates two types of wind regime.

The existence of this division for solar-type giants (luminosity classes IV to II), is interpreted as the change in structure of coronal loops as the temperature and gravity decrease. Consider a corona made up of loops of height H and temperature T. A static model of the loop is given by equating the non-radiative energy supply E (heating) and the radiative losses (Rosner et al. 1978): $E(N_e, T) = N_e^2 \Lambda(T)$, where the function $\Lambda(T)$ has the form given in Fig. 2.9. Let $H_P = kT/\mu g$ be the scaleheight at the apex of the loop. The energy equation of the loop has two classes of solution whose stability has been studied by Antiochos and Noci (1986):

SUPERGIANTS

O3 BO   AO FO   GO   KO   MO   M8

GIANTS

O3 BO   AO FO   GO   KO   MO   M8

MAIN SEQUENCE

O3 BO   AO FO   GO   KO   MO   M8

Hot Winds
X-Rays
ζ Pup   * α Car   SUPERGIANTS   * α Ori

□ α Aqr   * α¹ Sco
* HD 37017   Extended Chromospheres
* σ Ori E   Cool Winds

□ α TrA
* Vega   ○ Capella   * α Boo * α Tau
Ab
Sirius A *   Coronal and
Altair   ○ λ And   Transition Region
γ Vir N * Procyon   "Dividing Line"
* χ¹ Ori
Sun *   * ξ Boo A
· Eri *
61 Cyg A   MAIN
EQ Vir *   SEQUENCE

* Single or Widely
Separated Stars
○ RS CVn Systems   AD Leo * * Wolf 630 A
□ Hybrid Stars
Fully
Convective
▨ Definitely Solar-like   Stars

▧ Almost Certainly Solar-like

▒ Probably Solar-like

ABSOLUTE VISUAL MAGNITUDE

B - V

V - R

Fig. 7.16. HR diagram (absolute visual magnitude – colour index) showing the regions where solar-type chromospheres are detected (Adapted from J.L. Linsky, *Sol. Phys* **100**, 333 (1985). Reproduced by kind permission of Kluwer Academic Publishers)

(a) A "hot loop" solution, which for given H and E implies a maximum temperature T_m above 10^6 K and a quasi-isobaric loop. The hot-loop solution is only possible *and* stable if $\theta = H_P(10^5)/H$ is much smaller than 1.

(b) A "cold loop" solution, which retains hydrostatic equilibrium only if P and N_e decrease with height; in this case the energy equation is only satisfied if $\Lambda(T)$ is an increasing function of T, requiring $T < 10^5$ K. This case occurs if $\theta = 1$. For low-gravity stars H_P varies as R_*^2, and if we assume that the height H of the loops varies as R_*, θ grows linearly with R_*. As a result, without invoking any change in the nature of the heating across the Linsky–Haisch line, we can understand the absence of X-ray emission

to the right of the line: the coronal loops are filled with material which is cool ($< 10^5$ K) rather than hot ($> 10^6$ K).

Pre-main-sequence stars (T Tauri stars and Herbig stars) do not appear on Fig. 7.16. Classical T Tauri stars and Herbig stars are very active, but differ from active solar-type stars in having a dense wind, with a terminal velocity of order 300 km s^{-1}. The "weak-line" T Tauri stars are similar to active solar-type stars.

Figure 7.17 gives an HR diagram with the same coordinates as the previous figure, showing the stars detected by the *Einstein* satellite in the first survey. All spectral types on the main sequence are present, and we note the detection of some white dwarfs. This diagram gave the first plausible indication of the existence of X-ray emitting coronae around hot stars as well as around solar-type stars. This interpretation has however been disputed, and various production mechanisms for X-rays in instabilities and shocks in an otherwise cool wind have been proposed. However, we should
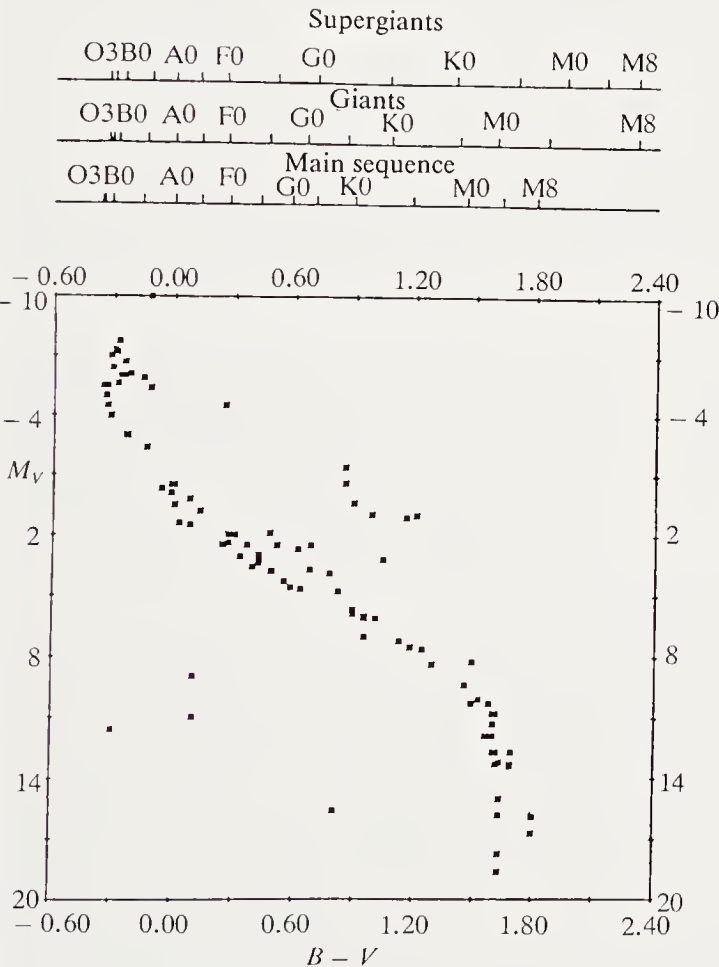
Fig. 7.17. HR diagram (absolute magnitude – $B - V$ index) showing the regions where X-ray emission has been detected. (From G.S. Vaiana et al., *Astrophys. J.* **245**, 163 (1981). Reproduced by kind permission of *The Astrophysical Journal*, published by The University of Chicago Press; © 1981 The American Astronomical Society)
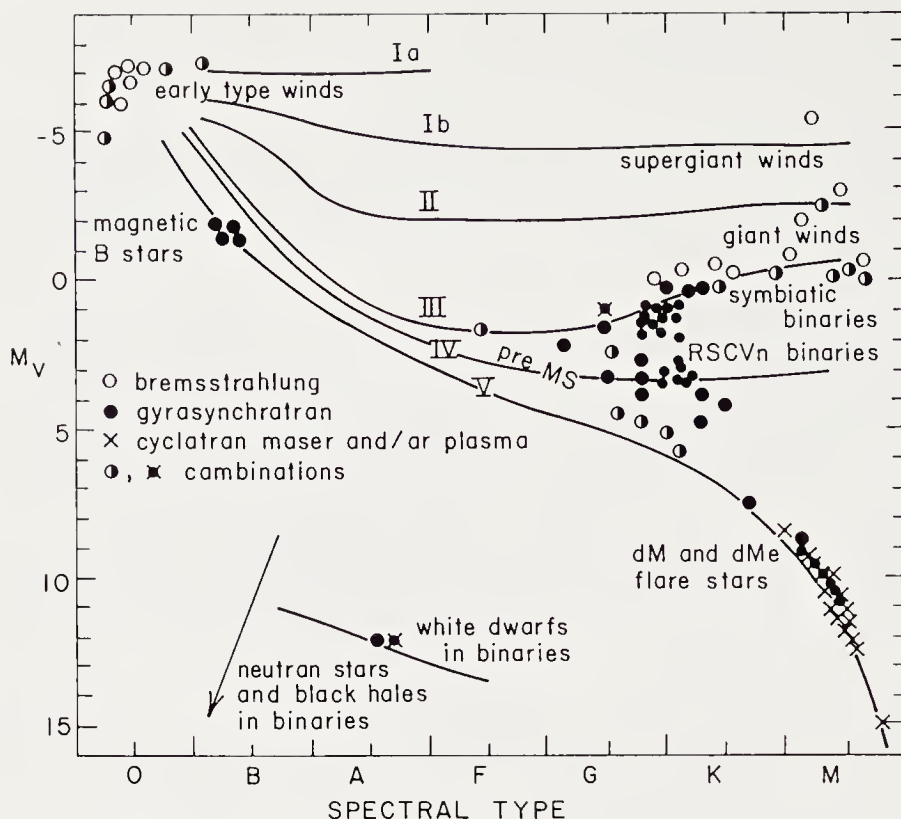
Fig. 7.18. HR diagram (absolute visual magnitude – spectral type) showing the regions where radio emission is detected from stars. (From G. Dulk, in *Advances in Space Research* **6**, 8 (1986) p. 96 (Pergamon), reproduced by kind permission of the Committee on Space Research – COSPAR)

not neglect the possibility of magnetic activity in hot stars. Uchida (1986) assumed that there are *types* of activity on the main sequence according to the generation or just the maintenance of the magnetic field. For hot stars, whose ages (10^6–10^7 y) are less than the diffusion time $\tau_D = 4\pi\sigma R_*^2/c^2$ of the magnetic field (where σ is the conductivity, R_* the stellar radius, and c the velocity of light) the dynamo mechanism need not have a regenerative phase, unlike the solar-type case.

Figure 7.18 is an HR diagram for stars detected in radio waves. What radiation mechanism should we appeal to here? Bremsstrahlung occurs in hot stars with fast dense winds and in cool giants and supergiants to the right of the Linsky–Haisch line. In O and B stars the radiation may also be non-thermal in origin; the magnetic field would then be the primordial field enclosed within the star. The star's rotation stretches the fieldlines and makes them toroidal at the surface. The field is not regenerated, in contrast to what is thought essential in solar-type dynamos. There are several O and WR stars whose radio spectral index is not compatible with purely thermal emission. These observations support the suggestion of magnetically driven activity in hot stars.

7.5.2 Parameters Influencing Stellar Activity

The Relation Between Activity Indices. We compare the luminosity in various activity indicators for active solar-type stars. We find

$L_{\text{MgII}} \sim L_{\text{CaII}}$ (holds for all chromospheric lines),
$L_{\text{CIV}} \sim L_{\text{CaII}}^{1.5}$ (holds for all transition zone lines),
$L_x \sim L_{\text{CaII}}^3$.

The same power laws hold for the Sun when we correlate the observed luminosities simultaneously observed at different altitudes in either quiet or active regions. L_{CaII} is a measure of chromospheric activity. As seen in Sect. 7.4 L_{CaII} in the Sun is proportional to $B_{\parallel}$ in the photosphere. Assuming that the calibration of this relation is generally valid, there result empirical scaling laws between the activity indicators at various heights in the outer atmosphere and the photospheric magnetic field. These laws, which hold for dwarfs or giants of spectral types F to M, imply empirical constraints on the heating mechanism for the outer layers of these stars.

The Relation Between Activity, Rotation, and Spectral Type. In contrast to other stellar properties, activity does not depend simply on the effective temperature and gravity. It is well established that activity indicators strengthen as the rotation rate Ω increases, saturating for the close W UMa binary systems. However, the activity indicators also depend on spectral type, strengthening towards later types. The mass decreases from $1.5\,M_\odot$ to $0.2\,M_\odot$ as we go from F0 to M5 on the main sequence, with a continuous increase in the thickness D of the convection zone. At a mass of $0.2\,M_\odot$ (M5) the star is fully convective. There have thus been efforts to correlate activity indicators with parameters describing both rotation and convection. We recall from Chap. 5 that, at least in solar-type stars, the magnetic field is thought to be generated by dynamo action, i.e. induction in a turbulent conducting fluid.

A useful parameter is the Rossby number Ro, the ratio of the characteristic rotation time $1/\Omega$ to the convective turnover time $\tau_c = l/V$, where l is, say, the mixing length and V a typical convection velocity. The Rossby number $V/l\Omega$ or P_{rot}/τ_c is connected to the characteristic dynamo number Dy by $Dy = Ro^{-2}$. In $\alpha - \omega$ models, Dy grows with the dynamo efficiency (Chap. 5). The intensity of the emission in the H and K lines of Ca II is well correlated with the Rossby number calculated at the base of the convection zone of solar-type stars, while the X-ray emission is correlated with an effective Rossby number, defined at the depth of the maximum convection velocity (see Fig. 5.10).

The stellar age also controls the activity level of stars, in that young stars show many intense signs of activity. However, it appears that *age is a less important factor than rotation*. This is demonstrated by the RS CVn binaries, in which one component is a sub-giant, usually rotating synchronously

with the orbit at a fairly high rate (< 10 d period). These stars are extremely active, despite their age, being comparable to the T Tauri stars in activity. Moreover, the subdwarf HD 103095, which is an old star of low metallicity, shows a magnetic cycle of large amplitude, similar to that of the Sun.

8. The Last Stages of Stellar Evolution

8.1 Minimum-Energy States

The late stages of stellar evolution require special discussion, mainly because of the occurrence of brief phases of rapid dynamical evolution. The reference objects for this discussion are stars in minimum-energy states: white dwarfs, neutron stars, and possibly black holes. The endpoints are reached once these objects have exhausted all their reserves of gravitational and nuclear energy and can radiate no more.

In practice these objects are observed directly or indirectly by the radiation they emit, either in the last stages of their evolution, or in their interaction with their environments (binary systems). In the latter case the observable phenomena reflect either (a) the conversion of energy into radiation in the outer layers of these stars, involving processes in rarified media (cataclysmic variables) or (b) surface phenomena, occurring at optical depths large enough to allow a treatment similar to that of the internal structure (gamma-ray bursts).

It thus seems advisable to take as reference objects stars of minimum energy and zero temperature.

There are two possible approaches. The first is to use the equation of state to find the (M, R) relation to order of magnitude, and deduce the main properties of these zero-temperature stars; the other is to solve exactly the hydrostatic-equilibrium equation to get the best (M, R) relation possible, and in particular the maximum masses of white dwarfs and neutron stars. The first method is useful in making clear the physics underlying the properties of dense and ultradense stars; the second method is of course indispensable for accurate comparison with observational data.

8.2 The Physics of Minimum-Energy States

8.2.1 The Equation of State ($T = 0$)

We require the properties of the equation of state $P = P(\rho)$ at zero temperature, now including effects neglected in Chap. 3. The equation of state of a degenerate electron gas (Fermi gas) has to be corrected for microscopic (mainly nuclear) effects and those of neutronisation and general relativity.

An ultrarelativistic gas of free particles of mass m has pressure P and internal energy U given by

$$P = \frac{mc^2}{\lambda^3} f(x),$$ (8.1)

$$U = \frac{mc^2}{\lambda^3} g(x)$$ (8.2)

with the particle number density

$$n = \frac{8\pi p_F^3}{3h^3} = \frac{x^3}{3\pi^2 \lambda^3}$$ (8.3)

and the usual definitions

$$\lambda = \frac{\hbar}{mc}$$ (8.4)

(the Compton wavelength) and

$$x = \frac{p_F}{mc}$$ (8.5)

(with p_F the Fermi momentum), as well as the expressions

$$f(x) = \frac{1}{8\pi^2} \left\{ x(1+x^2)^{1/2} \left(\frac{2x^2}{3} - 1 \right) + \ln \left[x + (1+x^2)^{1/2} \right] \right\},$$ (8.6)

$$\lim f(x \gg 1) = \frac{1}{12\pi^2} \left(x^4 - x^2 + \frac{3}{2} \ln 2x \ldots \right),$$ (8.7)

$$g(x) = \frac{1}{8\pi^2} \left\{ x(1+x^2)^{1/2}(1+2x^2) - \ln \left[(x + (1+x^2)^{1/2} \right] \right\},$$ (8.8)

$$\lim g(x \gg 1) = \frac{1}{4\pi^2} \left(x^4 + x^2 - \frac{1}{2} \ln 2x \ldots \right).$$ (8.9)

These relations apply equally for free electrons and free neutrons. We will use them later to show the existence of a maximum mass for white dwarfs and neutron stars.

As the density increases, the energy of electrons of momentum p_F

$$E = (p_F^2 c^2 + m_e^2 c^4)^{1/2}$$ (8.10)

may become large enough to allow inverse β reactions, electron captures of the form

$$^A_Z X + e^- \rightarrow ^{\ \ A}_{Z-1} X + \nu.$$ (8.11)

Table 8.1. Neutronisation thresholds

Inverse β reactions	Neutronisation energy (^{a}) (MeV)	Density (g cm^{-3})
$^1_1H + e^- \rightarrow \, ^1_0n + \nu$	0.782	$1.22.10^7$
$^{32}_{16}S \rightarrow \, ^{32}_{15}P \rightarrow \, ^{32}_{14}Si$	1.710	$1.47.10^8$
$^{56}_{26}Fe \rightarrow \, ^{56}_{25}Mn \rightarrow \, ^{56}_{24}Cr$	3.695	$1.14.10^9$
$^{28}_{14}Si \rightarrow \, ^{28}_{13}Al \rightarrow \, ^{28}_{12}Mo$	4.643	$1.97.10^9$
$^{24}_{12}Mg \rightarrow \, ^{24}_{11}Na \rightarrow \, ^{24}_{10}Ne$	5.513	$3.16.10^9$
$^{20}_{10}Ne \rightarrow \, ^{29}_9F \rightarrow \, ^{20}_8O$	7.026	$6.21.10^9$
$^{16}_8O \rightarrow \, ^{16}_7N \rightarrow \, ^{16}_6C$	10.419	$1.90.10^{10}$
$^{12}_6C \rightarrow \, ^{12}_5B \rightarrow \, ^{12}_4Be$	13.370	$3.90.10^{10}$
$^4_2He \rightarrow \, ^3_1H + n \rightarrow 4\,n$	20.596	$1.37.10^{11}$

(^{a}) After subtraction of the electron rest-mass energy.

The opposite β-decay reactions are forbidden because all the transla-
tional electron states are occupied. The equilibrium is thus displaced to-
wards the formation of neutron-rich elements, with a decrease in the elec-
tron number. If Q is the energy requirement for the reaction (8.11), the
required Fermi momentum is $p_F = (Q^2 - m_e^2 c^4)^{1/2}/c$. Equation (8.3) then
gives the density at the capture threshold.

A realistic study of degenerate electron capture by nuclei must take
account of the chemical composition. Table 8.1 gives the neutronisation
thresholds for the most abundant elements in the interior of white dwarfs,
arranged in order of increasing density, and calculated from the capture
threshold of an electron for an isolated nucleus. We shall see later the sig-
nificance of these for white dwarfs.

These capture thresholds do not suggest thermodynamic equilibrium
between the electrons and nuclei. Thermodynamic equilibrium must be at-
tained in the minimum-energy state: physically this can only occur in a
zero-temperature medium through neutron-producing reactions of the type

$$^{12}C + \, ^{12}C \rightarrow \, ^{23}Na + n \tag{8.12}$$

or at high temperatures, where reactions of the type (p, e^-, ν, n) can occur.

In principle the study of minimum-energy states uses the following
method. Let the mass of a nucleus containing Z protons and $A - Z$ neu-
trons be $M(A, Z)$. If n_N, n_n, n_e are the number densities of nuclei (A, Z),
neutrons, and free electrons, the total energy per unit volume is

$$U = n_N M(A, Z)c^2 + (U_e - n_e m_e c^2) + U_n + U_L, \tag{8.13}$$

where U_L is the energy of the neutron lattice (which at absolute zero becomes a centred cubic lattice), U_n the energy of the neutron gas (including the rest-mass energy of the neutrons), and U_e the energy of the electron gas. Under the constraint of conservation of baryon number

$$n(\text{baryons}) = A n_N + n_n \,, \tag{8.14}$$

we have to minimise the energy,

$$\frac{\partial U}{\partial (n_n/n)} = 0 \,, \tag{8.15}$$

and, using the equation of state and the pressure of the Fermi gases, we find the Fermi energy of the free neutrons.

Writing $F(\text{neutrons})$, $F(\text{electrons})$ for the Fermi energies of the neutrons and electrons, we have

$$F(\text{neutrons}) = \frac{1}{A} \left\{ M(A, Z) c^2 + Z \left(F(\text{electrons}) - m_e c^2 \right) + \frac{4}{3} Z \frac{U_L}{n_e} \right\} \,. \tag{8.16}$$

We calculate $F(\text{neutrons})$ for each A and Z and a given baryon density. Conversely, to find the neutron-evaporation threshold for a given chemical species we set $F(\text{neutrons}) = 0$ and calculate $F(\text{electrons})$, giving the density. Above this density free neutrons are captured and a new chemical species appears, until it too is superseded by a species richer in neutrons.

Above a certain density neutron-rich nuclei can only exist in the presence of a free-neutron gas, the neutrons evaporating spontaneously from the nuclei. Reactions of the type $^A_Z X \rightarrow ^{A-1}_{Z-1} X + n$ are forbidden if the free-neutron energy states are occupied. The medium then consists of nuclei surrounded by free electrons and free neutrons.

The calculation of successive neutronisation states depends on the function $M(A, Z)$, extrapolated from known nuclei to neutron-rich nuclei. The results obtained differ slightly depending on the choice of the function

Table 8.2. Equilibrium nuclei before neutron evaporation

Nuclei	$\rho_{\max}$ (g cm^{-3})	Nuclei	$\rho_{\max}$ (g cm^{-3})
$^{56}_{26}$Fe	$8.1.10^6$	$^{78}_{28}$Ni	$1.6.10^{11}$
$^{62}_{28}$Ni	$2.7.10^8$	$^{76}_{26}$Fe	$1.8.10^{11}$
$^{64}_{28}$Ni	$1.2.10^9$	$^{124}_{42}$Mo	$1.9.10^{11}$
$^{84}_{34}$Se	$8.2.10^9$	$^{122}_{40}$Zr	$2.7.10^{11}$
$^{82}_{32}$Ge	$2.2.10^{10}$	$^{120}_{38}$Sr	$3.7.10^{11}$
$^{80}_{30}$Zn	$4.8.10^{10}$	$^{118}_{36}$Kr	$4.3.10^{11}$

$M(A, Z)$. Table 8.2 gives the values of the critical density ρ_{max} for a sequence of neutron-rich nuclei (Baym, Pethick, and Sutherland 1971), above which a new nuclear species is in equilibrium with the medium.

The ratio Z/A decreases along the sequence.

Between the density ($\rho = 4.3 \times 10^{11}$ g cm^{-3}) at which neutron evaporation allows an equilibrium nuclei $\rightleftharpoons$ neutrons + electrons, and nuclear density ($\rho \simeq 2.8 \times 10^{14}$ g cm^{-3}), the equation of state is fairly well understood. Above the latter, nucleon–nucleon interactions become very important and there is currently no agreement about what theory to adopt.

8.2.2 The Equation of State for $T \neq 0$

Supernovae are associated with the collapse of a star or stellar core of the order of one solar mass. The temperature increase associated with the collapse brings about an equilibrium nucleus $\rightleftharpoons$ nucleus$'$ $+n + e^-$, which determines the equation of state. The relaxation time for nuclear equilibrium is so short (10^{-20} s) compared with the collapse timescale (10^{-3} s) that we can regard the medium as being in local thermodynamic equilibrium and in a minimum-energy state at each instant. We note that the assumption of local thermodynamic equilibrium implies a different treatment according to whether we assume the medium opaque or transparent to neutrinos. In fact the medium becomes opaque to neutrinos at a density of a few times 10^{11} g cm^{-3} (below nuclear densities) and for a stellar radius of 100 km (well before the end of the collapse). We shall return to this question later (p. 367).

At finite temperature, and below nuclear densities, we have to consider the statistical equilibrium between various species of nuclei in the presence of a fluid of protons, neutrons, electrons, and alpha particles. El Eid and Hillebrandt (1980) considered an ensemble of 457 nuclear species, from $Z = 10$ to $Z = 32$. The energy density of the mixture includes the contribution of all these species and that arising from nucleon–nucleon interactions. If n is the baryon number density, the pressure is found from the total energy density using the thermodynamic relation

$$P = -\frac{\partial(U/n)}{\partial(1/n)}. \tag{8.17}$$

Above the density ($\rho_0/10$) (where ρ_0 is the nuclear density) we have to take account of the interaction between the nucleus and the gas of free particles (protons and neutrons). The calculation requires a model of the nucleon–nucleon interaction and the Hartree–Fock potential acting on each particle.

There have been two calculations: (a) using a liquid-drop model of the nucleus (Lamb, Lattimer, Pethick, and Ravenhall 1978); (b) using a periodic lattice, each cubic cell containing Z protons and $(A - Z)$ neutrons (Bonche and Vautherin 1981), fixing the entropy per nucleon at $(S/Ak) \simeq 1$.

Table 8.3. Atomic weight at very high density for $S/Ak = 1$ (from Bonche and Vautherin)

Temperature (MeV)	P (MeV fm^{-3})	Density (fm^{-3})	Density (g cm^{-3})	A
3.80	0.1357	0.02	$3.3 \cdot 10^{13}$	300
4.35	0.1309	0.03	$5.0 \cdot 10^{13}$	400
4.75	0.3356	0.04	$6.7 \cdot 10^{13}$	500
5.00	0.4482	0.05	$8.3 \cdot 10^{13}$	600
5.34	0.5655	0.06	$1.0 \cdot 10^{14}$	900
5.53	0.6851	0.07	$1.16 \cdot 10^{14}$	1 000

Very neutron-rich nuclei are produced (Table 8.3) where the density is given in fermi^{-3} or fm^{-3}, 1 fm = 10^{-13} cm). Once the density reaches one half of the nuclear value, neutron-rich nuclei are replaced by bubbles full of free neutrons and protons, while the bound neutrons and protons collect on the walls and edges of the cubic lattice.

For a certain value of the density (depending on the temperature) there is a phase transition to a single homogeneous phase, well represented by a Saha equation with a binding energy of 16 MeV per nucleon for the dense phase (in the lower-density domain). In the higher-density domain, the transition occurs when the nuclei touch (nuclear density). The boundary is independent of the temperature since the nucleons form a degenerate gas whose Fermi energy is about 60 MeV per nucleon, well above the temperature of the medium (Fig. 8.1).

It is very difficult to establish the equation of state above nuclear densities. An essential property is that the sound speed should be less than the speed of light. The relation $dP/d\rho < c^2$ implies the upper limit $P = \rho c^2$,
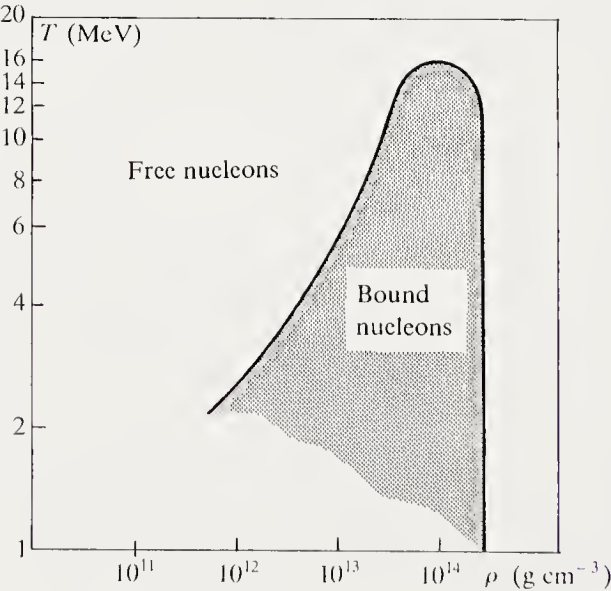
Fig. 8.1. Phase diagram for dense matter. The region of bound nuclei is limited on one side by high temperatures and on the other by high densities (above nuclear density)

which we shall use later. Once neutrinos are trapped in dense matter, the equation of state should include the pressure of their Fermi gas.

8.2.3 Maximum Mass of White Dwarfs

Relativistic Effects. A very simple argument based on a uniform-density model allows us to estimate the general-relativistic correction.

For a sphere of radius R containing N nucleons of atomic weight A and charge Z the baryon mass is

$$M_B = N \, A m_U \,, \tag{8.18}$$

where m_U is the atomic mass unit 1.66×10^{-24} g, and the electron number density is given by

$$\tfrac{4}{3}\pi R^3 N_e = N(Z/A) \,. \tag{8.19}$$

The total energy is then

$$W = -\frac{3}{5}\frac{GM^2}{R} + M_B c^2 + \frac{4}{3}\pi R^3 U_e \,. \tag{8.20}$$

We must include the electron energy in the gravitational mass:

$$M = M_B + \frac{4}{3}\frac{\pi R^3 U_e}{c^2} \,, \tag{8.21}$$

and we then have

$$W = -\frac{3}{5}\frac{G}{R}\left(M_B + \frac{4}{3}\frac{\pi R^3 U_e}{c^2}\right)^2 + M_B c^2 + \frac{4}{3}\pi R^3 U_e \,. \tag{8.22}$$

We have to find the minimum of W. Differentiating with respect to R, using the relation between R and x (8.3, 19), we obtain a relation $M_B = M_B(x)$. The maximum mass occurs for $\partial M_B/\partial x = 0$. To a first approximation, the mass is given by Chandrasekhar's limiting value

$$M_{B\,Ch} = \left(\frac{\hbar c}{G}\right)^{3/2}\left(\frac{Z}{A m_U}\right)^2 \,, \tag{8.23}$$

where m_U is the mass unit, so that $M_B \simeq 1.5\,M_\odot$. The value of $x = p_F/m_e c$ is

$$x = \left(\frac{4}{3}\frac{A}{Z}\frac{m_U}{m_e}\right)^{1/3} \,, \tag{8.24}$$

where $x \simeq 17$; the corresponding density is $\rho = 9.48 \times 10^9$ g cm^{-3}; the radius is

$$R \simeq \left(\frac{\hbar^3}{Gc}\right)^{1/2} \frac{1}{(m_U^2 m_e)^{2/3}}, \tag{8.25}$$

or about 440 km (about one half of the exact result). The relativistic correction gives

$$M_{BGR} = M_{BCh} \left(1 - 1.5\frac{GM_B}{Rc^2}\right), \tag{8.26}$$

amounting to a correction of 1.5 %.

The exact calculation uses the hydrostatic-equilibrium equation in general relativity; in spherical coordinates these are the Oppenheimer–Volkoff equations:

$$\frac{dm}{dr} = 4\pi r^2 \rho, \tag{8.27}$$

$$\frac{dP}{dr} = -\frac{\rho m}{r^2}\left(1 + \frac{P}{\rho c^2}\right)\left(1 + \frac{4\pi P r^3}{mc^2}\right)\left(1 - \frac{2GM}{rc^2}\right)^{-1}, \tag{8.28}$$

and the condition on the metric is easily satisfied at the surface. Approximating the solution as an $n = 3$ polytrope we can calculate the internal energy, the gravitational energy, and the general-relativistic correction. Writing

$$P = K\rho^{4/3}, \tag{8.29}$$

for K constant we find (see Appendix)

$$E_{\text{int}} = AM\rho_c^{1/3} + CM\rho_c^{-1/3}, \tag{8.30}$$

$$E_{\text{grav}} = -BM^{5/3}\rho_c^{1/3}, \tag{8.31}$$

$$\Delta E_{RG} = -DM^{7/3}\rho_c^{2/3}. \tag{8.32}$$

with

$$A = k_1 K, \qquad B = k_2 G,$$
$$C = k_3 \frac{m_e^2 c^3}{\hbar(\mu_e m_U)^{2/3}}, \qquad D = k_4 \frac{G^2}{c^2}, \tag{8.33}$$

and

$$k_1 = 1.75579, \qquad k_2 = 0.63900,$$
$$k_3 = 0.519723, \qquad k_4 = 0.918294,$$

where these values come from assuming the density distribution is given to a first approximation by an $n = 3$ polytrope. We thus find the mass, radius, and critical density, including the relativistic correction:

344
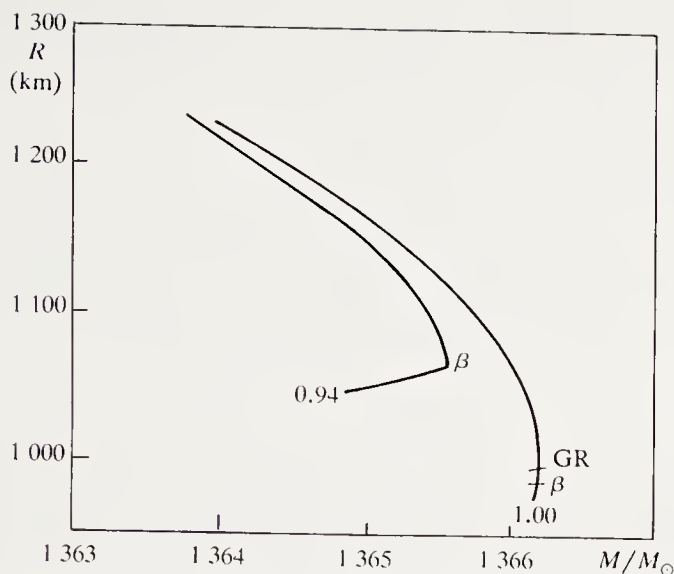
Fig. 8.2. Mass–radius relation for carbon–oxygen white dwarfs of homogeneous chemical composition. The inverse beta-decay instability is reached before the relativistic instability for carbon concentrations less than 0.94. β denotes the point where the electron-capture instability appears, and GR the appearance of the general-relativistic instability

$$M_{\mathrm{crit}} = 1.457 \, (\mu_e/2)^{-2} \left(1 - 15.7 \frac{GM}{Rc^2} \right) M_\odot \,, \tag{8.34}$$

$$R_{\mathrm{crit}} \simeq 10^8 \ \mathrm{cm} \,, \tag{8.35}$$

$$\rho_c = \frac{CB^2}{DA^2} = 2.646 \times 10^{10} \left(\frac{\mu_e}{2} \right)^2 \ \mathrm{g \, cm^{-3}} \,, \tag{8.36}$$

where $\mu_e = A/Z$. The relativistic correction to the mass is 0.034.

Nuclear Effects (Beta Captures). The capture of electrons by nuclei reduces the central electron pressure. This capture occurs in oxygen for a density below the critical relativistic density ($\rho_c(\mathrm{oxygen}) = 1.90 \times 10^{10}$ g cm^{-3}), while in carbon and helium it occurs above above the critical density ($\rho_c(\mathrm{carbon}) = 3.90 \times 10^{10}$ g cm^{-3}; $\rho_c(\mathrm{helium}) = 1.37 \times 10^{11}$ g cm^{-3}) (Table 8.1). Figure 8.2 give the mass–radius relations for white dwarfs of different chemical composition. For a carbon white dwarf one has to reach a carbon concentration $X(^{12}\mathrm{C}) = 0.94$, so that the unstable branch begins at the relativistic critical mass. For a pure-oxygen white dwarf the unstable branch begins before the critical relativistic mass is reached. In other words, an oxygen-rich white dwarf whose mass grows by accretion can become unstable before reaching the relativistic mass.

Pycnonuclear Reactions. The possibility of pycnonuclear reactions (see Chap. 3) imposes constraints on the chemical composition of white dwarfs. The reaction time for a pure carbon white dwarf is, from Van Horn and Salpeter (1969)

$$\tau = 2.5 \times 10^{-46} \lambda^{-7/4} \exp(2.64\lambda^{-1/2}), \qquad \text{with} \tag{8.37}$$

$$\lambda = \frac{1}{A} \left(\frac{1}{Z^7} \frac{1}{\mu_e} \frac{\rho}{1.36 \times 10^{11}\,\text{g cm}^{-3}} \right)^{1/3}. \tag{8.38}$$

This time reaches 10^{17} s (3 billion years) for $\rho \sim 10^{10}$ g cm^{-3}. This central density is reached for white dwarfs of about $1.3\,M_\odot$. We conclude that carbon-rich white dwarfs cannot have masses above about $1.3\,M_\odot$.

8.2.4 The Maximum Mass of Neutron Stars

Use of an Approximate Equation of State. The maximum-mass problem for neutron stars is more difficult, because we do not know the exact equation of state of ultradense matter above nuclear densities.

The essential constraint that the sound speed should be less than that of light allows an estimate of the maximum mass of a neutron star. The assumptions are

— The equilibrium is represented by the Oppenheimer–Volkoff equations.

— The pressure is an increasing function of the density. This is a constraint of microscopic stability, since there can be no small condensations of matter.

— The causality condition is written

$$\frac{dP}{d\rho} \leq c^2.$$

— We refer the equation of state to a density ρ_0 for which the pressure is known.

A plausible assumption is

$$P = P_0 + (\rho - \rho_0)c^2 \qquad \rho \geq \rho_0 \tag{8.39}$$

with $\rho_0 = 4.6 \times 10^{14}$ g cm^{-3} and $P_0 = 7 \times 10^{33}$ dyne cm^{-2} (Rhoades and Ruffini 1974). With this pressure–density relation we can estimate the maximum mass. The gravitating mass includes the baryons and their kinetic energy. If $\varepsilon = \rho c^2$ is the energy density, n the baryon number density and P the pressure, the adiabatic assumption leads to

$$d\left(\frac{\rho c^2}{n} \right) + P d\left(\frac{1}{n} \right) = 0, \tag{8.40}$$

giving

$$\frac{d\rho}{dn} = \frac{\rho + P/c^2}{n}. \tag{8.41}$$

We look for an approximate solution of the form $\rho = $ constant. Then the mass is

$$M = 4\pi \int_0^R \rho r^2 \, \mathrm{d}r = \frac{4}{3}\pi R^3 \rho \tag{8.42}$$

and the total baryon number (here with $G = c = 1$) is

$$
\begin{aligned}
A &= 4\pi \int_0^R \frac{n r^2 \, \mathrm{d}r}{(1 - 2m(r)/r)^{1/2}} \\
&= 2\pi n \left(\frac{3}{8\pi\rho} \right)^{3/2} (\chi - \sin\chi \cos\chi),
\end{aligned}
\tag{8.43}
$$

where $\sin\chi$, defined by (8.44), is the ratio of the Schwarzschild radius to the radius of the star:

$$\sin\chi = \left(\frac{8\pi\rho}{3} \right)^{1/2} R = \left(\frac{2M}{R} \right)^{1/2} \tag{8.44}$$

(with $G = c = 1$).

The equilibrium configuration is reached when the energy is a minimum at constant baryon number, i.e.

$$\left(\frac{\partial M}{\partial \chi} \right)_A = 0. \tag{8.45}$$

We differentiate (8.42), (8.43), and (8.44) with respect to χ. Eliminating $\mathrm{d}n/\mathrm{d}\chi$, and $\mathrm{d}\rho/\mathrm{d}\chi$ using (8.41) gives

$$\left(\frac{P}{\rho} \right) = \zeta(\chi) = \frac{6\cos\chi}{9\cos\chi - 2\sin^3\chi/(\chi - \sin\chi\cos\chi)} - 1. \tag{8.46}$$

For a given relation $\rho(n)$ we can then calculate $R(\chi)$ and $M(\chi)$. The latter is a function of χ with a maximum at $\chi = \chi_m$. For $\chi < \chi_m$ the stability condition is

$$\left(\frac{\partial^2 M}{\partial \chi^2} \right)_A \geq 0 \tag{8.47}$$

at constant baryon number. The maximum mass at $\chi = \chi_m$ is given by the vanishing of the second derivative in (8.47).

We can write (8.47) in the form

$$\Gamma > \Gamma_{\text{crit}}(\chi), \tag{8.48}$$

where

$$\Gamma = \frac{\partial \ln P}{\partial \ln n} = \left(1 + \frac{\rho}{P} \right) \frac{\mathrm{d}P}{\mathrm{d}S} \tag{8.49}$$

and

$$\Gamma_{\text{crit}} = (\zeta + 1)\left\{1 + \frac{3\zeta + 1}{2}\left(\frac{\zeta + 1}{6\zeta}\tan^2\chi - 1\right)\right\}.$$ (8.50)

The limiting mass is given by $dP/d\rho = 1$, and

$$\Gamma = \frac{n}{P}\frac{dP}{dn} = \left(1 + \frac{1}{P/\rho}\right)\frac{dP}{d\rho},$$ (8.51)

so that for this mass

$$\Gamma = 1 + \frac{1}{(P/\rho)} = \Gamma_{\text{crit}},$$ (8.52)

implying

$$1 = \zeta\left\{1 + \frac{3\zeta + 1}{2}\left(\frac{\zeta + 1}{6\zeta}\tan^2\chi - 1\right)\right\}.$$ (8.53)

This condition is satisfied for $\zeta = 0.364, \chi = 1.12, M/R = 0.405$ and

$$M_{\text{max}} \simeq 3.6M_\odot.$$ (8.54)

With more realistic equations of state we find maximum masses between 1.5 and $2.7\,M_\odot$. As an example, the equations of state of Bethe and Johnson and of Friedman and Pandharipande give the respective results

$$M_{\text{max}} = 1.85M_\odot, \qquad R = 9.8\,\text{km}, \qquad GM/Rc^2 = 0.28,$$
$$M_{\text{max}} = 1.93M_\odot, \qquad R = 9.4\,\text{km}, \qquad GM/Rc^2 = 0.305.$$

Any compact object with a mass above about $3\,M_\odot$ is a black-hole candidate.

Neutron-Star Masses: Observations. There are two sources of information on neutron-star masses, corresponding to two types of binary: (a) X-ray binaries, and (b) binary pulsars.

For X-ray binaries we have the standard case of a spectroscopic binary. If we can measure the period P and the radial velocity amplitudes of both components, v_X and v_0, we have the mass functions

$$f_X = \frac{(M_0 \sin i)^3}{(M_X + M_0)^2} = \frac{Pv_X^3}{2\pi G},$$
$$f_0 = \frac{(M_X \sin i)^3}{(M_X + M_0)^2} = \frac{Pv_0^3}{2\pi G},$$ (8.55)

giving the mass ratio

$$q = \frac{M_X}{M_0} = \frac{v_0}{v_X}$$ (8.56)

and the mass of the X-ray component

$$M_X = \frac{f_X q (1+q)^2}{\sin^3 i} . \tag{8.57}$$

The mass determination depends on finding $\sin i$. This can be estimated from the eclipse duration or the light-curve form. In the latter case the estimate is crude.

For binary pulsars the data come from the measurement of the pulse arrival times. In the case of PSR 1913+16 for example, we have only one mass function, since the companion is a dense object which is not directly observable. However, the very small separation of the components gives rise to measurable relativistic effects such as periastron advance, giving more information and allowing one to calculate the masses of both components $M_{pulsar} = 1.41 \pm 0.06 \, M_\odot$, $M_{companion} = 1.41 \pm 0.06 \, M_\odot$. The period of PSR 1913+16 is 59 ms, the orbital period is 27.906 s, and the periastron advance is 4.22 degrees per year.

Figure 8.3 summarises the present data on neutron-star masses. The masses are all less than the maximum mass; they are compatible with the masses found in models of neutron-star formation in supernova explosions (see Sect. 8.5). They are also close to the maximum mass found using more realistic equations of state.

These methods do not apply to low-mass X-ray binaries, which in general do not pulse; they do not apply to binaries containing a black-hole candidate, such as Cyg X-1, LMC X-3, and A0620-00. The first two are massive systems, and the mass limit for the companion is deduced from an
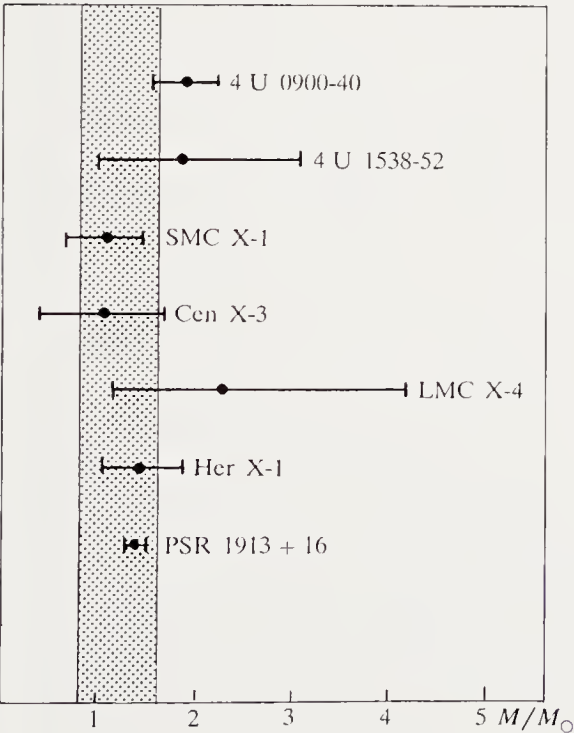
Fig. 8.3. Summary of observational data for neutron-star masses (X-ray binaries, binary pulsar). The abscissa gives the mass and the ordinate the various stars for which the mass has been measured. The *horizontal lines* are the error bars. The *vertical dotted* region shows the mass range deduced from various production scenarios for neutron stars

assumption about the mass–luminosity relation of the visible companion; the third is a low-mass system with period 7.7 hours. In this case the mass function of the invisible star is already $3.2\,M_\odot$, giving a lower limit to the mass of the dense star.

8.3 White Dwarfs

8.3.1 Properties

Spectral Types. The various spectral types of white dwarf indicate differences of internal structure. Two main classes are known, the DA, characterised by the presence of hydrogen in their spectra, and the DB, characterised by helium, and sometimes showing traces of metals or the $\lambda\,4670$ band of C_2. The DB sequence extends to low temperatures to a class of stars without spectral lines at all. From their continua they are classified as DC stars. At high temperatures the stars are classified as DO, by analogy with O stars.

The temperature scale is established using model atmospheres of pure hydrogen or helium. DA temperatures go from 7000 K to 30 000 K. Lines of neutral helium (He I) are seen in the DB stars from 12 000 K to 30 000 K.

The DA are helium-poor (He/H $< 10^{-3}$). In contrast, the hydrogen abundance is very low in DB stars (He/H $\gtrsim 10^5$). Hydrogen has been found in some DB stars, called DBA stars, with He/H $\simeq$ 3–10 $\times 10^3$.

The masses of the DA stars are of the order of 0.6 to $0.8\,M_\odot$, while the DB stars have masses of the order of $0.3\,M_\odot$.

The DA and DB white dwarfs appear to have different origins and to have been formed in these two types respectively. About 15 % of white dwarfs are of DB type or its variants such as DC. Their velocity distribution suggests that white dwarfs belong to an intermediate or old population. A small fraction come from population I stars, implying that less than 50 % of white dwarfs come from stars with mass larger than $1.5\,M_\odot$, i.e. with ages less than 3 billion years.

Gravitational Separation. The differences in chemical composition between the DA and DB stars appear to be associated with their different origins. The existence of hot (up to 40 000 K) helium stars shows that, as for the DA stars, we are dealing with a continuous cooling sequence of stars.

The absence of helium in the DA stars and the low abundance of metals and carbon in the DB stars are caused by gravitational settling. The effect of settling appears when we require that in equilibrium the diffusion velocity should vanish (see Sect. 5.5.4). In isothermal equilibrium the contribution of the term ∇T is zero. In statistical equilibrium the mixing height depends a little on the degree of degeneracy. Once settling of hydrogen and other ions has occurred, the difference in the ratios $A/Z\,[(A/Z)_H = 1; (A/Z)_{ion} \simeq 2]$

means that the mixing height is always of the order of the scaleheight of the ions and small compared with the the pressure scaleheight, and thus small compared with the star's radius.

When settling of helium and other ions occurs the situation is very different depending on whether the medium is degenerate or not. In a non-degenerate or weakly degenerate medium the mixing height is smaller than the pressure scaleheight but of the same order; in contrast, in a strongly degenerate medium the mixing height grows with the degeneracy parameter Λ, given by

$$\ln \Lambda = \left(\frac{3\sqrt{\pi}}{8} \frac{\rho}{\mu_e m_H} \right)^{2/3} (h^2/2\pi m_e kT)$$
$$= 3.094 \times 10^5 \left(\frac{\rho}{\mu_e} \right)^{2/3} \frac{1}{T},$$

once $A_1/Z_1 - A_2/Z_2$ is zero. This is caused by the decoupling of the electron and ion pressure gradients in strongly degenerate conditions. If the charge per unit mass is the same for all ions there is no settling at all.

DA white dwarfs, with hydrogen envelopes, allow very efficient settling between hydrogen and the other elements. By contrast, in DB white dwarfs the helium envelopes favours less efficient settling.

Structure. Immediately below the surface radiative zone, at very small optical depth ($\tau < 0.1$) there is a convection zone, which may extend down to layers of quite high temperature. For decreasing surface temperatures convection appears in DA stars at $T_{\rm eff} \simeq 16\,500$ K and in DB stars at $50\,000$ K.

As a star with a helium envelope cools, the convection zone becomes deeper. We might suppose that metals would appear at the surface once the convection zone became deep enough to dredge them up. However, the timescale for gravitational separation is so short compared with the cooling time that settling has stopped by the time the convection zone reaches deep layers. The metals therefore cannot be dredged up to the surface. Metals may be added by accretion. This idea meets some difficulties, and at present there is no fully satisfying theory of the presence of metals in DB spectra.

8.3.2 Evolution and Cooling

Cooling. We have already seen (Sect. 3.4.5) that the energy source for white dwarfs is their thermal energy. We assume (as is indeed very plausible) that white dwarfs are formed as a result of a planetary nebula in which a star of initial mass less than 6–8 $M_\odot$ is reduced to a degenerate core of temperature $\gtrsim 10^8$ K in which nuclear reactions have stopped. This core then evolves by radiating its thermal content, at first rapidly via neutrino emission, then more and more slowly by radiative transfer through a thin

non-degenerate envelope. After about 10^5 y the matter ejected by the central star of the planetary nebula has dispersed enough to be undetectable, and the object passes from the planetary-nebula nucleus stage to the white-dwarf stage. The star follows a line of constant radius in the HR diagram and its luminosity decreases over 10^{10} y to about $\log(L/L_\odot) \simeq -4.5$.

Luminosity Function. Study of white-dwarf cooling allows us to construct a theoretical luminosity function, i.e. the number of white dwarfs per unit volume (of the galactic disc) and per interval of bolometric magnitude. Because they outnumber all the others, we calculate the luminosity function for $0.6\,M_\odot$ white dwarfs.

While the white-dwarf plasma remains liquid we can use the relation found in Sect. 3.4.5,

$$L \sim T^{3.5} \sim V_v M_{WD} \frac{\mathrm{d}T}{\mathrm{d}t}, \qquad (8.58)$$

giving the luminosity function $\Phi \sim (|\mathrm{d}\log L/\mathrm{d}t|)^{-1} \sim L^{-5/7}$. In the $(\log L, \log \Phi)$ representation the luminosity function is a straight line of slope $-5/7$, which agrees roughly with observations of the brighter white dwarfs (Fig. 8.4).

To do better than this requires a more detailed model of the outer layers and estimates of the electrostatic energy's contribution to the specific heat, the latent heat of crystallisation, and the change of specific heat when the temperature of the central regions falls below the Debye temperature. In this last phase the specific heat is lower (it deceases as T^3), accelerating the cooling and flattening the luminosity function (Fig. 8.4, for $\log(L/L_\odot) \simeq -2.5$).
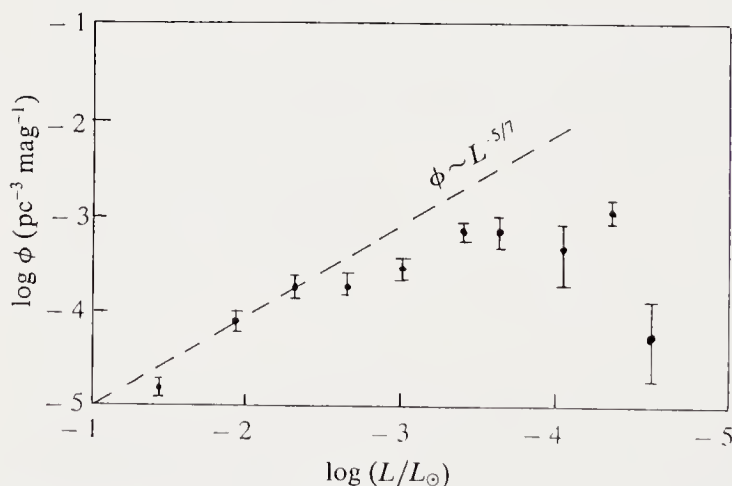
Fig. 8.4. White-dwarf luminosity function (from the data of Winget et al. *Ap. J.* **315**, L77 (1987)) and the theoretical relation $L^{-5/7}$ for cooling of a liquid plasma (log–log scale). For bright white dwarfs, neutrino cooling puts the observational points below the line; for faint white dwarfs the decreased heat content flattens the distribution below the line

The observed luminosity function uses the relation between visual and bolometric magnitudes. The latest results on white-dwarf bolometric corrections make $\log(L/L_\odot) = -4.5$ correspond to visual magnitude $M_V = 15.7$. Assuming that white-dwarf number counts are correct to this magnitude, there appears to be an abrupt decrease of the luminosity function, with a total absence of white dwarfs of luminosity less than $\log(L/L_\odot) = -4.5$. But for this luminosity, and a mass of $0.6\,M_\odot$, present theory gives an age of 9×10^9 y, much less than the ages of globular clusters (1.8×10^{10} y). This difference raises several questions.

(a) Is the physics of white-dwarf interiors understood? Is there another energy source which would increase the lifetime of white dwarfs with $\log(L/L_\odot) = -4.5$ by about 5 billion years?

(b) Are the white-dwarf number counts complete? All of those counted are nearby, so are we correctly accounting for their galactic distribution, particularly as the scaleheight perpendicular to the galactic plane increases with the age of the population?

(c) Is the ratio of white-dwarf formation to stellar births independent of time?

8.3.3 White Dwarfs in Binaries

The presence of white dwarfs in many binaries introduces the physics of the interaction of the white dwarf with a companion. The main phenomenon is accretion from this star on to the white dwarf. Depending on the chemical composition of the white dwarf and the accreted material as well as the accretion rate, a great variety of phenomena can occur: cataclysmic variables, novae, type I supernovae, and possibly the formation of at least some pulsars.

The designation cataclysmic variable is applied to a wide variety of stars which show abrupt increases in brightness at irregular intervals. This group includes the novae, dwarf novae, and recurrent novae. Typical novae are also probably recurrent, but at intervals of the order of 10^5 y; recurrent novae have outbursts separated by a few decades, and dwarf novae (U Gem or SS Cyg type) have weaker outbursts (a few magnitudes) at intervals of weeks. The physical processes involved are very different, typical nova explosions being thermonuclear in origin, while dwarf-nova outbursts are caused by the production of a shock wave due to an increase in the rate of accretion on to the white dwarf of matter originating on the companion.

We note here an important property of the orbital periods of cataclysmic variables, namely the lack of systems with periods between two and three hours, and the absence of systems with periods less than 80 minutes. Long-period systems are characterised by electromagnetic angular-momentum losses (possibly of the type described for single stars in Sect. 5.5). These losses shrink the binary orbit and thus cause mass to be transferred from the companion to the white dwarf. At a period of about 3 hours,

the companion (which fills its Roche lobe) has a period of about $0.3\,M_\odot$ and becomes fully convective. The currently favoured hypothesis is that the angular-momentum losses are then severely reduced, and with them the mass transfer and the variability.

We shall describe here only some results relevant to the study of accretion on to a white dwarf and the production of thermonuclear explosions in novae.

Accretion on to White Dwarfs.   We describe the physics of the process. The accreted matter contracts gravitationally, which raises the temperature. Energy is transported outwards by radiation and inwards by conduction. Nuclear burning of hydrogen (usually the main constituent of the accreted material) begins and becomes significant once the temperature at the base of the accreted layer exceeds 10^7 K. The situation then depends on the accretion rate.

(a) For the lowest accretion rates, contraction is slow and nuclear reactions are late in appearing. The accumulated mass is enough to make the material partially degenerate, so that hydrogen burning is explosive. This leads to the ejection of the accreted envelope: this is the nova phenomenon.

(b) For intermediate accretion rates nuclear flashes are more frequent but weaker.

(c) High accretion rates allow stable hydrogen burning at the rate the material accretes. The star becomes a red giant.

(d) There is a limiting accretion rate at which the gravitational energy release produces enough radiation pressure to balance the weight of the accreting material. If the accreting layer is optically thin this corresponds to the Eddington luminosity $L_{\mathrm{Edd}} = 4\pi c G M_{WD}/\kappa$, where M_{WD} is the mass of the white dwarf. If the opacity is due to Thomson scattering, $L_{\mathrm{Edd}} = 1.4 \times 10^{38} M_{WD}$ and the limiting accretion rate is $\dot{M} = 10^{-3}(R_{WD}/R_\odot)$ $M_\odot$ y^{-1}.

The critical accretion rate above which the star becomes a red giant can be found by comparing the growth rate of the degenerate core by accretion with the growth rate of the degenerate core of a red giant with hydrogen shell burning,

$$\left(\frac{\mathrm{d}M}{\mathrm{d}t}\right)_{RG} = \frac{L_H}{X_e \varepsilon_H},\tag{8.59}$$

where X_e is the hydrogen concentration of the envelope, and $\varepsilon_H = 6 \times 10^{18}$ erg g^{-1} is the energy yield per unit mass of hydrogen. With the interpolation formula

$$L = 5 \times 10^4 \left(M - \tfrac{1}{2}\right) L_\odot\tag{8.60}$$

we get, identifying L and L_H,
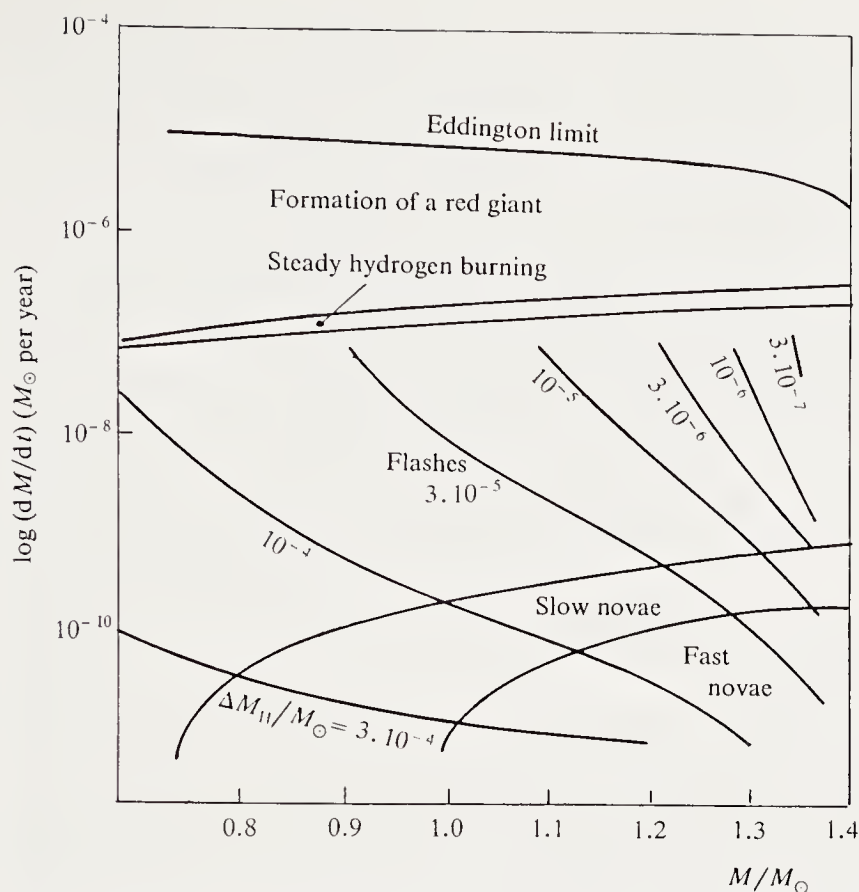
Fig. 8.5. Modes of hydrogen shell burning as a function of accretion rate $\dot{M}_H$ of hydrogen-rich material. In descending order we find the Eddington limit, the formation of a red giant envelope, the region of weak flashes, and the domain of slow and fast novae. ΔM_H is the mass of hydrogen-rich matter accreted at the point of ignition of hydrogen

$$\left(\frac{\mathrm{d}M}{\mathrm{d}t}\right)_{RG} = 6.75 \times 10^{-7}(M_{WD} - 0.5)M_\odot/\text{year}. \tag{8.61}$$

We call $\dot{M}_{RG}$ the critical accretion rate for the transition to the red-giant state. It seems that accretion is stable in the range 0.4–$1\dot{M}_{RG}$, flashes occurring at lower accretion rates. Figure 8.5 gives the various regimes in the $(M_{WD}, \dot{M})$ plane; for the flash regime the mass of hydrogen accreted at the point of ignition is given (implying the time interval between flashes).

Novae. Novae are characterised not only by energy production but also by the ejection of matter. The transition from the cataclysmic-variable state to a nova requires enough thermonuclear energy to produce the ejection at velocities of several hundreds to a thousand km s^{-1} (slow and fast novae respectively).

For carbon–oxygen white dwarfs the efficiency of thermonuclear reactions depends on the CNO concentration of the accreted matter. Rapid novae can only occur in CNO-rich accreted matter. The ejection condition

Table 8.4. Abundances in some novae

	Year	H	He	C	N	O	Ne
RR Pic	1925	0.53	0.43	0.0039	0.022	0.0058	0.011
HR Del	1967	0.45	0.48		0.027	0.047	0.0030
T Aur	1891	0.47	0.40		0.079	0.051	
V 1500 Cyg	1975	0.49	0.21	0.070	0.075	0.13	0.023
V 1668 Cyg	1978	0.45	0.23	0.047	0.14	0.13	0.068
V 693 Cr A	1981	0.29	0.32	0.0046	0.080	0.12	0.17
DQ Her	1934	0.34	0.095	0.045	0.23	0.29	
V 1370 Aql	1982	0.053	0.005	0.031	0.095	0.061	0.47

is satisfied once the pressure at the base of the accreted mass exceeds a critical value. For $Z(\mathrm{CNO}) = 0.507$, we require

$$P = \frac{GM_{WD}M_{\mathrm{flash}}}{4\pi R_{WD}^4} > 2 \times 10^{19} \text{ dyne cm}^{-2}, \tag{8.62}$$

where M_{flash} is the accreted mass. The regions of slow and fast novae are shown in Fig. 8.5.

A significant fraction of white dwarfs among cataclysmic variables are rich in neon, as is shown by the analysis of the chemical composition of the ejecta (Table 8.4). This neon cannot have been produced by the thermonuclear reactions involved in the explosion; it must have been present in the white dwarf, turbulent mixing in the explosion carrying some of the white dwarf material off with the ejecta.

8.3.4 Supernovae and White Dwarfs

Classification. Several hundred supernovae have been observed in other galaxies. They are classified as SN II or SN I depending on the presence or absence of hydrogen lines. The frequency of SN I and SN II depends on the type of galaxy, and even on the position within the galaxy. Tammann's statistics (Table 8.5) lead to the following conclusions:

Table 8.5. Supernova rate (per century and $10^{10} M_\odot$) (Tammann 1982)

Type of galaxy	All SN	SN 1	SN II
E	0.22	0.22	0
S0	0.12	0.12	0
S0a Sa	0.28	0.28	0
Sab Sb	0.69	0.37	0.32
SBc Sc Scd Sd	1.38	0.77	0.61
Sdm Sm Im	1.02	0.83	0.19
10	not determined		

(a) SN I occur in all galaxies.
(b) The SN I rate is higher for elliptical and late-type spiral galaxies.
(c) SN II occur only in Sab galaxies.
(d) The SN I and SN II rates are comparable in the most numerous group of spirals, i.e. Sbc, Sc, Scd, Sd.

SN II are thus clearly associated with the young population of galaxies. Conversely, the SN I come mostly from older stars. More exactly, SN II are located in spiral arms and correlated with neutral hydrogen. In constrast the SN I are associated neither with spiral arms nor hydrogen.

Both the SN I and SN II classes can be subdivided, implying a dispersion in the properties of the progenitor stars. 80 % of the SN I form a homogeneous group, the SN Ia, with remarkably similar light curves. The others differ in spectral characteristics and could have more massive progenitors, rather like the SN II.

The significance of Tammann's statistics may be altered by the recent suggestion of a correlation between SN II and bursts of star formation.

The Spectra of SN Ia.   The spectra of SN Ia are remarkably uniform. They evolve in time in such a predictable way that the age of an SN Ia can be found by comparing it with a well-observed SN Ia, such as SN 1972E in NGC 5253. The essential feature of SN I spectra is the almost total absence of hydrogen lines. Near maximum light the energy distribution in the spectrum is well represented by a fairly cool continuum (about 15 000 K), corresponding to a photospheric radius of 10^{15}; the expansion velocity is about 11 000 km s^{-1}.

The light curve is characterised by a very rapid rise to a luminosity of 10^{43} erg s^{-1}. During this phase most of the light comes from the continuous spectrum. As the expansion proceeds the photosphere cools rapidly and the luminosity falls to 10^{42} erg s^{-1} in a month. From then on the light curve shows a slow exponential decline with a timescale of about 60 days, which can last up to 700 days after maximum light. During this luminosity decrease the continuous spectrum continues to weaken and disappears, and the spectrum is dominated by broad emission bands. Model-atmosphere calculations well reproduce the observed bands, which seem to be due mainly to groups of iron lines (Fe^+ and Fe^{++}). Near maximum, by contrast these lines appear in absorption (Fig. 8.6).

This result agrees very well with the theory of energy production in SN I through the radioactive decay $^{56}Ni \rightarrow {}^{56}Co \rightarrow {}^{56}Fe$ with decay times 6 and 77 days respectively. The production of ^{56}Ni during the collapse of a white dwarf (see below) appears very plausible since this nucleus is tightly bound and stable in the presence of dense degenerate electron gas. The β radioactivity of ^{56}Co contributes to the luminosity maximum, but mainly heats the expanding gas and the electrons exciting the emission lines.

The standard model of the SN I starts with a carbon–oxygen white dwarf in a binary system with an accretion rate sufficiently great as to ensure that
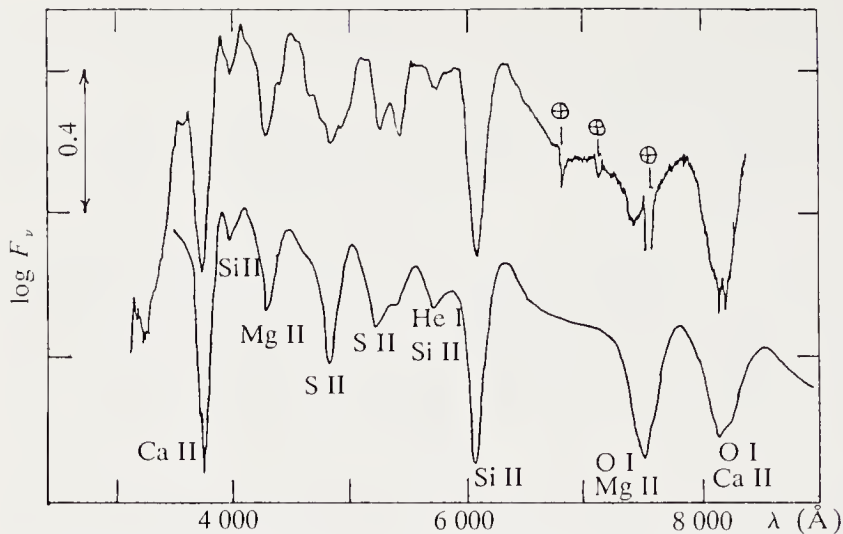
Fig. 8.6. *Top*: the spectrum of SN I 1981, in NGC 4536, 17 days after maximum light (McDonald Observatory). *Below*: a synthetic spectrum with $T_{\text{eff}} = 8000$ K, $v = 11\,000$ km s^{-1}, with visual extinction $A_V = 0.3$. The *crossed circles* indicate telluric lines. (From D. Branch, in *Supernovae: a Survey of Current Research*, ed. by M.J. Rees and R.J. Stoneham, Reidel 1982, p. 271. Reproduced by kind permission of Kluwer Academic Publishers)

its mass grows to the Chandrasekhar limit. The details of this phase have been calculated rather precisely for some cases. Given a $1\,M_\odot$ white dwarf with internal temperature 3×10^7 K, an accretion rate of $4 \times 10^{-8}\,M_\odot$ y^{-1} causes a series of flashes accompanied by the growth of a helium layer at the rate $\dot{M}_{He}$. The white dwarf contracts and the nuclear reaction rate, initially negligible, grows progressively.

The energy ε_{CC} liberated by carbon burning is at first removed by neutrino losses ε_ν, until the ignition line defined by $\varepsilon_{CC} = \varepsilon_\nu$ is crossed. Thermal instability appears before gravitational instability, the mass of the white dwarf being $M_{WD} = 1.378\,M_\odot$, with central density and temperature 2.6×10^9 g cm^{-3} and 3×10^8 K. The overpressure resulting from nuclear burning is only a few percent because of the strong degeneracy of the medium. A detonation wave cannot occur, but a deflagration wave appears, with a propagation velocity of the order of $(K/C\tau)^{1/2}$, where K/C is the thermal diffusivity and τ the characteristic time of the nuclear reactions.

It is important to know what type of remnant is left after the SN I explosion. According to some models, all of the star disperses, ejecting about $0.6\,M_\odot$ of ^{56}Ni. In other models the deflagration wave is slow enough to allow time for beta captures to occur followed by the collapse of the core and the production of a neutron star of between 0.4 and $0.6\,M_\odot$.

The mass of nickel ejected is directly related to the luminosity radiated by a type I supernova. Branch (1982) proposed the relation

$$H_0 = 40(M_{Ni}/M_\odot)^{-1/2} \text{ km s}^{-1} \text{ Mpc}^{-1}$$

between the Hubble constant (defined from the distance modulus of SN I) and the nickel mass, suggesting a nickel mass of the order of $0.25\,M_\odot$.

In our Galaxy the iron abundance is a test of the amount of ^{56}Ni ejected by SN I. The amount of iron produced in massive stars gives an upper limit to the amount of iron ejected by SN I (one event per century) of about $0.2\,M_\odot$ per explosion.

A Critique of SN I Models. We have still to check whether the scenarios suggested above are in fact realised. This depends on the population of white dwarfs in the various regions of the $(M_{WD}, \dot{M}_{He})$ plane, where $\dot{M}_{He}$ is the average net rate of mass increase resulting from accretion and mass loss in recurrent nova explosions. The distribution function of masses in binaries containing white dwarfs has a maximum at $M_{WD} \simeq 1\,M_\odot$, higher than for isolated white dwarfs (two maxima, at $0.6 - 0.7\,M_\odot$ and $0.2 - 0.3\,M_\odot$). The mass distribution is poorly known at higher masses. The rate of explosive collapses, which occur only for $M_{WD} \gtrsim 1.2\,M_\odot$, may be much smaller than the SN I frequency.

We examine systematically the accretion regimes for white dwarfs in binaries.

(a) For systems consisting of a white dwarf and a red giant or an AGB star filling its Roche lobe, mass transfer occurs on a dynamical timescale and leads to the formation of a common envelope which is ultimately lost by the system, except for the case $M_{WD} > 1.3\,M_\odot$. Accretion on to a white dwarf is possible only in the latter case, considerably reducing the frequency of this scenario, to the order of $1/100$ of the observed SN I frequency in the Galaxy.

(b) In systems consisting of a white dwarf and an AGB star, but where accretion occurs via a stellar wind, it is possible to produce SN I, but with the requirement of efficient capture ($\dot{M}_{He} > \dot{M}/3$). This condition has not been verified.

(c) In cataclysmic variables, consisting of a white dwarf and a main-sequence star, it is unclear if the mass increases at all for $\dot{M} < 10^{-8}\,M_\odot\,\mathrm{y}^{-1}$. Systems with higher accretion rates may not be very common.

A possible variant is the following scenario. In binaries consisting of two white dwarfs, gravitational radiation decreases the separation of the stars on a timescale of 10^{10} y for an initial separation of $3R_\odot$. Ultimately the less massive of the two white dwarfs fills its Roche lobe, leading to coalescence of the two white dwarfs on a dynamical timescale. If the total mass is large enough this may produce an SN I, but not necessarily. However, this depends in a complicated fashion on the chemical composition of the two components and the mass and angular-momentum exchange mechanisms.

If we assume that high-mass white dwarfs are produced by stars of 8–$10\,M_\odot$, we can estimate their number in the Galaxy. Assuming that all these white dwarfs become supernovae gives a rate of SN I comparable with

the observed one. This estimate suggests that there is a selection effect favouring the discovery of massive white dwarfs in binaries.

8.4 Neutron Stars

We have already mentioned the problem of neutron-star structure. We discuss their formation and properties here, and compare the latter with those of pulsars, X-ray sources, and gamma-ray-burst sources.

8.4.1 Structure

To understand the properties of pulsars, which we discuss later, we have to invert the historical order of discovery and first discuss models of neutron stars. Even without entering into quantitative details it is clear that the material of a neutron star is not the same at the surface as at the centre.

(a) *Close to the surface* ($\rho < 10^6$ g cm^{-3}) is a region about 10 m thick where the temperature and strong magnetic fields greatly affect the equation of state.

(b) We then have the *outer crust* ($10^6 \leq \rho \leq 4.3 \times 10^{11}$ g cm^{-3}), about 0.1 km thick, where the matter forms a Coulomb solid of neutron-rich nuclei in beta equilibrium with a relativistically degenerate electron gas.

(c) The *interior crust*, of thickness about 1 km ($\rho < 2 \times 10^{14}$ g cm^{-3}) consists of a Coulomb solid of neutron-rich nuclei in equilibrium with a neutron gas and a degenerate electron gas. The neutron gas is a superfluid (see later).

(d) The *neutron fluid* extends to the centre (about 10 km) and consists mainly of a neutron superfluid with a small concentration of superfluid protons and degenerate electrons.

The superfluidity of the neutron liquid is caused by the formation of neutron pairs with oppositely directed spin and momentum. These neutron pairs or quasi-particles form a boson gas which has superfluid properties if the temperature is low enough. This occurs if the thermal energy is less than the latent heat associated with pair formation. Nuclear research shows that neutrons and protons form pairs in heavy nuclei, and the latent heat is about 1 MeV. At the temperatures of a few keV inside neutron stars pair formation must occur and the neutron liquid behaves as a superfluid.

Superfluidity, as observed in ^{4}He below 2.19 K, is characterised by almost zero viscosity. In neutron stars this implies a very weak coupling between the neutron liquid and the solid crust. This is important for explaining the behaviour of pulsars during abrupt changes of the spin period.

8.4.2 Pulsars

These were discovered in 1967, during radio observations designed to detect interstellar scintillation. They are astronomical objects with radio emission pulsed at periods of the order of a second (Hewish et al. 1968).

The remarkable stability of the period required an emitter with high inertia. The very small period of 33 ms for the Crab pulsar required a small size. The later discovery of a 1.6 ms pulsar reinforced this, leading rapidly to the idea that the radio emission came from rotating neutron stars.

There are three main problems: (1) the emission mechanism of the radio waves, (2) the braking mechanism, and (3) the cause of period changes.

Radio Emission. About 500 pulsars are known (1988) in our Galaxy. They all show pulsed radio emission over a wide spectral band. The pulse amplitude varies widely and sometimes pulses are missing. The pulse duration is small compared with the period (1 to 5 %). At scales below a millisecond the pulses are very complex and show structure at a scale of 10 μs. However, the average profile over several hundred pulses is remarkably stable. Measurement of arrival times of average pulses shows that the pulsar, a rotating neutron star, is an extremely accurate clock. The period of some pulsars is known to a precision $\Delta P/P$ of the order of 10^{-13}.

The radio intensity has a power-law spectrum, $I_\nu \sim \nu^\alpha$, with $\alpha \simeq -1.5$ for $\nu < 1$ GHz, or even steeper for higher frequencies. A typical intensity is 0.1 Jy at 400 MHz (1 jansky = 10^{-23} erg s^{-1} cm^{-2}Hz^{-1}). The radiated power is between 2×10^{25} and 8×10^{30} erg s^{-1}.

Many pulsars show strong linear polarisation, up to 100 % in some cases. The degree and position angle vary with time during the pulse.

The radio pulses are interpreted using an oblique rotator model, in which the Earth is periodically swept by a directed radio beam which rotates with the pulsar. This requires a certain geometry of the emitting region, and the polarisation is explained if the radiation is from high-energy electrons in a magnetic field. The intensity is too great to be thermal, and coherent emission is the only plausible hypothesis.

There is no consensus at present about the mechanism responsible for pulsar emission, and the interpretation of polarisation properties is difficult. A possible model is as follows.

We assume a dipole magnetic field, with the dipole axis inclined to the rotation axis. Radio emission occurs partly along the dipole axis, from electrons accelerated along it. This core emission dominates around 400 MHz. The rest of the emission is in a hollow cone from electrons accelerated perpendicular to the fieldlines, the polarisation depending on the curvature of the fieldlines (Fig. 8.7). The change of position angle during the passage of the radio beam is explained if the emission in the hollow cone dominates around 1 GHz. Figure 8.8 shows the average pulse profile of PSR 1821 + 05 at various frequencies, clearly showing emission in a hollow cone at high frequencies of strength comparable to that of the magnetic core.
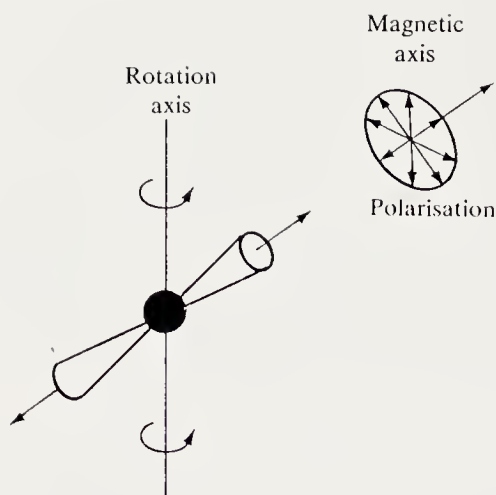
Fig. 8.7. Pulsar radio emission: core and hollow cone. Electrons propagate along the magnetic field, and the curvature of the fieldlines is responsible for the core emission (one pulse each time the beam crosses the Earth). Electrons gyrating around the fieldlines emit in a hollow cone (two pulses each time the beam crosses the Earth)

PSR 1821 + 05   DM = 66.8   pc – cm^{-3}

2 380 MHz

1 665 MHz

1 387 MHz

932 MHz

600 MHz

430 MHz

318 MHz

293 MHz

273 MHz

229 MHz

228 MHz

Flux

Pulse phase (milliperiods)
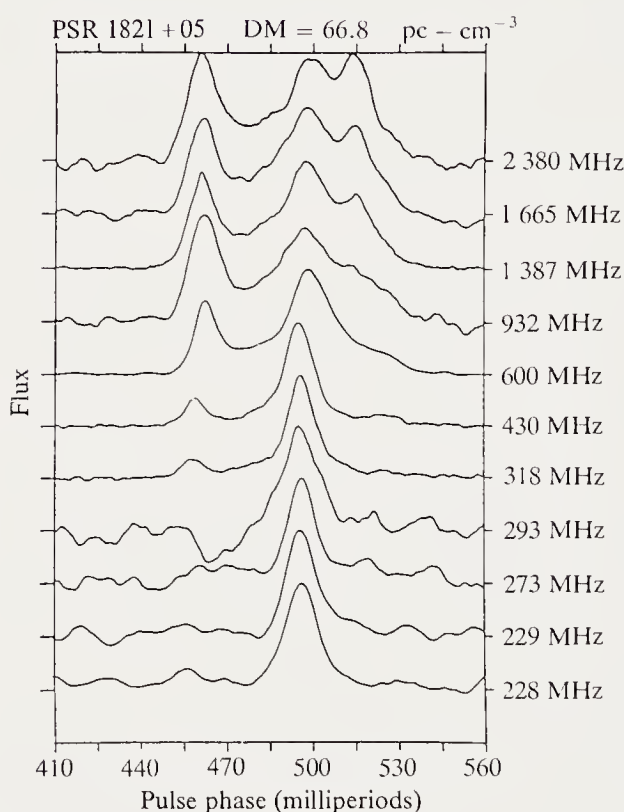
Fig. 8.8. Average pulse profiles in PSR 1821 + 05 for various frequencies between 228 and 238 MHz (see text) (J.H. Taylor and D.R. Stinebring, *Ann. Rev. Astron. Astrophys.* **24**, 308 (1986). Reproduced by kind permission of Annual Review of Astronomy and Astrophysics: © 1986 Annual Reviews Inc.)

Magnetic Fields. The presence of strong magnetic fields in neutron stars is shown by direct and indirect data:

— Pulsed emission in a hard X-ray line in Her X-1 and 4U 0115-63 can be interpreted as due to a cyclotron emission line correponding to fields of 4–6×10^{12} G in Her X-1 and 2×10^{12} G in 4U 0115-63.

— Contraction of a solar-type star with a surface magnetic field of 100 G down to the size of a neutron star gives a field of 10^{12} G if the field is frozen in. This order of magnitude is compatible with estimates of the surface

fields. However, the evolution to the neutron-star state is complicated and this agreement is not conclusive.

— The slowing of the rotation of pulsars is mainly caused by the emission of low-frequency waves (at the rotation frequency itself). The presence of a strong electrostatic field near the pulsar surface creates a charge-separated plasma around the neutron star. Study of this medium, corotating with the star, involves very difficult non-linear problems. The radiated power is however of the same order as for an oblique rotator in vacuum. Assuming that we have a dipole inclined to the rotation axis at an angle α and rotating in vacuum, we find a radiated power

$$W_{\text{dipole}} = -\frac{B_P^2 R^6 \Omega^4 \sin^2 \alpha}{6c^3}, \tag{8.63}$$

where Ω is the angular frequency and B_P the polar fieldstrength of the dipole. From the observed spindown of the Crab pulsar we get

$$W_{\text{cin}} = I\Omega\dot{\Omega} = 6.4 \times 10^{38} \text{ erg s}^{-1}, \tag{8.64}$$

where I is the moment of inertia. With $M = 1.4\,M_\odot$, $R = 12$ km, $I = 1.4 \times 10^{45}$ g cm^2. This loss of kinetic energy is comparable to the power radiated by the Crab nebula, $W_{\text{obs}} = 5 \times 10^{38}$ erg s^{-1}. Using the expression for W_{dipole} we calculate $B_P \simeq 5 \times 10^{12}$ G.

Pulsar spindown is not entirely caused by electromagnetic radiation. A small asymmetry can for example lead to significant gravitational radiation.

Period Glitches. Lengthy observations of the Crab and Vela pulsars reveal abrupt accelerations, followed by a slowing of the rotation, after which the spindown rate returns to the value observed before the glitch.

In the Crab pulsar the period initially decreases by $\Delta P/P \simeq -10^{-8}$, and in Vela by -2×10^{-6}. In the Vela pulsar the original spindown rate is resumed after 50 days, and in the Crab pulsar after about 10 days.

For the Crab pulsar it is currently thought that the spindown creates tensions in the solid crust of the neutron star. A starquake adjusts the shape of the crust to the flattened equilibrium form, decreasing the moment of inertia and thus increasing the angular frequency. The motion of the crust is first communicated to charged particles via the magnetic field (characteristic timescale 100 s), then to the neutron superfluid, with a longer relaxation time. From the observed rate of period increase we can calculate the time between two starquakes using the theoretical properties of the solid crust. The result is satisfactory for the Crab pulsar, with an interval of about 10 years between glitches.

On the other hand the calculated interval for the Vela pulsar is of the order of 10^5 y, while the four glitches observed were separated by 2.5, 4.1, and 2.8 years. One must envisage other processes, for example connected to the quantum properties of a rotating superfluid. The quantisation of fluid

circulation leads to the existence of quantised vortices, each one having a fluid circulation

$$\oint v dl = \frac{h}{2m_n},$$ (8.65)

where v is the fluid velocity and $2m_n$ the mass of a neutron pair. These quantised vortices can be pinned to the crust, and the glitches would be caused by catastrophic unpinning of vortex lines.

8.4.3 Gamma-Ray Bursts

The timing of simultaneous detections of gamma-ray bursts by several satellites (at least three) allows one to find the direction of the emission by stereoscopy. Several gamma-ray-burst sources have been localised in this way. In all there are about 5000 events per year in the Galaxy with flux above 5.5×10^{-9} erg cm^{-2}, and 10^5 events with flux above 10^{-10} erg cm^{-2}. The spectrum is a combination of thermal radiation ($kT \sim$ keV), inverse Compton radiation, and synchrotron emission. The total energy emitted is of the order of 10^{38} erg per burst.

The main hypothesis is as follows. In a binary system, a small amount of mass ($10^{-13} M_\odot$) is captured by a neutron star and flows down fieldlines to accumulate near the magnetic poles. This matter is confined by the magnetic field and undergoes nuclear reactions which become explosive and produce the gamma-ray burst. Each neutron star can thus produce a burst every thousand years, which is compatible with the estimated number of neutron stars in the Galaxy (about 10^9) and the frequency of observed gamma-ray bursts.

The Gamma Ray (or Compton) Observatory, launched in 1991, records seconds-long gamma pulses several times a day. The sources of the bursts are distributed isotropically over the sky, which is hardly compatible with an origin in our Galaxy.

8.4.4 X-ray Sources

Observations of the sky with X-ray instruments of good angular resolution (the *Uhuru* mission 1970, and 10 missions since then) have revealed a large number of point X-ray sources, of which many belong to binary systems and some show pulses.

The X-rays in some cases are thermal emission from an accretion disc around a neutron star or possibly a black hole. The disc is formed by capture of matter from the companion star, which often fills the Roche lobe. The disc luminosity is of the same order as the Eddington limit, about 10^{38} erg s^{-1}.

The accreted matter eventually falls on to the neutron star. In pulsing sources accretion is accompanied by spinup, just as expected from the extra angular momentum supplied by the accreting matter.

8.5 Type II Supernovae

As we have seen, type II supernovae are characterised by strong hydrogen lines. The total ejected mass can be estimated from spectroscopic data and is of the order of ten solar masses. The idea that SN II result from the evolution of massive stars is supported by their presence in the arms of spiral galaxies. These are young objects originating in rapidly evolving, i.e. massive, stars. The problem thus reduces to studying the evolution of massive stars up to the point when the explosion disperses most of the stellar mass.

Depending on its mass, and parameters such as rotation, the progenitor star may have rather different characteristics, and the light curve in particular can vary widely. However, in all cases the main events occur in a central region of mass no more than about $1.3\,M_\odot$.

8.5.1 Pre-supernova Models

As we have seen in Chap. 3, the evolution of massive stars involves a competition between the mass loss rate and the speed with which the outer radiative region mixes with the convective core. We can follow the evolution of such stars given a good phenomenological description of the mixing process. The result, confirmed by studies of the Per OB1 association, is that stars rotating more rapidly than about 300 km s^{-1} evolve homogenously and remain as O stars, while more slowly rotating stars evolve towards the red-giant branch. These stars have large hydrogen-rich envelopes, whereas the homogeneous stars evolve as blue and helium rich.

It is conventional to study the evolution of a non-rotating star without mass loss, from the formation of the helium core up to the collapse of the iron core. The beginning of the collapse will be discussed in Sect. 8.5.2. We note that this last process is independent of the outer layers and any mass loss from them.

Using the results of Chap. 3 we summarise the evolution of massive stars in the pre-supernova stage as follows.

(a) *In the range* 8–11 $M_\odot$ there are a great variety of evolutionary tracks. Near $8\,M_\odot$ ignition of a degenerate carbon core becomes possible. Near $11\,M_\odot$ there are 6 successive nuclear burning stages (hydrogen, helium, carbon, neon, oxygen, and silicon).

In the range 8–10 $M_\odot$ a helium core of 2.2 to $2.5\,M_\odot$ forms. Helium burning is followed by carbon burning (mainly ^{12}C + ^{12}C → ^{20}Ne + ^{4}He). Neon 20 is less bound than oxygen 16 so that for increasing temperature the main subsequent reaction is

$$^{20}\text{Ne} + \gamma \to {}^{16}\text{O} + {}^4\text{He}\,.$$
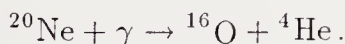
A degenerate oxygen–neon core results finally.

Contraction continues to a density of 2.5×10^{10} g cm^{-3}. Oxygen burning ($^{16}\text{O} + {}^{16}\text{O} \rightarrow {}^{28}\text{Si} + {}^{4}\text{He}$) soon begins in the contracting core because of electron capture by ^{20}Ne and ^{24}Mg. The collapse continues and the unstable core (near the Chandrasekhar limit) reaches the nuclear density without undergoing silicon burning and formation of elements in the iron peak.

For 10 to 11 $M_\odot$ stars with a helium core of 2.5 to 2.8 $M_\odot$, a hydrostatic iron core can form, with complicated instabilities associated with iron burning at the boundary.

(b) *For stars more massive than* 11 $M_\odot$, the main process is slow silicon burning (main reaction $^{28}\text{Si} + \gamma \rightarrow {}^{24}\text{Mg} + {}^{4}\text{He}$), followed by the capture of alpha particles up to the iron peak in the stellar core. The final result gives the mass of the iron core at the start of neutronisation. To find the mass of the iron core (which plays a critical role in the supernova phenomenon) requires the use of the best available microscopic data, including as complete a reaction network as possible (incorporating in particular the abundance of ^{30}Si coming from the beginning of neutronisation) and a choice of the initial abundance of the elements (we take "cosmic" abundances). It is also necessary to use very fine spatial and temporal grids in the numerical calculation of the evolution. In the reactions near the iron peak it is the speed of the reaction $^{45}\text{Sc}(p,\gamma)^{46}\text{Ti}$ which controls the rate of iron formation.

During the silicon-burning phase the core is convective, and the silicon-burning layer moves outwards, surrounded by a convective layer, at a speed determined by the formation rate of the iron core. Figure 8.9 gives the abundance distribution at the beginning of the collapse ($v \simeq 1000$ kms^{-1}) for a 25 $M_\odot$ star.

The mass of the iron core is 1.41 $M_\odot$ and 1.35 $M_\odot$ for stellar masses of 20 and 25 $M_\odot$ respectively. The entropy per baryon at the centre (whose importance will be clarified in Sect. 8.5.2) is 0.69 and 0.73; and the numbers Y_e of electrons per baryon at the centre are 0.422 and 0.423 respectively.

8.5.2 Collapse

Electron Capture. The basic problem in treating the collapse is the electron capture rate. These captures decrease the electron pressure and allow core collapse. Electrons are mainly captured by nuclei rather than protons, which have low abundance. The number Y_e of free electrons per baryon decreases slowly. The fractional electron pressure deficit

$$F = \frac{4}{3} \frac{Y_e^{(i)} - Y_e}{Y_e^{(i)}} \tag{8.66}$$

(where $Y_e^{(i)}$ is the value of Y_e at the start of the collapse) gives the infall velocity through

$$\frac{\text{d}^2 R}{\text{d}t^2} = -\frac{GMF}{R^2}. \tag{8.67}$$
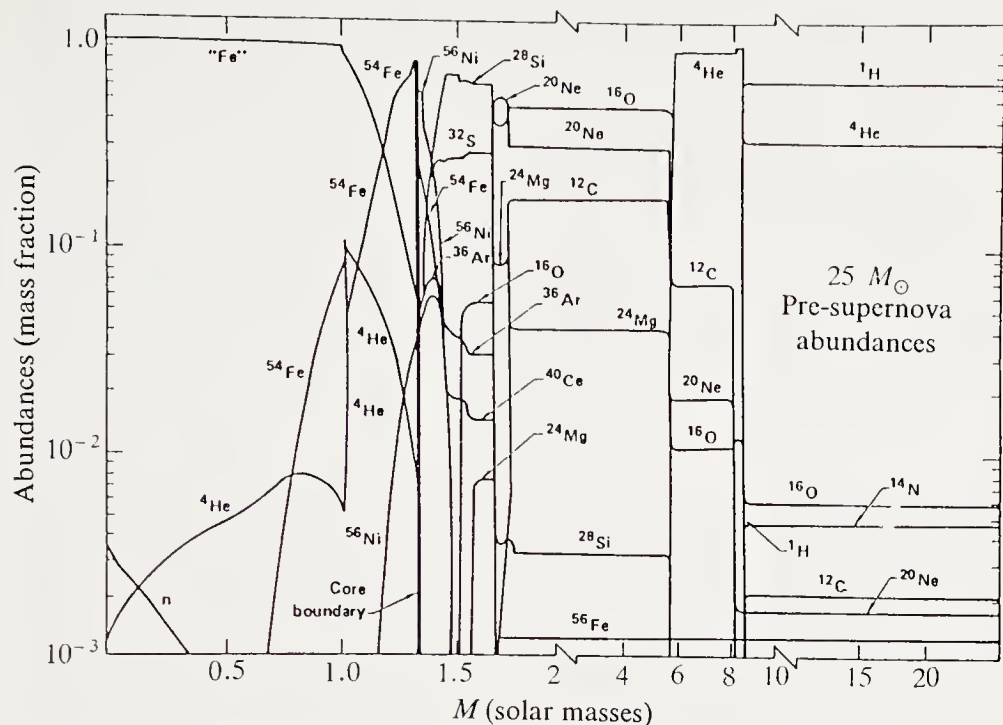
Fig. 8.9. Abundances in a pre-supernova of 25 $M_\odot$. The chemical composition is given at the point where the collapse velocity has reached 1000 km s^{-1} in all zones. The *horizontal lines* represent abundances in (or created in) convection zones. The curve marked "Fe" is an exception. This gives the sum of all the mass fractions of the iron group richer in neutrons than ^{54}Fe. The central density is 2×10^9 g cm^{-3} and the central temperature is 5×10^9 K. (From S.E. Woosley, in *Problems of Collapse and General Relativity*, ed. by D. Bancel and M. Signore, NATO series, Reidel 1984. Reproduced by kind permission of Kluwer Academic Publishers)

For $F \simeq 0.1$ the infall velocity is less than free fall. At the point where neutrinos are trapped in the moving matter, $\rho \simeq 5 \times 10^{11}$ g cm^{-3} (see Sect. 5.2.2) the infall velocity is 2.45×10^9 cm s^{-1}, and the core radius is 40 km.

Neutrino Trapping. The neutrino mean free path, according to Bethe, Brown, Applegate, and Lattimer (1979) (BBAL) is

$$\lambda_\nu = 10_{km} \left(\frac{100}{\rho_{10}} \right) \left(\frac{\overline{N}^2 X_h}{6A} + X_n + \frac{5}{6} X_p \right)^{-1} \left(\frac{10\,MeV}{\varepsilon_\nu} \right)^2 , \qquad (8.68)$$

where X_h, X_n and X_p are the mass fractions of heavy nuclei, neutrons, and protons. $\overline{N}$ is the average number of neutrons per heavy nucleus with atomic weight A and ε_ν the neutrino energy.

In statistical equilibrium BBAL give $\rho_{10} = 50$ (where $\rho_{10} = \rho/(10^{10}$ g cm^{-3}), $A = 10^5$, and $\overline{N} = 67$. After time t, neutrinos have diffused a distance

$$r_{\text{diff}} \simeq \left(\frac{\lambda_\nu ct}{3} \right)^{1/2} \tag{8.69}$$

corresponding to a diffusion velocity

$$v_{\text{diff}} = \frac{r_{\text{diff}}}{t} . \tag{8.70}$$

At the same time the infall velocity of the homologously contracting core boundary is (8.66, 67)

$$u = - \left(\frac{2FGM}{r} \right)^{1/2} \tag{8.71}$$

with (from Sect. 5.2.2)

$$\rho_{10} = 3 \times \left(\frac{10^7}{r} \right)^3 . \tag{8.72}$$

We regard the neutrinos as trapped once their diffusion velocity is smaller than the infall speed. With $M = 1\,M_\odot$, $F = 0.1$ we find that neutrinos are trapped once $\rho_{10} = 30$, corresponding to a radius $r = 46$ km and an infall velocity of about $20\,000$ km s^{-1} at the core boundary.

Once the neutrinos are trapped the total lepton number $Y_l = Y_e + Y_\nu$ remains fixed. Also there is equilibrium between direct and inverse beta decays, so that the chemical potentials μ_e, μ_r, μ_n and μ_p satisfy

$$\mu_e - \mu_r = \mu_n - \mu_p . \tag{8.73}$$

The Chandrasekhar mass changes from the point that the neutrinos are trapped because of their contribution to the pressure. If Y_L is the lepton number per baryon before trapping and Y_ν this number after trapping, and $f = Y_\nu/Y_L$, the Chandrasekhar mass is immediately multiplied by

$$((1 - f)^{4/3} + 2^{1/3} f^{4/3})^{3/2} .$$

For $f \simeq 0.2$ to 0.25, the Chandrasekhar mass is reduced by a factor of 0.84. This reduction must be taken into account in calculating the energy liberated in the explosion.

Simplifed Collapse Dynamics (Self-similar Collapse). Many partial-differential equations involving time and one space variable admit solutions of the form $f(x^\alpha t^\beta)$. This method, discovered by Sedov, gives exact asymptotic solutions. For core collapse of a supernova we have to make a further simplification, by replacing the equation of state by a polytropic relation

$$P = K\rho^\gamma ,$$

where P is the pressure, ρ the density, γ the adiabatic compressibility, and K a constant whose value depends only on the entropy of the system. This approximation is justified to the extent that calculations before neutrino trapping show that the specific entropy is effectively constant. From this point on it remains constant because there are no neutrino energy losses.

Self-similar solutions are useful in that parameters such as the initial central density only affect the early stages of the motion, memory of them being lost at late times. The solution then depends only on two physical quantities, the constant K (depending only on the entropy) and the gravitational constant G. We introduce the dimensionless combination

$$X = K^{-1/2} G^{(\gamma-1)/2} r(-t)^{\gamma-2} . \tag{8.74}$$

The time origin is the singular point at which the density becomes infinite. The self-similar solution neglects the rebound produced once the central density reaches the nuclear value. To describe the density, velocity, mass within radius r, and the total energy within r we introduce the dimensionless functions $D(X), V(X), M(X), E(X)$ and the necessary dimensional combinations of K and G:

$$\begin{aligned}
\rho &= G^{-1}(-t)^{-2} D(X), \\
\nu &= K^{1/2} G^{(1-\gamma)/2}(-t)^{1-\gamma} V(X), \\
m(r) &= K^{3/2} G^{(1-3\gamma)/2}(-t)^{4-3\gamma} M(x), \\
\varepsilon(r) &= K^{5/2} G^{(3-5\gamma)/2}(-t)^{6-5\gamma} E(X),
\end{aligned} \tag{8.75}$$

where $M(X)$ and $E(X)$ are defined by

$$M(X) = 4\pi \int_0^X X^2 D(X) \mathrm{d}X, \tag{8.76}$$

$$E(X) = 4\pi \int_0^X X^2 D(X) \left[\frac{1}{2} V^2(X) + \frac{1}{\gamma-1} D^{\gamma-1}(X) - \frac{M(X)}{X} \right] \mathrm{d}X.$$

The latter expression contains contributions from the kinetic, internal, and gravitational energies. We find differential equations for D, V, M, E by inserting the expressions for $\rho, v, m(r)$, and $\varepsilon(r)$ into the equations of hydrodynamics (conservation of mass, momentum and energy, and the equation of state). We have to satisfy the following boundary conditions:

(a) As t tends to zero and X tends to infinity the density must remain finite, forcing the asymptotic relation

$$D(X) \sim X^{-2/(2-\gamma)}, \tag{8.77}$$

eliminating t from the expression for the density, which tends to zero at large times.
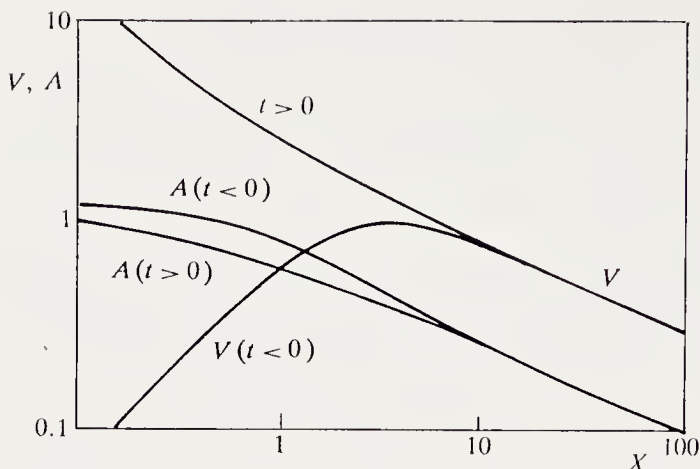
Fig. 8.10. The reduced velocity V and sound speed A before $(t < 0)$ and after $(t > 0)$ the rebound

(b) Across the singular point (of the same type as in the theory of stellar winds) the solution coming from infinity must match to the regular solution as X tends to zero.

The *critical solution* found in this way automatically satisfies the condition of zero total energy.

We can thus use the equations for $t > 0$ (giving the post-explosion solution) if we take as initial conditions the density and velocity distributions from the asymptotic solution for large X $(t = 0)$. As an example, Fig. 8.10 shows the behaviour of the reduced velocity and sound speed $V(X), A(X)$ for $\gamma = 1.3$. We define the inner core by the velocity maximum. This region behaves homologously, with infall speed essentially proportional to X; the outer core has a velocity distribution similar to that of an accretion flow.

The mass of the inner core changes slowly with time. For $\gamma = 1.3$ it varies as $(-t)^{0.1}$. The velocity inside the sonic point $(V + A = 0)$ is subsonic, allowing pressure perturbations to travel ahead of the collapse and cause the homologous rearrangement of the density and velocity. The velocity in the outer core is by contrast highly supesonic and close to free fall.

Defining the Chandrasekhar mass M_{Ch} as that mass of a static polytrope with the same central density as the collapsing inner core we find that the mass of the inner core is related to M_{Ch} by

$$M_{\mathrm{IC}} = (1 + f)M_{\mathrm{Ch}} , \tag{8.78}$$

where f is fixed by the average value of $(4 - 3\gamma)$ in the collapsing inner core. For realistic equations of state

$$f = 0.1 . \tag{8.79}$$

Rebound. We assume for simplicity that the equation of state becomes stiffer once the matter reaches the nuclear density $\rho = 2.7 \times 10^{14}$ g cm^{-3}.

370

Matter becomes almost incompressible and accumulates in a sphere close to hydrostatic equilibrium. The pressure pulse caused by the increased stiffness of the equation of state cannot propagate spatially beyond the sonic point while the density and velocity are continuous there. The collapse proceeds, with the pressure inside the sonic point continuing to grow until a shock wave forms.

For zero total energy we can assume that the binding energy B of the core and the shock energy have zero sum, so that the energy available for the rebound is of the order of $-B$. To calculate the binding energy requires an equation of state. We can however estimate it as the gravitational binding energy of the excess mass over the Chandrasekhar mass. For a rigid constant-density sphere we get

$$B = -\frac{3}{5}\frac{GM^2}{R}f \tag{8.80}$$

with $4\pi R^3 \rho_0 /3 = M$. Taking $\rho_0 = 2.7 \times 10^{14}$ g cm^{-3}, $f = 1$ and $M = 0.8\,M_\odot$ we find

$$-B = 6.2 \times 10^{51}\ \mathrm{erg}\,, \tag{8.81}$$

which is indeed of the order given by more sophisticated calculations.

We would like to know if the shock can eject the envelope and produce a supernova. There are three problems in calculating the shock propagation velocity.

(a) The high post-shock temperature causes further electron captures, decreasing the electron pressure and the shock propagation velocity.

(b) Once the density ahead of the shock falls below $\rho \simeq 10^{11}$ g cm^{-3} neutrinos can leave the shock region and remove energy.

(c) An iron-rich region forms around the inner core of mass M_{IC}. The temperature is at least 20 MeV, causing iron to be photodissociated at 8.9 MeV per nucleon. We thus need a dissociation energy E_D

$$E_D = 17(M_{Fe} - M_{\mathrm{IC}}) \times 10^{51}\ \mathrm{erg}\,, \tag{8.82}$$

where M_{Fe}, M_{IC} are in solar masses. The mass of iron depends on the entropy at the start of the collapse (hence the importance of this quantity). The initial shock energy $-B$ is evidently just enough to supply the neutrino losses (about 2×10^{51} erg) and the iron photodissociation energy. Woosley's latest models give $M_{Fe} = 1.35$; using $M_{\mathrm{IC}} = 1\,M_\odot$ we find a binding energy $-B \simeq 9 \times 10^{51}$ erg and an iron photodissociation energy $E_D \simeq 7 \times 10^{51}$ erg. There remain only 10^{51} erg to push the shock to the star's surface.

Element Synthesis. Heavy elements are formed from small quantities of iron in three main processes, clearly visible in the nuclear $(A - Z, Z)$ plane (Fig. 8.11).
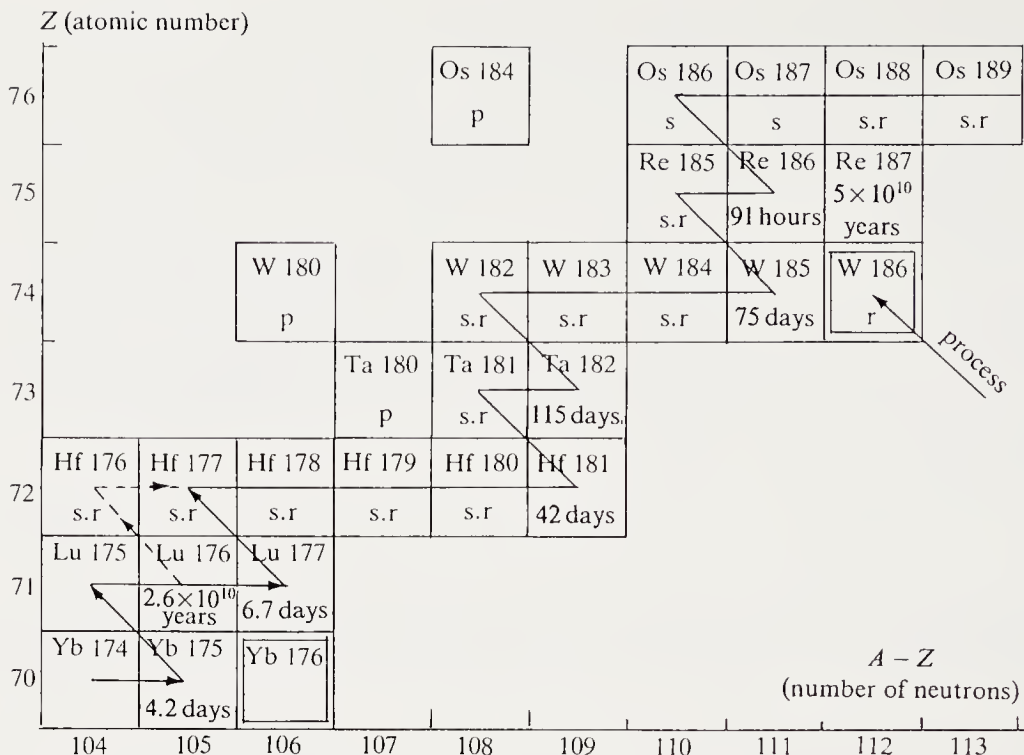
Fig. 8.11. Part of the nuclear (N, Z) plot. The *continuous zig-zagging line* shows the path of *slow* (*s*) processes. In this part of the plot the two species ^{186}W and ^{176}Yb (tungsten and ytterbium) can only be formed by *rapid* (*r*) processes

The first process is slow (*s* process). The capture timescale for a neutron is of the order of 10 to 100 years. If a radioactive element is formed, beta decay occurs before another neutron is captured. This mechanism gradually populates the valley of stability in the direction of increasing atomic weight. These reactions occur at 200–300 milion degrees, requiring a neutron source. The *s* process may occur mainly in pulses of intermediate-mass stars during the AGB phase.

Proton captures (*p* process) occur during high-temperature phases and should be the main nucleosynthetic process during an SN II explosion.

The rapid (*r*) process requires a significant neutron flux. Neutron-rich elements form up to the point that beta decays, occurring in less than a second, bring the element back into the stable region. Many elements can only be formed by the *r* process, in particular radioactive elements (uranium) and some others such as those in Fig. 8.11 (^{176}Yb and ^{186}W). If we calculate the detailed abundances produced by the *s* process and subtract them from solar abundances, we find a curve whose form agrees perfectly with the abundances produced *only* by the *r* process (Fig. 8.12). Supernova explosions may be the favoured site of the *r* process, although there are doubts because of the difficulty of producing enough neutrons.

372
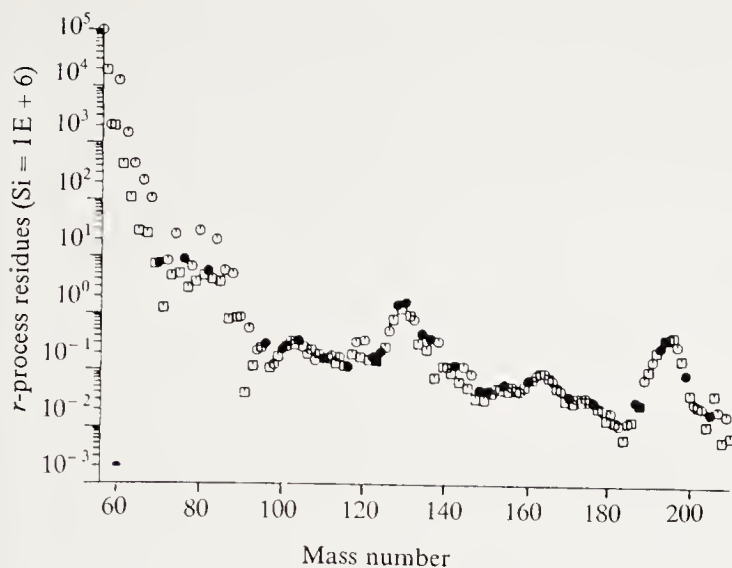
Fig. 8.12. Contribution of the r process to element formation. The figure shows the abundances found by subtracting s process abundances from solar abundances. The *solid dots* show elements accessible only via r processes. The whole curve shows the importance of the r process, whose site is still uncertain. (From H. Beer, in *Advances in Nuclear Astrophysics*, ed by E. Vangioni-Flam et al. Editions Frontières 1986, Fig. 5, p. 382)

8.5.3 Supernova 1987A

On 1987 February 23 an International Astronomical Union telegram announced Ian Shelton's discovery of a 5th magnitude object in the Large Magellanic Cloud (LMC), obviously a supernova. This supernova has three remarkable features.

(a) The progenitor star was a blue supergiant, while the typical SN II progenitors are red supergiants, the enormous radius being important in the formation of the usual light curve.

(b) The light curve of SN 1987A is completely atypical. Maximum light is four magnitudes fainter than for typical SN II; the brightness in the blue began to increase a month after the explosion, reached a secondary maximum after three months, and did not begin to decline until the beginning of June.

(c) There were simultaneous neutrino detections in Japan (Kamiokande II: 11 events), in the USA (Irvine–Michigan–Brookhaven, Cleveland: 8 events), and in the USSR (Baksan: 5 events).

These features raise many problems for our understanding of the structure of the pre-supernova and the theory of the explosion. The observations are extremely important and detailed; we summarise the main conclusions.

The Nature of the Progenitor.   We have first to explain why the progenitor was a blue supergiant. This may indicate different evolutionary tracks for stars of low metallicity, as in the LMC, and high metallicity, as in our Galaxy.
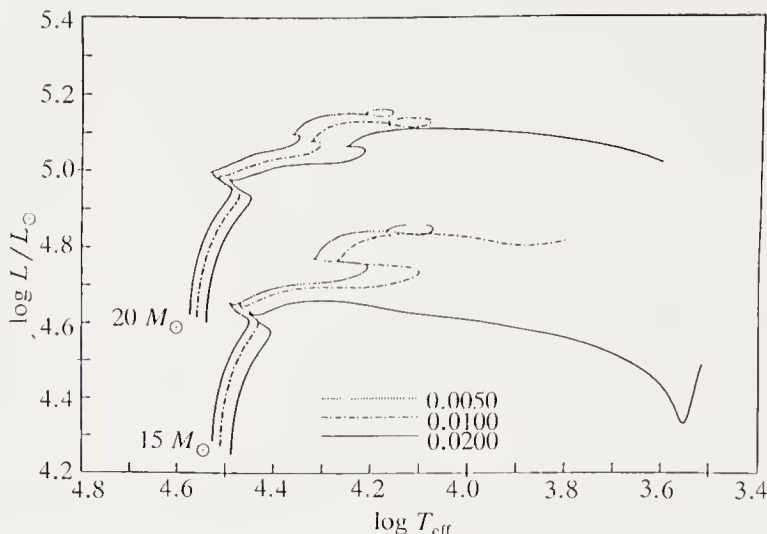
Fig. 8.13. Evolutionary tracks in the HR diagram for stars with various heavy-element abundances. (From W. Hillebrandt et al., *Nature* **327** (1987) p. 597. Reproduced by kind permission of *Nature*; © 1987 Macmillan Magazines Ltd.)

Homogeneous evolution would give a massive helium-rich star, whose explosion would result in an SN Ia. These SN Ia resemble SN I in their lack of hydrogen but SN II in the mass of their ejecta.

A strong stellar wind, leaving a blue naked remnant star has been suggested, with the progenitor then being a Wolf–Rayet star in its final stages. However, the progenitor, Sanduleak 62212, was a blue B3I supergiant and not a Wolf–Rayet.

Evolutionary tracks on the HR diagram show that changing Z from 0.02 to 0.005 is enough to make the star finish its evolution at an effective temperature near 16 000 K (Fig. 8.13). Ignition of ^{12}C–^{12}C can begin in a 15–25 $M_\odot$ star while it still appears as a blue supergiant. Because the thermal timescale is so short (comparable to the nuclear timescale) we do not expect evolution to the red-giant branch.

Light Curve. The rise in the light curve seems to be caused by the arrival at the photosphere of the expanding gas of the disintegration energy of radioactive nickel ^{56}Ni and then radioactive cobalt ^{56}Co, with a timescale of 77.12 days. The production of 0.075 $M_\odot$ of ^{56}Co predicted by Woosley's models (collapse of a 6 $M_\odot$ helium core) agrees with what one can conclude from the light curve after 8 months of observation. Cobalt and nickel are identified in the infrared mainly via a Co II line at 10.5 microns and a Ni II line at 6.6 microns. The iron ^{56}Fe found in the expanding gas results from the rapid radioactive decay of ^{56}Ni and corresponds to a mass of 0.07 $M_\odot$.

Neutrinos. The detectors at Kamiokande, Cleveland (IMB), and Baksan all work on the same principle: a large volume of very pure water is surrounded by thousands of photomultipliers and sited in a deep mine to min-

imise the cosmic-ray background. The IMB detector contains 6800 tons of water, and Kamiokande II 2140 tons, but it has photomulipliers covering a larger fraction of its surface. The detection threshold of Kamiokande II is 8 MeV while those of IMB and Baksan are 20 and 10 MeV, the latter despite a much smaller number of free protons. The main reaction is antineutrino capture by protons:

$$p + \tilde{\nu}_e \rightarrow n + e^+$$

whose cross-section is about 100 times larger than the elastic reaction $\nu_e + e^- \rightarrow n + e^-$.

There are two main questions about the observed event numbers: (a) Does collapse and explosion theory of SN II correctly predict these numbers? (b) Can theory explain the dispersion in arrival times of these antineutrinos?

(a) With the capture cross-section

$$\sigma_{\tilde{\nu}_e p} \simeq 7.5 \times 10^{-44} E_1^2 (\text{MeV}) \, \text{cm}^2 = \sigma_0 E_{\text{MeV}}^2$$

we can calculate the total number of neutrinos emitted by the supernova and the total neutrino energy, assuming an energy spectrum for them.

Letting $x = E_0/kT$, where E_0 is the detector threshold, d the source distance, and n the number of events at a detector containing n_p free protons, we find a total energy

$$W = \frac{24\pi d^2 n}{\sigma_0 n_p} \frac{1\,\text{MeV}}{E_0(\text{MeV})} \frac{e^x}{x(x^4 + 4x^3 + 12x^2 + 24x + 24)},$$

assuming a thermal non-degenerate neutrino spectrum. This has a minimum value of 0.401 for $x = 4.1335$. This corresponds to a minimum energy of 0.443×10^{55} erg. With a threshold around 10 MeV this implies a minimum emission temperature of 2.41 MeV.

The average energy of the antineutrinos detected at Kamiokande was about 14.66 MeV. It is related to the threshold energy E_0 by the approximate relation

$$\frac{\langle E \rangle}{E_0} = \frac{5}{x} \left(1 + \frac{x^5}{120} \frac{1}{1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}} \right).$$

With a 10 MeV threshold we find a temperature $T = 2.21$ MeV, and $x = 4.511$. The total energy is than 0.462×10^{53} erg. Assuming equipartition of energy between the 6 neutrino and antineutrino species this implies a liberated energy of 2.77×10^{53} erg.

We have seen that the total energy of the collapsing core is close to zero. This implies internal energy equal to $-E_{\text{grav}}$. This energy is available for neutrino emission and can be calculated for present models. The most

likely give $-E_{\mathrm{grav}}$ between 2.5 and 3×10^{53} erg, compatible with the results from Kamiokande, IMB, and Baksan, assuming a distance $d = 5$ kpc for the LMC.

(b) Neutrino escape is a diffusion process. If we simplify drastically and estimate the diffusion time from the relation

$$\tau_{\mathrm{diff}} = \frac{R^2}{D}$$

with

$$D = \tfrac{1}{3} \lambda_\nu c \,,$$

we find, with $T = 2.5$ MeV, $\overline{N}^2/A \simeq 4000$ (from Bonche and Vautherin 1981), $\rho = 2.5 \times 10^{14}$ g cm^{-3}, and $R = 15$ km,

$$(R^2/D) \cong 2.4\,\mathrm{s} \,,$$

which is indeed of the observed order.

Appendix

Calculation of the Relativistic Correction for White-Dwarf Stability

To discuss the stability of a white dwarf, taking account of general relativity, we write the total energy in the form

$$E = E_{\text{int}} + E_{\text{grav}} + \Delta E_{\text{int}} + \Delta E_{GR} . \tag{A.1}$$

To a first approximation only the first two terms are important. We can evaluate them for a polytrope. With $P = K\rho^\gamma, n = 1/(\gamma - 1)$

$$E_{\text{int}} = K\rho_c^{(1/n)} M \frac{n}{|\xi_1^2 \theta'|} \int_0^{\xi_1} \xi^2 \theta^{n+1} \mathrm{d}\xi , \tag{A.2}$$

$$E_{\text{grav}} = (4\pi\rho_c)^{1/3} \frac{GM^{5/3}}{|\xi_1^2 \theta'|^{5/3}} \int_0^{\xi_1} \xi^3 \theta' \theta'' \mathrm{d}\xi , \tag{A.3}$$

$$= (\rho_c)^{1/3} GM^{5/3} \frac{3}{5-n} \frac{|4\pi\xi_1^2\theta'|^{1/3}}{\xi_1} . \tag{A.4}$$

The term ΔE_{int} represents the contribution to the internal energy from the non-zero temperature. Per unit mass this contribution is

$$u = \frac{\varepsilon_e - M_e c^2 n_e}{\rho} . \tag{A.5}$$

Using the degeneracy parameter x,

$$x = \left(\frac{3\pi^2 \rho \lambda_e^3}{\mu_e m_U} \right)^{1/3} , \tag{A.6}$$

where λ_e is the de Broglie wavelength of the electron and m_U the atomic mass unit, we have

$$u = \frac{3}{4} \frac{m_e c^2}{\mu_e m_U} \left(x - \frac{4}{3} + \frac{1}{x} + \dots \right) . \tag{A.7}$$

The first term is just $3P/\rho$ and is used to calculate E_{int}. The second (constant) term does not appear in the variational calculation. This leaves the term in $1/x$, which gives

$$\Delta E_{\text{int}} = \frac{3}{4} \frac{m_e c^2}{\mu_e m_U} \int \frac{\mathrm{d}m}{x}, \tag{A.8}$$

where

$$\Delta E_{\text{int}} = \frac{m_e^2 c^3}{\hbar(\mu_e m_U)^{2/3}} M \rho_c^{-1/3} k_3 \tag{A.9}$$

with

$$k_3 = \frac{3}{4} \frac{1}{(3\pi^2)^{1/3}} \frac{1}{|\xi_1^2 \theta'|} \int_0^{\xi_1} \xi^2 \theta^2 \mathrm{d}\xi. \tag{A.10}$$

Writing ρ_0, m_0 for the baryon density and mass, and $r_0 = (3v/4M)^{1/3}$ for the radius of the sphere of volume v defined by

$$\mathrm{d}v = \left(1 - \frac{2m}{r}\right)^{-1/2} 4\pi r^2 \mathrm{d}r, \tag{A.11}$$

we can express the relativistic correction (assumed small) as

$$\Delta E_{GR} = \int_0^R \rho_0 \mathrm{d}v \left[-u\frac{m}{r} - \frac{1}{2}\left(\frac{m}{r}\right)^2 + \frac{m_0}{r_0} - \frac{m}{r} \right], \tag{A.12}$$

and we find in standard units

$$\Delta E_{GR} = -k_4 \frac{G^2}{c^2} M^{7/3} \rho_c^{2/3}$$

with

$$k_4 = \frac{(4\pi)^{2/3}}{(5-n)|\xi_1^2 \theta'|^{7/3}} \left\{ -\frac{5+2n-n^2}{n+1} 2 \int_0^{\xi_1} \xi^3 \theta' \theta^{n+1} \mathrm{d}\xi + \frac{3}{2}(n-1) \right.$$

$$\left. \times \int_0^{\xi_1} \xi^4 \theta'^2 \theta^n \mathrm{d}\xi \right\}. \tag{A.13}$$

With the parameters A, B, C, D defined in equations (8.33), we find

$$E = (AM - BM^{5/3})\rho_c^{1/3} + CM\rho_c^{-1/3} - DM^{7/3}\rho_c^{2/3}. \tag{A.14}$$

Differentiating with respect to ρ_c and setting the result equal to zero we find the equilibrium mass as a function of ρ_c. Differentiating again we find the stability limit. For $A/Z = \mu_e$ this gives a central density

$$\rho_c = 2.646 \times 10^{10} (\mu_e/2)^2 \text{ g cm}^{-3}.$$

378

Bibliography

Chapter 1

Basic data on stars and stellar atmospheres is given in the following:

Allen, C.W. (1976) *Astrophysical Quantities* (Athlone, London)

Böhm–Vitense, E. (1989) *Introduction to Stellar Astrophysics, vols 1–3* (Cambridge University Press)

Golay, M. (1984) *Introduction to Astronomical Photometry* (Reidel, Dordrecht)

Gray, D.F. (1976) *The Observation and Analysis of Stellar Photospheres* (Wiley, New York)

Jaschek, C., Jaschek, M. (1987) *The Classification of Stars* (Cambridge University Press)

Mihalas, D. (1978) *Stellar Atmospheres*, 2nd edn (Freeman, San Francisco)

Mihalas, D., Binney, J. (1981) *Galactic Astronomy* (Freeman, San Francisco)

Pecker, J.C., Schatzman, E. (1959) *Astrophysique Générale* (Masson, Paris)

Unsöld, A. (1968) *Physik der Sternatmosphären* (Springer, Berlin Heidelberg)

Stellar spectra of various types are given in

Abt, H.A., Meinel, A.B., Morgan, W.W., Tapscott, J. (1969) *An Atlas of Low-Dispersion Stellar Spectra* (Kitt Peak Nat. Obs. and Yerkes Obs.)

Absolute magnitude and distance determinations are discussed in

Hauck, B. (1973) in *Problems of Calibration of Absolute Magnitudes and Temperature of Stars*, ed. by B. Hauck, B. Westerlund, p. 117

Hayes, D.S., Pasinetti, L.E., Davis Philip, A.G. (eds.) (1985) *Calibration of Fundamental Stellar Quantities*, IAU Symp. 111 (Reidel, Dordrecht)

Heck, A. (1978) *Vistas Astron.* **22**, 221

Van de Kamp, P. (1981) *Stellar Paths* (Reidel, Dordrecht)

See also the following review articles:

Böhm-Vitense, E. (1981) *Annu. Rev. Astron. Astrophys.* **19**, 295

Dravins, D. (1982) *Annu. Rev. Astron. Astrophys.* **20**, 61

Mould, J. (1982) *Annu. Rev. Astron. Astrophys.* **20**, 91

Popper, D.M. (1980) *Annu. Rev. Astron. Astrophys.* **18**, 115

Strömgren, B. (1966) *Annu. Rev. Astron. Astrophys.* **4**, 433

A classic book on stellar evolution is

Schwarzschild, M. (1958) *Structure and Evolution of the Stars* (Princeton University Press)

The interaction of radiation and matter is covered in

Oxenius, J. (1986) *Kinetic Theory of Particles and Photons* (Springer, Berlin Heidelberg)
Thomas, R.N. (1965) *Some Aspects of Non-equilibrium Thermodynamics in the Presence of a Radiation Field* (University of Colorado Press, Boulder)

A basic article on photoelectric photometry is

Young, A.T. (1974) *Methods Exp. Phys.* **12A**, 123

Astrophysical observing techniques, stellar and non-stellar, are covered in

Léna, P. (1988) *Observational Astrophysics* (Springer, Berlin Heidelberg)

The reference catalogue for the properties of nearby stars is

Gliese, W. (1969) Catalogue of Nearby Stars, *Veröffent. Astron. Rechen-Inst., Heidelb.*, no. 22

Other works cited in the text are

Beals, C. (1949) *Publ. Dom. Astrophys. Obs. Victoria* 9, 1
Blackwell, D.E., Shallis, M.J. (1977) *Mon. Not. R. Astron. Soc.* **183**, 285
Cayrel, R. (1986) *Astron. Astrophys.* **168**, 81
Cramer, N. (1982) *Astron. Astrophys.* **112**, 330
De Jager, C., Nieuwenhuijzen, H. (1987) *Astron. Astrophys.* **177**, 277
Sandage, A.R. (1957) *Astrophys. J.* **125**, 435

Chapter 2

Books giving a general coverage of solar physics are

Athay, R.G. (1976) *The Solar Chromosphere and Corona: Quiet Sun* (Reidel, Dordrecht)
Jordan, S. (ed.) (1981) *The Sun as a Star* (CNRS-NASA, NASA SP-450)
McLean, D.J., Labrum, N.R. (eds.) (1985) *Solar Radiophysics* (Cambridge University Press)
Priest, E.R. (1982) *Solar Magnetohydrodynamics* (Reidel, Dordrecht)
Stix, M. (1989) *The Sun* (Springer, Berlin Heidelberg)
Sturrock, P.A. (ed.) (1986) *Physics of the Sun* vols. 1–3 (Reidel, Dordrecht)
White, O.R. (ed.) (1977) *The Solar Output and Its Variations* (University of Colorado Press, Boulder)

Two books of more historical interest for the study of the Sun are

Pecker, J.C., Schatzman, E. (1959) *Astrophysique Générale* (Masson, Paris)
Thomas, R.N., Athay, R.G. (1961) *Physics of the Solar Chromosphere* (Interscience, New York)

A recent review of Solar System problems is

Encrenaz, T., Bibring, J.P. (1990) *The Solar System* (Springer, Berlin Heidelberg)

Models of the photosphere and chromosphere of the Sun are given in

Avrett, E.H. (1977) in *The Solar Output and Its Variations*, ed. by O.R. White (University of Colorado Press, Boulder) p. 327
Holweger, H., Muller, E. (1974) *Sol. Phys.* **39**, 19
Kurucz, R.L. (1974) *Sol. Phys.* **34**, 17
Kurucz, R.F. (1991) in *Stellar Atmospheres: Beyond Classical Models*, ed. by L. Crivellari, I. Hubeny, NATO ASI Series (Kluwer, Dordrecht)
Stenflo, J.D. (ed.) (1990) *Solar Photosphere: Structure, Convection and Magnetic Field*, IAU Symp. 138 (Kluwer, Dordrecht)
Vernazza, J.E., Avrett, E.H., Loeser, R. (1981) *Astrophys. J. Suppl. Series* **45**, 635

The theory of radiative transfer, stellar atmospheres and various models are discussed in

Borsenberger, J., Gros, M. (1978) *Astron. Astrophys. Suppl.* **31**, 291
Heyvaerts, J., Roueff, E., Encrenaz, P. (not yet published) Processus Physiques et Astrophysiques I (InterEditions/Editions du CNRS, Paris)
Kurucz, R.L. (1979) *Astrophys. J. Suppl. Ser.* **40**, 1
Maltby, P. et al. (1986) *Astrophys. J.* **306**, 284 (Appendix A)
Mihalas, D. (1972) *Non-LTE Model Atmospheres for B and O Stars* (NCARTN/STRG)
Mihalas, D. (1978) *Stellar Atmospheres* (Freeman, San Francisco)
Omont, A., Smith, E.W., Cooper, J. (1972) *Astrophys. J.* **175**, 185

Radiative losses are discussed in

Athay, R.G. (1986) in *Physics of the Sun*, ed. by P.A. Sturrock (Reidel, Dordrecht) vol. II, p. 1
McWhirter, R.W.P., Thonemann, P.C., Wilson, R. (1975) *Astron. Astrophys.* **40**, 63

The first paper on the chemical composition of the Sun is

Russell, H.N. (1929) *Astrophys. J.* **70**, 11

Other works cited in the text are

Athay, R.G., Skumanich, A. (1968) *Astrophys. J.* **152**, 211
Ayres, T.R., Linsky, J.L. (1976) *Astrophys. J.* **205**, 874
Brueckner, G.E. (1981) *Solar Active Regions*, ed. by F.Q. Orall (University of Colorado Press, Boulder) p. 113
Dumont, S. (1967) *Ann. Astrophys.* **30**, 861
Pottasch, S.R. (1964) *Space Sci. Rev.* **3**, 816
Roudier, T., Muller, R. (1986) *Sol. Phys.* **107**, 11
Samain, D. (1980) *Astrophys. J. Suppl. Ser.* **44**, 273
Stenflo, J.O. (1973) *Sol. Phys.* **32**, 41

Chapter 3

The classic references on internal structure are still relevant: note particularly the historical analysis of the theory of internal structure in the book by Chandrasekhar.

Bouvier, P., Maeder, A. (eds.) (1977) *Advanced Stages in Stellar Evolution*, Saas-Fee Advanced Course 7 (Swiss Society for Astrophysics and Astronomy, Geneva Observatory, Sauverny), with contributions by Iben, Renzini, and Schramm

Hauck, B., Maeder, A. (eds.) *Astrophysical Processes in Upper Main Sequence Stars*, Saas-Fee Advanced Course 13 (Swiss Society for Astrophysics and Astronomy, Geneva Observatory, Sauverny), with contributions by Cox, Vauclair, and Zahn

Chandrasekhar, S. (1939) *Stellar Structure* (University of Chicago Press; reprinted by Dover)

Cox, J.P., Giuli, R.T. (1968) *Stellar Structure* vols. 1,2 (Gordon and Breach, London)

Eddington, A. (1930) *The Internal Constitution of the Stars* (Cambridge University Press)

Encrenaz, T., Bibring, J.P. (1990) *The Solar System* (Springer, Berlin Heidelberg)

Kippenhahn, R., Weigert, A. (1990) *Stellar Structure and Evolution* (Springer, Berlin Heidelberg)

Shapiro, S.K., Teukolsky, S.A. (1983) *Black Holes, White Dwarfs and Neutron Stars* (Wiley, New York)

Schatzman, E. (1958) *White Dwarfs* (North Holland, Amsterdam)

Schwarzschild, M. (1958) *Structure and Evolution of the Stars* (Princeton University Press)

One should add the following review articles:

Iben, I., Jr. (1967) *Annu. Rev. Astron. Astrophys.* **5**, 571

Iben, I., Jr. (1974) *Annu. Rev. Astron. Astrophys.* **12**, 215

Iben, I., Jr., Renzini, A. (1983) *Annu. Rev. Astron. Astrophys.* **21**, 271

Iben, I., Jr. (1984) *Phys. Rep.* **105**, no. 6, 329

The following references are mainly of historical importance.

Truncation of the summation over states through interaction with neighbouring ions was first discussed by

Pannekoek, A. (1938) *Mon. Not. R. Astron. Soc.* **98**, 694

The problem of opacities is constantly evolving. The first reference article is

Cox, A.N. (1965) Stellar absorption coefficient and opacities, in *Stellar Structure*, ed. by L.H. Aller, D.B. McLaughlin (University of Chicago Press) p. 195

See also

Cox, A.N., Tabor, J.E. (1976) *Astrophys. J. Suppl. Ser.* **31**, 271

The first article on the nuclear origin of the energy radiated by the stars is without doubt

Russell, H.N. (1919) *Pabl. Astron. Soc. Pac.* **31**, 205

closely followed by

Perrin, J. (1920) *Revue du Mois* **21**, 113

Tunnelling of potential barriers was discussed by

Gamow, G. (1928) *Z. Phys.* **51**, 204

The first estimates of the energy production rate are given by

Atkinson, R. d'E. (1931) *Astrophys. J.* **73**, 250, 308

The basic paper on thermonuclear energy production is

Bethe, H.A. (1939) *Phys. Rev.* **56**, 239

Several articles in *Annual Reviews of Astronomy and Astrophysics* give the best reaction rates. See also

Reeves, H. (1965) Stellar energy sources, in *Stellar Structure*, ed. by L.H. Aller, D.B. McLaughlin (University of Chicago Press) p. 113

The role of collective effects in potential barrier penetration was first described by

Schatzman, E. (1948) *J. Phys. Radium* **IX**, 46

A correct analysis of collective effects is given by

De Witt, H.E., Graboske, H.C., Cooper, M.J. (1973) *Astrophys. J.* **181**, 439

and subsequently by

Alastuey, A., Jancovici, B. (1978) *Astrophys. J.* **226**, 1034

Reaction rates in strongly correlated plasmas (Wigner solids) are given by

Salpeter, E.E., Van Horn, H.M. (1969) *Astrophys. J.* **155**, 183

The onset of evolution to the giant branch was found by

Schönberg, M., Chandrasekhar, S. (1942) *Astrophys. J.* **96**, 161

The discussion of the Hertzsprung gap (Fig. 3.11) is due to

Maeder, A. (1974) *Astron. Astrophys.* **32**, 177

There are many papers evaluating the ages of clusters using stellar-evolution models. The results are model-dependent. We have used

Maeder, A. (1974) *Astron. Astrophys.* **32**, 177

Important papers on neutrino energy losses are

Adams, J.B., Ruderman, M.A., Woo, C.H. (1963) *Phys. Rev.* **129**, 1383
Fester, G.G., Ruderman, M.A. (1969) *Phys. Rev.* **180**, 1277
Beaudet, G., Petrosian, V., Salpeter, E.E. (1967) *Astrophys. J.* **150**, 979

The notion of a limiting mass (Chandrasekhar limit) is given by

Chandrasekhar, S. (1935) *Mon. Not. R. Astron. Soc.* **95**, 207

The first paper mentioning the possibility of measuring the solar-neutrino flux is

Bahcall, J.N., Fowler, W.A., Iben, I., Jr., Sears, R.L. (1963) *Astrophys. J.* **137**, 344

The solar-neutrino problem has been discussed in many papers as the subject has evolved. A good description of Davis's experiment is given by

Bahcall, J. (1969) *Sci. Am.*, July, p. 28

An important paper on the effect of mass loss on evolution is

Fusi-Pecci, F., Renzini, A. (1976) *Astron. Astrophys.* **46**, 447

The effect of mass loss on evolution to the white-dwarf stage is studied by

Weidemann, V. (1977) *Astron. Astrophys.* **59**, 411

The most important papers on the evolution of binary systems under tidal effects are

Zahn, J.P. (1966) *Ann. Astrophys.* **29**, 313, 489, 685
Zahn, J.P. (1970) *Astron. Astrophys.* **4**, 452
Zahn, J.P. (1975) *Astron. Astrophys.* **41**, 329
Zahn, J.P. (1977) *Astron. Astrophys.* **57**, 383

For discussion of problems of mass exchange see in particular

Lubow, J.H., Shu, F.H. (1975) *Astrophys. J.* **198**, 383
Paczynski, B., Sienkiewicz, R. (1972) *Acta Astron.* **22**, 73

For T Tauri stars see

Bertout, C. (1989) *Annu. Rev. Astron. Astrophys.* **27**, 351

Other works cited in the text are

Abt, H.A. (1961) *Astrophys. J. Suppl. Ser.* 6, 37
Abt, H.A. (1965) *Astrophys. J. Suppl. Ser.* **11**, 429
Abt, H.A., Bidelman, W.P. (1967) *Astrophys. J.* **158**, 1091
Abt, H.A., Moyd, K.I. (1973) *Astrophys. J.* **182**, 809
Blaauw, A. (1961) *Bull. Astron. Inst. Neth.* **15**, 265
Combes, F. et al. (1991) *Galaxies et Cosmologie* (InterEditions/Editions du CNRS, Paris)
Cowling, T.G. (1941) *Mon. Not. R. Astron. Soc.* **101**, 367
Däppen, W., Mihalas, D., Hummer, D.G., Mihalas, B.W. (1988) *Astrophys. J.* **332**, 261
Hummer, D.G., Mihalas, D. (1988) *Astrophys. J.* **331**, 794
Jaschek, C., Gomez, A.E. (1970) *Publ. Astron. Soc. Pac.* **82**, 809
Kopal, Z. (1959) *Close Binary Systems* (Chapman and Hall, London)
Lamers, H.J.G.L.M. (1981) *Astrophys. J.* **245**, 593
Landau, L.D., Lifshitz, E.M. (1959) *Statistical Physics* (Pergamon Press, London – Paris)
Larkin, A.I. (1960) Thermodynamic functions of a low temperature plasma, *Sov. Phys. J.E.T.P.* **II**, 1363 (JETP **38**, 1896, 1960)
Lebreton, Y., Maeder, A. (1986) *Astron. Astrophys.* **161**, 119

Maeder, A. (1975) *Astron. Astrophys.* **40**, 303

Maeder, A. (1976) *Astron. Astrophys.* **47**, 389

Maeder, A., Meynet, G. (1987) *Astron. Astrophys.* **182**, 243

Mihalas, D., Däppen, M., Hummer, D.G. (1988) *Astrophys. J.* **331**, 815

Monet, D.G. (1980) *Astrophys. J.* **237**, 513

Odell, A.P. (1974) *Astrophys. J.* **192**, 417

Smith, L.F. (1968) *Mon. Not. R. Astron. Soc.* **138**, 109

Smith, L.F. (1973) in *Wolf–Rayet and High Temperature Stars*, IAU Symp. 49, ed. by
    M.K.V. Bappu, J. Sahade (Reidel, Dordrecht) p. 15

Sweigart, A.V., Gross, P.G. (1978) *Astrophys. J. Suppl. Ser.* **35**, 405

Underhill, A. (1967) *Bull. Astron. Inst. Neth.* **19**, 173

Van Albada, G.B. (1968) *Bull. Astron. Inst. Neth.* **20**, 47

Webbink, R.F. (1979) in *White Dwarfs and Variable Degenerate Stars*, IAU Colloq. 53,
    ed. by H. Van Horn, V. Weidemann (University of Rochester), p. 426

Weizsäcker, C.F. von (1937) *Phys. Z.* **38**, 176

Weizsäcker, C.F. von (1938) *Phys. Z.* **39**, 633

Chapter 4

The problem of mass loss and stellar winds is still under active discussion.
The first suggestion of a stellar wind is given by

Parker, E.N. (1958) *Astrophys. J.* **128**, 664

A discussion of observational data is given in

Renzini, A. (1977) in *Advanced Stages of Stellar Evolution*, Saas-Fee Advanced Course 7,
    ed. by A. Maeder, P. Bouvier (Swiss Society for Astrophysics and Astronomy, Geneva
    Observatory, Sauverny) p. 151

and in

Chiosi, C., Stalio, R. (eds.) (1981) *Effects of Mass Loss on Stellar Evolution*, IAU Colloq.
    59 (Reidel, Dordrecht)

A systematic theoretical analysis is

Holzer, T.E., Fla, T., Lear, E. (1983) *Astrophys. J.* **275**, 808

Holzer, T.E. (1987) in *Circumstellar Matter*, IAU Colloq. 122, ed. by J. Appenzeller,
    C. Jordan (Reidel, Dordrecht) p. 289

Hot stars are discussed by

Kudritzki, R.P., Pauldrach, A., Puls, J. (1986) in *New Insights in Astrophysics*, Proc.
    Joint NASA/ESA/SERC Conference, University College London (ESA SP-263)
    p. 247

See also

Chiosi, C., Stalio, R. (eds.) (1981) *Effects of Mass Loss on Stellar Evolution*, IAU Colloq.
    59 (Reidel, Dordrecht)

Iben, I., Jr., Renzini, I. (eds.) (1981) *Physical Process in Red Giants* (Reidel, Dordrecht)

Underhill, A.B., Michalitsianos, A.G. (eds.) (1984) *The Origin of Non-radiative Heating/Momentum in Hot Stars* (NASA 2858)

The subject of the solar corona has undergone many developments since the first papers by Biermann (1946, 1948), Schwarzschild (1948), and Schatzman (1949). See the review paper by

Kuperus, M., Ionson, J.A., Spicer, D.S. (1981) *Annu. Rev. Astron. Astrophys.* **19**, 7

and the update by

Ionson, J.A. (1985) *Sol. Phys.* **100**, 289

Other works cited in the text are

Hills, J.G., Dale, T.M. (1973) *Astrophys. J.* **185**, 937
Lamers, H.J.G.L.M. (1981) *Astrophys. J.* **245**, 593
Reimers, D. (1977) *Astron. Astrophys.* **61**, 217
Weidemann, V., Koester, D. (1983) *Astron. Astrophys.* **121**, 77

Chapter 5

The problems of convection are far from satisfactory solution. The classic discussion of mixing-length theory is given in

Cox, J., Giuli, R.T. (1968) *Principles of Stellar Structure* (Gordon and Breach, London) vol. II

See also the review papers by

Spiegel, E. (1971) *Annu. Rev. Astron. Astrophys.* **9**, 323
Spiegel, E. (1972) *Annu. Rev. Astron. Astrophys.* **10**, 261

Convective overshooting was first discussed by

Shaviv, G., Salpeter, E.E. (1973) *Astrophys. J.* **184**, 191

The modal theory analysis was made by

Massaguer, J.M., Latour, J., Toomre, J., Zahn, J.P. (1984) *Astron. Astrophys.* **140**, 1

Systematic treatments are given in

Baglin, A. (ed.) (1983) *Instabilités Hydrodynamiques et Applications Astrophysiques,* Ecole de Goutelas, 5–9 April 1983 (SFSA, 61 avenue de l'Observatoire, Paris)

Instabilities in radiative regions are discussed by J.P. Zahn in

Cox, A.N., Vauclair, S., Zahn, J.P. (eds.) (1983) *Astrophysical Processes in Upper Main Sequence Stars,* Saas-Fee Advanced Course 13 (Swiss Society for Astrophysics and Astronomy, Geneva Observatory, Sauverny) p. 253
Durney, B.R. (ed.) (1987) *The Internal Solar Angular Velocity* (Reidel, Dordrecht) p. 201

Goupil, M.J., Zahn, J.P. (eds) (1990) *Rotation and Mixing in Stellar Interiors* (Springer, Berlin Heidelberg)

Ledoux, P., Noels, A., Rogers, A.W. (eds.) (1974) *Stellar Instability and Evolution*, IAU Symp. 59 (Reidel, Dordrecht) p. 185

Circulations are reviewed by

Mestel, L. (1965) in *Stars and Stellar Systems*, vol. VIII, *Stellar Structure*, ed. by G.P. Kuiper, B.M. Middlehurst (University of Chicago Press) p. 465

See also

Tassoul, J.L. (1978) *Theory of Rotating Stars* (Princeton University Press)

For the dynamo problem see

Cowling, T.G. (1981) The present status of dynamo theory, *Annu. Rev. Astron. Astrophys.* 19, 115

and the Goutelas course:

Baglin, A. (ed.) (1985) *Champs Magnétiques Stellaires* (SFSA, 61 avenue de l'Observatoire, Paris)

Three important books are

Moffatt, H.K. (1978) *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge University Press)

Parker, E. (1979) *Cosmical Magnetic Fields* (Clarendon, Oxford)

Zeldovich, Ya.B., Ruzmaikin, A.A., Sokoloff, D.D. (1983) *Magnetic Fields in Astrophysics* (Gordon and Breach, London)

The idea of magnetic braking of stellar rotation is due to

Schatzman, E. (1962) *Ann. Astrophys.* 25. 18

Gravitational settling was studied by

Aller, L.H., Chapman, S. (1960) *Astrophys. J.* 132, 461

Radiation field effects were introduced by

Michaud, G. (1970) *Astrophys. J.* 160, 641

The limiting role of turbulent separation was described by

Schatzman, E. (1969) *Astron. Astrophys.* 3, 331

See also the review paper

Vauclair, S., Vauclair, G. (1982) Element segregation in stellar outer layers, *Annu. Rev. Astron. Astrophys.* 20, 37

The instability or Schwarzschild condition is given in

Schwarzschild, K. (1906) *Göttinger Nachrichten*, p. 41

von Zeipel's theorem is given in

von Zeipel, H. (1924) in *Festschrift f. H. v. Seeliger* (Springer, Berlin)

The Eddington–Sweet circulation velocity was established by

Eddington, A.S. (1925) *Observatory* **48**, 73

and

Sweet, P.A. (1950) *Mon. Not. R. Astron. Soc.* **110**, 69, 548

The effects of a molecular weight gradient are explained in

Mestel, L. (1953) *Mon. Not. R. Astron. Soc.* **113**, 716

Inversion of the sense of circulation was found by

Gratton, L. (1945) *Mem. Soc. Astron. Ital.* **17**, 5

and

Öpik, E.F. (1951) *Mon. Not. R. Astron. Soc.* **111**, 278

Mestel's proof is in

Mestel, L. (1957) *Astrophys. J.* **236**, 550

μ-currents are given by Mestel (1965).
The papers by Tassoul and Tassoul are

Astrophys. J. Suppl. Ser. **49**, 317, 1982;
Astrophys. J. **261**, 265, 273, 1982; **264**, 298, 1983; **267**, 334, 1983; **271**, 315, 1983; **279**, 384, 1984; **286**, 350, 1984

Instabilities: We have already given the reference to Zahn (1974). Fjörtoft's method is given in

Eliassen, A., Kleinschmidt, E. (1957) *Handb. Phys.* **48**, 64

The shear instability is given in

Townsend, A.A. (1958) *J. Fluid Mech.* **4**, 361
Dudis, J.J. (1974) *J. Fluid Mech.* **64**, 65

The Goldreich–Schubert–Fricke instability is in

Goldreich, P., Schubert, G. (1967) *Astrophys. J.* **150**, 571
Fricke, K. (1968) *Z. Astrophys.* **68**, 317

Diffusion processes are discussed in

Baglin, A. (1972) *Astron. Astrophys.* **19**, 45
Kippenhahn, R., Thomas, H.C. (1981) in *Fundamental Problems in the Theory of Stellar Evolution*, IAU Symp. 93, ed. by D. Sugimoto, D.Q. Lamb, D.N. Schramm (Reidel, Dordrecht) p. 237

For Cowling's theorem see

Cowling, T.G. (1934) *Mon. Not. R. Astron. Soc.* **94**, 39

The idea of a mean-field electrodynamics is due to

Steenbeck, M., Krause, F., Rädler, K.H. (1966) *Z. Naturf.* **21A**, 1285

following the model of Parker:

Parker, E.N. (1955) *Astrophys. J.* **122**, 293

For descriptions of solar-surface phenomena see Chap. 2 of the present book and

Jordan, S. (ed.) (1981) *The Sun as a Star* (CNRS-NASA, SP-450)

The discovery of the relation between spectral type and stellar rotation velocity is in

Struve, O., Elvey, C.T. (1931) *Mon. Not. R. Astron. Soc.* **91**, 663

and has resulted in many studies. The demonstration of the age–rotation-velocity relation is given in

Demarque, P., Roeder, R.C. (1967) *Astrophys. J.* **147**, 1188
Kraft, R.P. (1967) *Astrophys. J.* **150**, 551
Skumanich, A. (1972) *Astrophys. J.* **171**, 565
Wilson, O.C. (1966) *Astrophys. J.* **144**, 695

See also

Durney, B.R., Latour, J. (1978) *Geophys. Astrophys. Fluid Dyn.* **9**, 241
Greenstein, J.L., Sargent, A.I. (1974) *Astrophys. J. Suppl. Ser.* **28**, 157
Michaud, G. (1970) *Astrophys. J.* **160**, 641
Roxburgh, I.W. (1978) *Astron. Astrophys.* **65**, 281
Simon, T., Herbig, G., Boesgaard, A.M. (1985) *Astrophys. J.* **293**, 551
Spitzer, L. (1962) *Physics of Fully Ionized Gases* (Interscience, New York)
Zahn, J.P. (1991) *Astron. Astrophys.* **252**, 179

Chapter 6

The classic text is

Rosseland, S. (1949) *The Pulsation Theory of Variable Stars* (Clarendon, Oxford)

A fundamental treatment is

Ledoux, P., Walraven, T. (1958) Variable stars, *Handb. Physik* **50**, 353

together with

Ledoux, P. (1958) Stellar stability, *Handb. Phys.* **50**, 605

Two review articles represent an important advance:

Cox, J.P. (1967) The linear theory: initiation of pulsational instability in stars, in *Aero-dynamic Phenomena in Stellar Atmospheres*, IAU Symp. 28, ed. by R.N. Thomas (Willmer Brothers, Birkenhead, UK)
Christy, R.F. (1968) The theory of Cepheid variability, *Q. J. R. Astron. Soc.* **9**, no. 1, p. 13

The basic information source for individual variable stars is

Kukarkin, B.V., Kholopov, P.N., Efremov, Yu.N., Kukarkina, N.B., Kurochin, N.E., Medvedeva, G.I., Perova, N.B., Fedorovich, V.P., Frolov, M.S. (1969) *General Catalogue of Variable Stars*, 3rd edn (Nauka, Moscow)

and

Kukarkin, B.V., Kholopov, P.N., Artukhina, N.M., Federovich, V.P., Frolov, M.S., Goranskij, V.P., Gorynya, N.A., Kurochkin, E.A., Medvedeva, G.I., Perova, N.G., Ponomareva, G.A., Samus, N.N., Shugarov, S.Yul. (1982) *New General Catalogue of Suspected Variable Stars* (Nauka, Moscow)

Recent books on the theory of variable stars are

Cox, J.P. (1980) *Theory of Stellar Pulsations* (Princeton University Press)
Unno, W., Osaki, Y., Ando, H., Saio, H., Shibahashi, H. (1989) 2nd edn, *Non Radial Oscillations of Stars* (University of Tokyo Press)

An important collection of review papers translated from Russian is

Kukarkin, B.V. (ed.) (1975) *Pulsating Stars* (Wiley, New York)

An introduction to the application of dynamical systems theory is

Buchler, J.R., Perdang, J.M., Spiegel, E.A. (eds.) (1985) *Chaos in Astrophysics* (Reidel, Dordrecht)

An interesting discussion of the history of astronomical discoveries is given in

Harwit, M. (1981) *Cosmic Discovery: The Search, Scope and Heritage of Astronomy* (The Harvester press, Brighton)

Other works cited in the text are

Gelly, B., Grec, G., Fossat, E. (1986) *Astron. Astrophys.* **164**, 383
Noyes, R.W., Baliunas, S.I., Belsorene, E., Duncan, D.K., Horne, J., Widrow, L. (1984) *Astrophys. J.* **285**, L23

See also

Baade, W. (1926) *Astron. Nachr.* **228**, 359

Baade, W. (1954) *Trans. Intern. Astr. Union* **8**, 397

Buchler, J.R., Kovacs, G. (1987) *Astrophys. J. Lett.* **320**, L57

Buchler, J.R., Regev, O. (1982) *Astrophys. J.* **263**, 312

Castor, J.I. (1968) *Astrophys. J.* **154**, 793

Christy, R.F. (1966) Pulsation theory, *Annu. Rev. Astron. Astrophys.* **4**, 353

Cowling, T.G. (1941) *Mon. Not. R. Astron. Soc.* **101**, 367

Cox, A.N. (1980) The masses of Cepheids, *Annu. Rev. Astron. Astrophys.* **18**, 15

Eggen, O.J. (1975) *Astrophys. J.* **195**, 661

King, D.S., Cox, J.P. (1968) Pulsating stars, *Publ. Astr. Soc. Pac.* **80**, 365

Leavitt, H. (1912) *Harvard Circ.* 173

Mineur, H. (1944) *Ann. Astrophys.* **7**, 160

Moore, D.W., Spiegel, E.A. (1966) *Astrophys. J.* **143**, 871

Renzini, A. (1977) in *Advanced Stages of Stellar Evolution*, Saas-Fee Advanced Course 7, ed. by P. Bouvier, A. Maeder (Swiss Society for Astrophysics and Astronomy, Geneva Observatory, Sauverny) p. 151

Wesselink, T.J.H. (1946) *Bull. Astron. Inst. Neth.* **10**, 91

Zhevakin, S.A. (1954) *Astron. Zh.* **30**, 161; **31**, 335

Zhevakin, A. (1963) *Annu. Rev. Astron. Astrophys.* **1**, 367

and

Christensen-Dalsgaard, J., Duvall, T.L., Jr., Gough, D.O., Harvey, J.W., Rhodes, E.J., Jr. (1985) *Nature* **315**, 378

Deubner, F.L. (1975) *Astron. Astrophys.* **44**, 371

Deubner, F.L. (1977) in *The Energy Balance and Hydrodynamics of the Solar Chromosphere and Corona*, ed. by R.M. Bonnet, P. Delache (G. de Bussac, Clermont-Ferrand)

Duvall, T.L., Harvey, J.W., Libbrecht, K.G., Popp, B.D., Pomerantz, M.A. (1988) *Astrophys. J.* **324**, 1158

Dziembowski, W.A., Goode, P.R., Libbrecht, K.G. (1989) *Astrophys. J.* **337**, L53

Evans, J.W., Michard, R. (1962) *Astrophys. J.* **135**, 812

Ledoux, P. (1951) *Astrophys. J.* **114**, 373

Leighton, R.B., Noyes, R.W., Simon, G.W. (1962) *Astrophys. J.* **135**, 474

Pekeris, C.L. (1938) *Astrophys. J.* **88**, 189

Rhodes, E.F., Ulrich, R.K., Jr., Simon, G.W. (1977) *Astrophys. J.* **218**, 901

Roddier, F. (1965) *Ann. Astrophys.* **28**, 463

Stebbins, J. (1953) *Publ. Astron. Soc. Pac.* **65**, 118

Thomson, W. (Lord Kelvin) (1863) *Philos. Trans. R. Soc. London* **153**, 612

Chapter 7

The phenomenology of stellar activity is presented in

Hartmann, L.W., Noyes, R.W. (1987) *Annu. Rev. Astron. Astrophys.* **25**, 27

Solar and stellar radio emission is discussed in

Dulk, G.A. (1985) *Annu. Rev. Astron. Astrophys.* **23**, 169

Stellar X-ray emission is discussed in

Rosner, R., Golub, L., Vaiana, G.S. (1985) *Annu. Rev. Astron. Astrophys.* **23**, 413
Vaiana, G.S. et al. (1981) *Astrophys. J.* **245**, 163

Stellar activity cycles were discovered by

Wilson, O.C. (1978) *Astrophys. J.* **226**, 379

See also

Baliunas, S.L., Vaughan, A.H. (1985) *Annu. Rev. Astron. Astrophys.* **23**, 413

The following papers discuss solar and stellar magnetic fields:

Babcock, H.W. (1962) in *Stars and Stellar Systems*, vol. II, ed. by W.A. Hiltner (University of Chicago Press) p. 107
Brown, D.N., Landstreet, J.D. (1981) *Astrophys. J.* **246**, 899
Donati, J.F., Semel, M., Praderie, F. (1989) *Astron. Astrophys.* **225**, 464
Marcy, G.W. (1984) *Astrophys. J.* **276**, 286
Robinson, R.D. (1980) *Astrophys. J.* **239**, 961
Seares, F.H. (1913) *Astrophys. J.* **38**, 99
Semel, M. (1989) *Astron. Astrophys.* **225**, 459
Skumanich, A., Smythe, C., Frazier, E.N. (1975) *Astrophys. J.* **200**, 747
Saar, S.H. (1988) *Astrophys. J.* **324**, 441

Other works cited in the text are

Antiochos, S.K., Noci, G. (1986) *Astrophys. J.* **301**, 440
Byrne, R.B. (1989) Solar and Stellar Flares, IAU Colloq. 104, *Sol. Phys.* **121**, 61
Gondoin, P. (1986) *Astron. Astrophys.* **160**, 73
Hartmann, L. et al. (1981) *Astrophys. J.* **249**, 662
Linsky, J.L., Haisch, B.M. (1979) *Astrophys. J.* **229**, L27
Rosner, R., Tucker, W.H. Vaiana, G.S. (1978) *Astrophys. J.* **220**, 643
Thomas, R.N. (1957) *Astrophys. J.* **125**, 260
Uchida, Y. (1986) *Adv. Space Res.* **6**, no. 8, 29
Vaughan, A.H. et al. (1981) *Astrophys. J.* **250**, 276

Chapter 8

The title of this chapter is adapted from an IAU symposium:

Tayler, R.J., Hessler, J.E. (eds) (1974) *The Last Stages of Stellar Evolution* (Reidel, Dordrecht)

The objects studied here have a very large observational database. Some evolutionary stages (white dwarfs, neutron stars) are well understood, at least in principle; others in constrast are active research areas (novae, type I and II supernovae) where there are new important data (SN 1987A).

For basic data on stellar models see

Schatzman, E. (1958) *White Dwarfs* (North Holland, Amsterdam)

Clayton, D. (1968) *Principles of Stellar Evolution and Nucleosynthesis* (McGraw-Hill, New York)

Zeldovich, Ya.B., Novikov, I.D. (1971) *Relativistic Astrophysics* (University of Chicago Press)

Shapiro, S.I., Teukolsky, S.A. (1983) *Black Holes, White Dwarfs and Neutron Stars* (Wiley-Interscience, New York)

For the physics of neutron stars, see

Baym, G., Pethick, C. (1979) Physics of neutron stars, *Annu. Rev. Astron. Astrophys.* **17**, 415

Baym, G., Pethick, C. (1975) Neutron stars, *Annu. Rev. Nucl. Phys.* **25**, 27

There have been many colloquia and symposia devoted to the late stages of stellar evolution, beginning with

Shaler, A.J. (ed) (1939) *Les Novae et les Naines Blanches* (Hermann, Paris)

Particularly noteworthy (with no attempt at completeness!) are

Bancel, D., Signore, M. (eds.) (1984) *Problems of Collapse and Numerical Relativity* (Reidel, Dordrecht)

Burbidge, E.M., Burbidge, G.R., Fowler, W.A., Hoyle, F. (1957) *Rev. Mod. Phys.* **29**, 547

Cosmovici, C.B. (ed.) (1984) *Supernovae and Supernovae Remnants* (Reidel, Dordrecht)

Craig Wheeler, J. (1981) The origin of supernovae, *Reports on Progress in Physics* **44**, 85

Eggleton, E., Mitton, S., Whelan, J. (eds) (1976) *Structure and Evolution of Close Binary Systems*, IAU Symp. n°73 (Reidel, Dordrecht)

Gallagher, J.S., Starrfield, S. (1978) Theory and observation of classical novae, *Annu. Rev. Astron. Astrophys.* **16**, 171

Lingenfelter, R.E., Hudson, H.S., Worall, D.M. (eds.) (1982) *Gamma Ray Transients and Related Astrophysical Phenomena* (AIP, New York)

Livio, M., Shaviv, G. (eds.) (1983) *Cataclysmic Variables and Related Objects* (Reidel, Dordrecht)

Manchester, R.N., Taylor, J.H. (1977) *Pulsars* (Freeman, San Francisco)

Meyerott, R., Gillespie, G.H. (eds.) (1980) *Supernovae Spectra* (AIP, New York)

Mihalas, D, Winkler, K.H. (eds.) (1986) *Radiation Hydrodynamics in Stars and Compact Objects*, IAU Colloq. 89 (Springer, Berlin Heidelberg)

Petschek, A.G. (ed.) (1988) *Supernovae* (Springer, Berlin Heidelberg)

Rees, M.J., Stoneham, R.J. (eds.) (1982) *Supernovae: A Survey of Current Research* (Reidel, Dordrecht)

Schatzman, E. (ed.) (1980) Physics of dense matter, *J. Phys. (Paris) Suppl.* **41**, no. 3

Schramm, D.N., Arnett, W.D. (eds) (1973) *Explosive Nucleosynthesis* (University of Texas Press, Austin)

Schramm, D.N. (ed.) (1977) *Supernovae* (Reidel, Dordrecht)

Shklovsky, I.S. (1968) *Supernovae* (Wiley-Interscience, New York)

Taylor, J.H., Manchester, R.N. (1977) Pulsars, *Annu. Rev. Astron. Astrophys.* **15**, 19

Taylor, J.H., Stinebring, D.R. (1986) Recent progress in the understanding of pulsars, *Annu. Rev. Astron. Astrophys.* **24**, 285

Trimble, V. (1982) Supernovae, *Rev. Mod. Phys.* **54**, 1183, 1982

Van Horn, H.M., Weidmann, V. (eds.) (1979) *White Dwarfs and Variable Degenerate Stars*, IAU Colloq. 83 (University of Rochester)

Woosley, S.E., Weaver, T.A. (1986) The physics of supernovae explosions, *Annu. Rev. Astron. Astrophys.* **24**, 205

Amongst historically important papers we have already mentioned Chandrasekhar's work on white dwarfs (Chap. 3); the first paper giving relativistic corrections is

Kaplan, S.A. (1949) *Notes savantes de l'Université de Lwow* **15**, 110

The first papers on neutron stars are

Landau, L.D. (1932) *Phys. Z. Sowjetunion* **1**, 285

and

Oppenheimer, J.R., Volkoff, G.M. (1938) *Phys. Rev.* **55**, 374

The role of collapse towards a neutron star was pointed out by

Baade, W., Zwicky, F. (1934) *Proc. Nat. Acad. Sci.* **20**, 255

Self-similar solutions have been studied by Sedov. The first edition of his book appeared in 1943:

Sedov, I.I. (1959) *Similarity and Dimensional Methods in Mechanics* (Infosearch, London)

The role of shock waves in novae and supernovae was discussed by

Lebedinski, A.I. (1946) *A.J. URSS* **23**, 15
Muraour, H. (1931) *J. Obs.* **54**, 144
Muraour, H. (1945) *C.R.A.S.* **221**, 200
Rosseland, S. (1946) *Astrophys. J.* **104**, 329
Schatzman, E. (1946) *C.R.A.S.* **222**, 722
Schatzman, E. (1946) *Ann. Astrophys.* **9**, 199

A large number of articles discuss the "Supernova of the Century" (SN 1987A). As an introduction, see

Alloin, D., Schatzman, E. (1987) *La Recherche* **18**, 1494

A first summary is given in

Kafatos, M. (ed.) (1988) *Supernova 1987A in the Large Magellanic Cloud* (Cambridge University Press)

The five first years of study of Supernova 1987A are reviewed in

Chevalier, R.A. (1992) *Nature* **355**, 691

Other works cited in the text are

Baym, G., Pethick, C., Sutherland, P. (1971) *Astrophys. J.* **170**, 299
Bethe, H.A., Brown, G.E., Applegate, J., Lattimer, J.M. (1979) *Nucl. Phys.* **A 324**, 487

Bonche, P., Vautherin, A. (1981) *Nucl. Phys.* **A 372**, 496

El Eid, M.F., Hillebrandt, W. (1980) *Astron. Astrophys. Suppl. Ser.* **42**, 215

Hewish, A., Bell, S.J., Pilkington, J.D.H., Scott, P.F., Collins, R.A. (1968) *Nature* **217**, 709

Lamb, D.Q., Lattimer, J.M., Pethick, C.J., Ravenhall, D.G. (1978) *Phys. Rev. Lett.* **41**, 1623

Rhoades, C.E., Ruffini, R. (1974) *Phys. Rev. Lett.* **32**, 324

Van Horn, H., Salpeter, E.E. (1969) *Astrophys. J.* **155**, 183

Subject Index

Springer-Verlag
and the Environment

We at Springer-Verlag firmly believe that an international science publisher has a special obligation to the environment, and our corporate policies consistently reflect this conviction.

We also expect our business partners – paper mills, printers, packaging manufacturers, etc. – to commit themselves to using environmentally friendly materials and production processes.

The paper in this book is made from low- or no-chlorine pulp and is acid free, in conformance with international standards for paper permanency.

DATE DUE
DATE DE RETOUR

MAR 2 8 2002	
MAR 2 8 2002	

CARR M^CLEAN

38-296

Schatzman · Praderie   The Stars

This book is a modern treatment of the physics of the stars. It starts with a chapter on stellar fundamentals, and then the story of the lives and deaths of the stars begins. The authors take us on a journey from the sun, a moderately old star, to supernovae, manifestations of dramatic death. On the way, the reader, advanced undergraduate or beginning post-graduate, is presented with such topics as nuclear reactions in stars, stellar evolution, mass loss and stellar wind, the hydro-dynamics of stellar interiors, variability, and solar and stellar activity.

This comprehensive but rigorous text, building on the theoretical and observational advances of recent years, including, for instance, a discussion of astrophysical theory in the light of Supernovae 1987A, will become required reading for all serious students of astronomy.